DIGITAL Enlightenment STUDIES

VOLTAIRE FOUNDATION
UNIVERSITY OF OXFORD

Text reuse as cultural practice: intertextuality in the 18th-century digital archive

Glenn Roe, Sorbonne Université, glenn.roe@sorbonne-universite.fr

Roe G. 2024. 'Text reuse as cultural practice: intertextuality in the 18th-century digital archive'.
In: *Digital Enlightenment Studies* 2, 1–30.
DOI: 10.61147/des.23

Text reuse, encompassing direct citations, paraphrases and allusions, represents a key aspect of intertextuality – a concept central to literary theory since the 1960s. This paper highlights how computational methods, particularly automatic text-reuse detection, can illuminate the complex system of intertextual exchange that informs 18th-century literary culture, focusing on significant works like the *Encyclopédie* and Voltaire's correspondence. By employing advanced techniques such as sequence alignment and social network analysis, we uncover hidden patterns of influence, citation strategies and the subtle interplay between originality and imitation in Enlightenment literature. The paper also considers the implications of these findings for modern understandings of authorship, originality and textuality, drawing connections to contemporary digital humanities practices. The paper ultimately aims to recontextualise the Enlightenment as a period of intense intertextual productivity, where the reuse of texts was not merely a scholarly exercise but a dynamic and essential component of literary creation.

Keywords: intertextuality, text reuse, network analysis

A preliminary version of this article was originally given as the first Voltaire Foundation Lecture in Digital Enlightenment Studies, at the Bodleian Library, Oxford, 2 May 2024.

1. Introduction

Text reuse has a rich and varied history for literary culture, encompassing a range of practices from direct citation, commentary and criticism to indirect references, paraphrases and allusions. 'Nous ne faisons que nous entregloser' goes the line by Montaigne, who warns us that all literature is in essence just a rehashing of what came before (2004, p.1069). This underlying tension between commentary and creation, or imitation and innovation, is in many ways the dialectic that drives literary practice and theory forward. In France, this conflict comes to a head at the turn of the 18th century, with the *Querelle des Anciens et des Modernes* ushering in a long and sometimes painful process of literary modernisation. The 18th century is thus an important transitory period in the long history of text reuse, one in which related concepts such as copyright and plagiarism begin to be discussed and codified.

This article aims to explore text reuse from both a theoretical and technical perspective, while demonstrating the potential of computational methods for the automatic detection of text reuses as a means for exploring the growing digital archive of the 18th century. As digital resources and collections reach a critical mass for Enlightenment studies – a field that has benefited greatly from the mass digitisation efforts of the past two decades (Burrows and Roe 2020) – new data-intensive methods are needed in order to exploit these resources to their fullest. Text-reuse detection is one such method. Technically, it is deceptively simple in its approach to text analysis: find two similar strings in at least two different documents. Theoretically, however, this simple implementation of text reuse is in need of both contextualisation – providing identified passages with an appropriate amount of historicity – and interpretation, which requires new forms of computational hermeneutics (Dobson 2019).

More generally, the need for deep domain expertise – philologist, historicist, hermeneutical – is essential for understanding the output of even the most robust computational models and methodologies; and it is here that humanists have much to offer. This is especially true in our current age of so-called artificial intelligence (AI), when large language models (LLMs) capable of reading whole universes of text and generating human-like responses based on this almost infinite archive, are in desperate need of theoretical grounding from the humanities (Dobson 2023). These models are the fever-dream, I would argue, of post-structuralism and we need scholars conversant in *theory* to understand these new artificial textualities. This article is thus a modest attempt at bridging this computational-hermeneutical divide; providing examples of text-reuse practices 'in the wild' as it were, along with both technical and theoretical reflexions on text reuse as an important, and perhaps overlooked, component of 18th-century literary culture.

From a computational perspective, text-reuse detection as a technique is best understood as a text-based instance of the 'longest common subsequence' (LCS) problem in computer science (Bergroth et al. 2000). Determining the LCS in texts is a long-standing challenge for natural language processing and forms

the foundation for data comparison tools like the 'diff' utility (Hunt and McIlroy 1976). It is also a significant challenge in fields such as computational linguistics and bioinformatics. LCS algorithms are integral to version control systems like Git, which use them to merge various modifications made to a set of revision-controlled files (Blischak, Davenport and Wilson 2016). Similar approaches applied specifically to text reuse applications date back to the early 2000s, when Paul Clough et al. (2002) at the University of Sheffield began experimenting with measuring text reuse in large journalistic corpora. Its initial definition is quite pragmatic: 'A topic of considerable theoretical and practical interest is that of *text reuse*: the reuse of existing written sources in the creation of a new text' (p.152). Over the past 15 years or so, text-reuse projects and applications developed in the digital humanities have grown both in terms of number and methodological variety. Freely available tools for automatic text-reuse detection include BLAST,[1] passim,[2] Tesserae,[3] Text-PAIR,[4] textreuse[5] and TRACER,[6] among others. These tools were developed to address a wide range of scholarly use-cases ranging from: bioinformatic gene sequencing; reprint detection in antebellum American and Finnish newspapers; identifying allusions in classical poetry; early-modern French dictionaries and encyclopaedias; 18th-century print culture in England; and biblical reception, to name but a few.

In the field of 18th-century studies, text reuse has also found fruitful applications for understanding the growing digital cultural record. Mass datasets such as Gale's Eighteenth Century Collections Online (ECCO) have in particular inspired teams working in Chicago (Gladstone and Cooney 2020) and Helsinki (Ryan, Mahadevan and Tolonen 2023) to explore text-reuse practices from philosophical, editorial and book-historical perspectives at a scale frankly unimaginable just two decades ago. These projects have not only moved research forward in their respective disciplines, but also represent important technical interventions for the wider digital humanities community, highlighting concerns over sustainable tool development, the accessibility of proprietary data and the use and design of data-rich interfaces for complex humanities corpora (Rosson et al. 2023).

From a theoretical perspective, text reuse can be seen as a concrete form of intertextuality – a concept that emerged in the 1960s and 1970s in France. Julia Kristeva was the first to use the term, translating Bakhtin's notion of dialogism, and it subsequently gained traction within the *Tel Quel* circle (Kristeva 1969). As an expansive literary-theoretical concept, there are numerous and varied approaches to both the theory and practice of intertextuality.[7] A conservative perspective might align with Harold Bloom's theory of 'influence', where the art of criticism involves understanding the underlying connections that link one poem to another. In this view, writers grapple with the great giants of the past, necessarily borrowing from their predecessors to assert their own poetic authority. For Bloom: 'Criticism is the art of knowing the hidden roads that go from poem to poem' (1973, p.96).

[1] https://blast.ncbi.nlm.nih.gov/Blast.cgi. See also Vesanto et al. (2017); Salmi et al. (2020).

[2] https://github.com/dasmiq/passim/. See also Smith, Cordell and Mullen (2015).

[3] https://tesserae.caset.buffalo.edu/. See also Coffee et al. (2013).

[4] https://github.com/ARTFL-Project/text-pair/. See also Olsen, Horton and Roe (2011).

[5] https://github.com/ropensci/textreuse/. See also Li and Mullen (2020).

[6] https://www.etrap.eu/research/tracer/. See also Büchler et al. (2014); Franzini et al. (2018).

[7] For an excellent overview of Intertextuality see Samoyault (2005).

A more radical take on intertextuality, appearing at the same time, comes from Roland Barthes, who posits that: 'Tout texte est un intertexte; d'autres textes sont présents en lui, à des niveaux variables, sous des formes plus ou moins reconnaissables: les textes de la culture antérieure et ceux de la culture environnante; tout texte est un tissu nouveau de citations révolues' (1973). Somewhere between these two extremes lies a more modest middle ground, one that is perhaps more tractable in computational terms, and which aligns itself with Gérard Genette's later interpretation of intertextuality as (1982, p.7):

> [U]ne relation de coprésence entre deux ou plusieurs textes, c'est-à-dire, eidétiquement et le plus souvent, par la présence effective d'un texte dans un autre. Sous sa forme la plus explicite et la plus littérale, c'est la pratique traditionnelle de la *citation* (avec guillemets, avec ou sans référence précise); sous une forme moins explicite et moins canonique, celle du *plagiat* […] qui est un emprunt non déclaré, mais encore littéral.

2. 18th-century text reuse

Today, reuse is widely considered as a form of plagiarism. This stems from the modern belief that words (and by extension, ideas) belong to specific authors and cannot be reused without their permission (Mallon 1989). This was not always the case, however, especially in the 18th century, as we shall see. While accusations of plagiarism were somewhat commonplace in the early-modern period, the widespread practice of creative reuse was an established aesthetic paradigm (Gevrey 1995, p.5):

> Si, comme Furetière, on tient pour auteur 'celui qui n'a pas pris ouvrage d'un autre', bien peu d'écrivains du XVII^{e} siècle satisfont à cette condition. Corneille et Mme de Lafayette furent accusés de plagiat; on se plut à lire le *Virgile travesti* ou à reconnaître dans les *Caractères* de La Bruyère des pensées et des maximes empruntées. Dans les modèles exemplaires qu'il réécrivait, le XVII^{e} siècle chercha la mesure de la généralité indispensable aux sujets de morale.

Indeed, the culture of citation during this period oscillated between the opposing poles of originality and imitation. In the early 18th century, for example, the classical notion of imitation still held sway, and was predicated on the belief that the value of a work derived from its conformity to the great works of the past. Through these works, one could access truth, beauty, nature and other higher ideals that the ancients exemplified (Norman 2015). For instance, Alexander Pope, speaking of Virgil, conflates the ancient author with nature itself, both of which merit imitation (1713, v.133–36):

> Perhaps he [Virgil] seem'd *above* the Critick's Law,
> And but from *Nature's Fountains* scorn'd to draw:
> But when t' examine ev'ry Part he came,
> *Nature* and *Homer* were, he found, the *same*.

And yet, from the mid-18th century onwards literary works gradually began to be understood as the emanation of what was unique in each individual – their genius (Schöch 2011). Concurrently, poets developed a new and immediate relationship with nature, one that did not need to be mediated by representations of the ancients.

To observe this shift in action, we need only compare Racine's second preface to *Britannicus*, published in 1674, in which the classical playwright praises the conformity of his work with that of Tacitus: 'J'avais copié mes personnages d'après le plus grand peintre de l'Antiquité, je veux dire d'après Tacite, et j'étais alors si rempli de la lecture de cet excellent historien, qu'il n'y a pas un trait éclatant dans ma tragédie, dont il ne m'ait donné l'idée' (2015, p.9). Contrast this with Rousseau, who, just a century later, proudly declares in the opening of his *Confessions* that he had neither predecessor nor successor for his undertaking: 'Je forme une entreprise qui n'eut jamais d'exemple, et dont l'exécution n'aura point d'imitateurs' (Rousseau 2022, p.41). Never mind the fact that Rousseau's work is generically borrowed from Saint Augustine and, thematically, continues the semi-autobiographical work of Montaigne (*pace* Jean-Jacques). Indeed, Montaigne is perhaps best understood as a key figure in the history of text reuse, exemplifying both the inherent ambiguity of a writer deeply imbued with the tradition of imitation – inserting his 'fleurs étrangères'[8] into his text as he goes along – while simultaneously aware of his identity as a unique and original author, intimately presenting himself to the reader (Compagnon 1979; Compagnon 1980).

Montaigne's project, filtered through Pascal and the 17th-century moralists, reaches a certain apogee with Rousseau, whose work underscores the gradual transition away from a culture of imitation towards one of individual originality; an evolution that will ultimately transform our understanding of authorship and textuality as the literature of the Enlightenment gives way to the Romantic era. In this manner, the pre-Romantic notion of originality was in fact tied to the various practices of text reuse in the 18th century. Roland Mortier has demonstrated that the concept of originality is relatively new and, more significantly, should be understood as an offshoot of French Enlightenment aesthetics (Mortier 1981).

Following Mortier's argument, 'originality' as an aesthetic category did not emerge in French until the very end of the 17th century. Roger de Piles first uses the term in his *Abrégé de la vie des peintres* (1699) to describe the originality of paintings – unique works that could not be copied: 'S'il y a des choses qui semblent favoriser l'originalité d'un Ouvrage, il y en a aussi qui paroissent la détruire; comme la répétition du même Tableau, l'oubli où il a été durant beaucoup de tems, & le prix modique qu'il a coûté' (p.100). Diderot later employs the term to describe a person of a very original character, specifically the Danish minister to France, Baron de Holberg, in a 1759 letter to Sophie Volland: 'Nous dinâmes tous d'appétit. Notre baron, le nôtre, fut d'une folie sans égale. Il a de l'originalité dans le ton et dans les idées. Imaginez un satyre gai, piquant, indécent et nerveux, au milieu d'un groupe de figures chastes, molles et délicates. Tel il étoit entre nous' (1950, p.28).

[8] 'Comme quelqu'un pourroit dire de moy que j'ay seulement faict icy un amas de fleurs estrangeres, n'y ayant fourny du mien que le filet à les lier' (Montaigne 2004, p.1055).

As a lexicographical item, 'originality' made its first dictionary appearance in the 1743 edition of the *Dictionnaire de Trévoux*, with the Jesuit editors proudly proclaiming that 'Ce mot ne se trouve point encore dans les Dictionnaires' (*Dictionnaire universel françois et latin* 1743, vol.4, p.1491).[9] This was followed by a somewhat laconic entry in the fourth edition of the *Dictionnaire de l'Académie française*, which defined originality simply as the 'Caractère de ce qui est original. Il se dit Des personnes & des choses' (1762, vol.2, p.266). The semantic shift between 'original' as a descriptor and 'originality' as an aesthetic or epistemological category – moving from character to quality – reflects the evolution of a concept whose modern acceptation will only crystallise at the beginning of the 19th century.[10]

The great mid-century *Encyclopédie*, edited by Diderot and D'Alembert, in many ways exemplifies the above transition. While certainly still informed by the aesthetics of imitation inherited from their 17th-century forebears, the *encyclopédistes* nonetheless hint at its gradual exhaustion over the course of the work's 20-year publication. The unsigned article 'Original', for example, still holds firm to the doctrine of imitation, defining an original as 'le premier dessein, ou instrument authentique de quelque chose, & qui doit servir comme de modele ou d'exemple à être copié ou imité' (Diderot and D'Alembert 2022, vol.11, p.648). And yet, it also acknowledges the rarity of true originality, noting that 'aujourd'hui l'on trouve à peine aucun titre ancien de possession, inféodation, &c. qui soit original; ce ne sont que des *vidimus*, ou copies collationnées sur les originaux'.

The second article for the headword 'Original', in its grammatical sense, includes nothing more than a cross-reference to the article 'Originalité'. As we know, cross-references in the *Encyclopédie* are far from neutral, and Diderot's use (and abuse) of cross-references is particularly striking (Starobinski 1995). Thus, at the other end of the cross-reference – following an early-modern hyperlink as it were – we are brought to the article 'Originalité', which, while technically unsigned, is likely written by Diderot.[11] Here we find that originality is considered the 'maniere d'exécuter une chose commune, d'une maniere singuliere & distinguée: l'originalité est très-rare. La plupart des hommes ne sont en tous genres, que des copies les uns des autres. Le titre d'*original* se donne en bonne & en mauvaise part (Diderot and D'Alembert 2022, vol.11, p.648). Diderot thus underscores the fundamental tension between imitation and originality that plagues any work of compilation such as a dictionary or his very own *Encyclopédie*.

[9] 'ORIGINALITÉ, s. f. Qualité qui fait qu'une chose est originale. *Originalitas*. Ce mot ne se trouve point encore dans les Dictionnaires: c'est de Piles qui l'emploie dans un Ouvrage qui regarde les Peintres & la peinture, où il dit qu'il est bien difficile de connoître l'originalité d'un tableau.'

[10] See, for example, the entry in the Académie's sixth edition of 1835: 'ORIGINALITÉ. s. f. Qualité de ce qui est original; caractère de ce qui est neuf, sans modèle de même nature, digne de servir de modèle. *L'originalité est une des qualités qui constituent le beau dans les arts. L'originalité n'est pas la bizarrerie. L'originalité d'une pensée, d'une expression. Il a de l'originalité dans l'esprit. Son style a de l'originalité, un caractère d'originalité fort piquant*' (1835, vol.2, p.315).

[11] See the authorship attribution page of the ARTFL *Encyclopédie*: https://encyclopedie.uchicago.edu/node/163.

From the outset, Diderot and D'Alembert faced numerous accusations of impiety, improper use of authorities and outright plagiarism (Albertan 1992). D'Alembert responded to these detractors in the editors' foreword to the third volume (Diderot and D'Alembert 2022, vol.3, pp.xvii–xviii), explaining that:

> L'Encyclopédie doit donc par sa nature contenir un grand nombre de choses qui ne sont pas nouvelles [...] Parmi les différens ouvrages qu'on a accusé l'Encyclopédie d'avoir mis à contribution, on a sur-tout nommé les autres Dictionnaires [..] la ressemblance qui se trouve quelquefois entre un article de l'Encyclopédie & un article de quelque Dictionnaire, est forcée par la nature du sujet, sur-tout lorsque l'article est court, & ne consiste qu'en une définition ou en un fait historique peu considérable: cela est si vrai, que sur un grand nombre d'articles la plûpart des Dictionnaires se ressemblent, parce qu'ils ne sauroient faire autrement.

If we turn to the *Encyclopédie* article 'Compilateur', which was authored, rather ironically, by one of the great text reusers of the *Encyclopédie*, the Abbé Mallet, we see emerge an interesting counter perspective. Mallet defines a compiler as an 'écrivain qui ne compose rien de génie, mais qui se contente de recueillir & de répeter ce que les autres ont écrit (Diderot and D'Alembert 2022, vol.3, p.762). While Mallet's compiler is very much not an advocate of originality, he goes on to note that for the most part lexicographers are also merely compilers called by a different name. This viewpoint likely irked Diderot, and, to rebut any ambiguity concerning lexicographical compilation, a cross-reference is included at the end of Mallet's article, subtly directing readers to the entry 'Plagiaire'.

'Plagiaire' begins by defining the term as an 'écrivain qui pille les autres auteurs, & donne leurs productions comme étant son propre ouvrage'. However, in a clear rebuke of Mallet's previous assertion, the article author, most likely Diderot, elaborates further (Diderot and D'Alembert 2022, vol.12, p.680) that:

> Les Lexicographes, au moins ceux qui traitent des arts & des sciences, paroissent devoir être exemts des lois communes *du mien* & *du tien*. Ils ne prétendent ni bâtir sur leur propre fonds, ni en tirer les matériaux nécessaires à la construction de leur ouvrage. En effet le caractere d'un bon dictionnaire tel que nous souhaiterions de rendre celui-ci, consiste en grande partie à faire usage des meilleurs découvertes d'autrui: ce que nous empruntons des autres nous l'empruntons ouvertement, au grand jour, & citant les sources où nous avons puisé. La qualité de compilateurs nous donne un droit ou un titre à profiter de tout ce qui peut concourir à la perfection de notre dessein, quelque part qu'il se rencontre. Si nous dérobons, c'est seulement à l'imitation des abeilles qui ne butinent que pour le bien public, & l'on ne peut pas dire exactement que nous pillons les auteurs, mais que nous en tirons des contributions pour l'avantage des lettres.

Thus, in the context of the *Encyclopédie*, plagiarism is portrayed as both negative, in its most restrictive and legalistic sense, and, given the right context, potentially positive, especially when it comes to dictionary-making,

where the judicious reuse of past authorities becomes an almost moral imperative. This positive spin on plagiarism will evolve further in the work of later writers. Voltaire, for example, in his *Questions sur l'Encyclopédie* – a significant work of text reuse in its own right – redefines the term in a more commercial sense (2012, pp.438–39):

> Quand un auteur vend les pensées d'un autre pour les siennes, ce larcin s'appelle *plagiat*. On pourrait appeler *plagiaires* tous les compilateurs, tous les faiseurs de dictionnaires, qui ne font que répéter à tort et à travers, les opinions, les erreurs, les impostures, les vérités déjà imprimées dans des dictionnaires précédents; mais ce sont du moins des plagiaires de bonne foi; ils ne s'arrogent point le mérite de l'invention. Ils ne prétendent pas même à celui d'avoir déterré chez les anciens les matériaux qu'ils ont assemblés; ils n'ont fait que copier les laborieux compilateurs du seizième siècle. Ils vous vendent en *in-quarto* ce que vous aviez déjà en *in-folio.* Appelez-les, si vous voulez, libraires, et non pas auteurs. Rangez-les plutôt dans la classe des fripiers que dans celle des plagiaires.

In Panckoucke's *Supplément à l'Encyclopédie* (1777), Marmontel argues that being labelled a plagiarist should by no means give one pause, as it is 'une sorte de crime littéraire, pour lequel les pédans, les envieux, et les sots ne manquent pas de faire le procès aux écrivains célèbres' (p.388). While it might entail the theft of thoughts, the outcry against plagiarism is often exaggerated, according to Marmontel. On the contrary, he continues, 'Quiconque met dans son vrai jour, soit par l'expression, soit par l'à-propos, une pensée qui n'est pas à lui, mais qui sans lui serait perdue, se la rend propre en lui donnant un nouvel être, car l'oubli ressemble au néant' (p.389). In this context, the plagiarist serves a productive purpose, saving the best bits of our shared culture from obscurity. Much like public law allows the cultivation of fallow land for the common good, Marmontel thinks literature should follow the same principle (p.389):

> Dans le droit public, la propriété d'un terrain a pour condition la culture: si le possesseur le laissait en friche, la société aurait droit d'exiger de lui qu'il le cédât ou qu'il le fît valoir. Il en est de même en littérature; celui qui s'est emparé d'une idée heureuse et féconde, et qui ne la fait pas valoir, la laisse comme un bien commun, au premier occupant qui saura mieux que lui en développer la richesse.

Viewed as such, plagiarism – or, more generously, text reuse – becomes not only permissible for Enlightenment writers, but can also be seen as a useful tool for shedding light on otherwise forgotten authors, both ancient and modern. Moreover, the productive poetics inherent in this expanded notion of text reuse is also in line with Diderot's preference for philosophical 'eclecticism' over and above the more systematic philosophies of the 17th century (Mulsow 1997). The ambition of the eclectic, Diderot tells us (Diderot and D'Alembert 2022, vol.5, p.270):

> [E]st moins d'être le précepteur du genre humain, que son disciple; de réformer les autres, que de se réformer lui-même; de connoître la vérité, que de l'enseigner. Ce n'est point un homme qui

> plante ou qui seme; c'est un homme qui recueille & qui crible. Il joüiroit tranquillement de la récolte qu'il auroit faite, il vivroit heureux, & mourroit ignoré, si l'enthousiasme, la vanité, ou peut-être un sentiment plus noble, ne le faisoit sortir de son caractere.

Caught somewhere between the pull of imitation and tradition and the lure of originality and genius, the eclectic will become a transitory literary device of the late Enlightenment, applying Diderot's philosophical ideas to other, more varied domains such as aesthetics, literary creation and, as we shall see, 'illuminated' mysticism (Antoine-Mahut 2024).

3. Identifying text reuse in the *Encyclopédie*

This historical divagation leads us back to text-reuse detection as a viable method for uncovering the various intertextual practices of 18th-century authors. With colleagues at the University of Chicago, we began testing these methods on the *Encyclopédie* itself, the text of which was first digitised in the late 1990s by the ARTFL Project and has been freely available on the web since 2007.[12] Our first experiments were aimed at automatically identifying potential sources of the *Encyclopédie*, testing various tried-and-true information retrieval techniques such as lexical similarity, distributional semantics and vector-space analysis to identify similar articles in other dictionaries. While these 'bag-of-words' techniques yielded fairly interesting results, confirmation that 'similar' articles were in fact borrowed or reused from previous dictionaries continued to elude us. Thus, we began developing an interest in more literal text-reuse detection, experimenting with the BLAST and Text-PAIR tools mentioned earlier, in order to apply sequence-based alignment (similar to those used in bioinformatic DNA analysis) to our datasets (see Allen et al. 2010; Olsen, Horton and Roe 2011; Roe 2012).

Building off these initial experiments, we have continued using the Text-PAIR system to identify regions of text reuse in the *Encyclopédie* and, progressively, other literary-historical corpora. A robust and flexible system for text-reuse detection, Text-PAIR effects scalable and fast comparisons of text collections of varying size and quality. Conceived specifically for digital humanities applications, Text-PAIR is highly fault-tolerant in its matching parameters, which allows it to find similar passages with significant changes and/or errors in transcription such as those generated by optical character recognition (OCR).[13] Although the identified pairwise passages represent a deceptively straightforward notion of intertextuality – where the same text (or roughly the same text) appears in two different places – our experience over the years has indicated that this simple approach can reveal a large and varied range of text-reuse practices, from direct quotations to literary allusions, and everything in between.

[12] See https://encyclopedie.uchicago.edu/.

[13] First conceived in 2009 by Mark Olsen at ARTFL and written in the Perl programming language, Text-PAIR has since 2015 been rewritten in Python and maintained by Clovis Gladstone. I would like to thank both Mark and Clovis for their invaluable input on improving and modifying the system over the years. See https://artfl-project.uchicago.edu/text-pair.

For example, it is unsurprising that the *Encyclopédie* cites or uses Pierre Bayle's *Dictionnaire historique et critique* (1697), a known predecessor and inspiration for the project. And indeed, we find a long sequence at the end of the article 'Spinosa, philosophie de' (Diderot and D'Alembert 2022, vol.15, pp.473–74) that is clearly derived from Bayle. What is surprising, however, is that the *encyclopédiste* (thought to be the Abbé Yvon) concludes the article in Bayle's voice, reusing the first-person pronoun that begins the sentence, 'Je finis par dire…', without any quotation marks to indicate a direct quotation. While Bayle is mentioned at the beginning of the article, this blending of voices signifies something more nuanced: Bayle's voice becomes part of the authorial fingerprint of the *Encyclopédie* in this and other articles. Whether readers at the time recognised this remains uncertain; but it is now clear that this is an example of a non-explicit reuse, which was only unearthed using computational methods.

This phenomenon is observed with numerous other works. The use of John Locke, for instance, in the *Encyclopédie* is particularly intriguing. Locke's *Second Treatise* was banned and burned in France and was put on the *Index* of prohibited books. Explicitly citing Locke would have attracted the censor's attention, risking the loss of publishing privilege for the *Encyclopédie*. However, the chevalier de Jaucourt – author of some 17 000 *Encyclopédie* articles – seemed determined to incorporate Locke's ideas into his contributions and did so surreptitiously. The article 'Gouvernement' (Diderot and D'Alembert 2022, vol.7, pp.788–91), for example, is littered with passages taken from the French translation of Locke and woven into Jaucourt's arguments seamlessly with no attribution whatsoever. This covert incorporation of Locke further exemplifies how text reuse in the 18th century functioned as a means of generating meaning through the often-unacknowledged redeployment of authoritative or subversive models.

These examples were identified as part of a collaborative research project which uncovered a sophisticated system of citation strategies within the *Encyclopédie* (Edelstein, Morrissey and Roe 2013). The use of loose, or non-existent, citations and naming conventions, coupled with other subtle cues, meant that the *encyclopédistes* were able to smuggle in passages from prohibited works while avoiding (sometimes) the censor's oversight. This system of implicit references may have been clear to some 18th-century readers but is less obvious to contemporary audiences. Specialists might discern these nuances, but for the most part we must rely on computational analysis to reveal these intricate layers of text reuse. Viewed in this light, the *Encyclopédie* functions not only as a repository of knowledge but also as a complex network of interwoven texts, connected through both direct and indirect citations, each reflecting the deep intertextual currents of its time.

Uncovering these hidden voices in the *Encyclopédie* helped us better understand its status as a polyphonic philosophical work, wherein authors such as Bayle and Locke become *de facto encyclopédistes* in their own right. While approximately 140 named contributors to the *Encyclopédie* are known, there are likely many more whose contributions remain unidentified. One such figure is Émilie Du Châtelet. As Judith Zinsser has stated, many *Encyclopédie* authors: 'simply took whole passages

from [Du Châtelet's] *Institutions de physique* for subjects as general as "Time" and "Hypothesis" and as specific as the descriptions of the metaphysical concepts of Leibniz and Wolff. Few identified the original author, and so Du Châtelet acquired the title of so many other women before and after her: "anonymous" ' (2006, p.285). Using a digitised version of the *Institutions*, we tested this hypothesis using Text-PAIR, discovering some 13 articles that contain substantial content 'borrowed' from Du Châtelet (Roe 2018). The article 'Contradiction', to take but one example, is actually 95 per cent Du Châtelet. Its stated author, Samuel Formey, was the Secretary of the Academy of Berlin, and originally intended to create his own philosophical dictionary but, after abandoning the idea, he later handed over his papers to D'Alembert. Unbeknownst to D'Alembert, Formey's articles heavily cribbed from Du Châtelet's *Institutions* (and undoubtedly other sources), and these unattested borrowings were then incorporated into the *Encyclopédie* with Formey credited as sole author. This complicated lineage begs the question of Du Châtelet's status as an author in the *Encyclopédie*, restoring a lost voice to the polyphonic text that had been muted for the last few centuries.

4. Communication circuits and correspondences

Moving beyond the *Encyclopédie*, and working with colleagues in Oxford, we wanted to explore text-reuse practices in other text collections and contexts, tracing their participation in the broader communication networks of 18th-century print culture, and drawing inspiration from Robert Darnton's now-famous communication circuit model (see Figure 1), which demonstrates the flow of information through various media channels in 18th-century France (Darnton 1996, p.189). With this model in mind, our aim was to use text reuse as sort of relay through which literary value was formed and debated in both the public and private spheres. To do this, we turned our focus to private correspondences, in particular those assembled by the *Electronic Enlightenment* project at Oxford.[14] As we know, private letters acted as a platform for sociability and a means of diffusing information and building literary authority. They were also, as we will see, a great vector for text reuse.

Perhaps no author in the 18th century was more attuned to the literary and social uses of correspondence than Voltaire. Recently, we have begun exploring the extraordinary richness of his more than 21 000-letter correspondence using digital methods (Cronk and Roe 2020). Unsurprisingly, turning Text-PAIR loose on this corpus also provides excellent examples of how text reuse functions in the (semi-)private sphere. Throughout his letters, Voltaire exhibits extraordinary reading and sharing practices, acting much like an early-modern *Reader's Digest*, excerpting works and sharing these passages with his friends. Take, for example, Helvétius' book *De l'homme*, published posthumously between 1771

[14] https://www.e-enlightenment.com/. See also Cronk (2020).

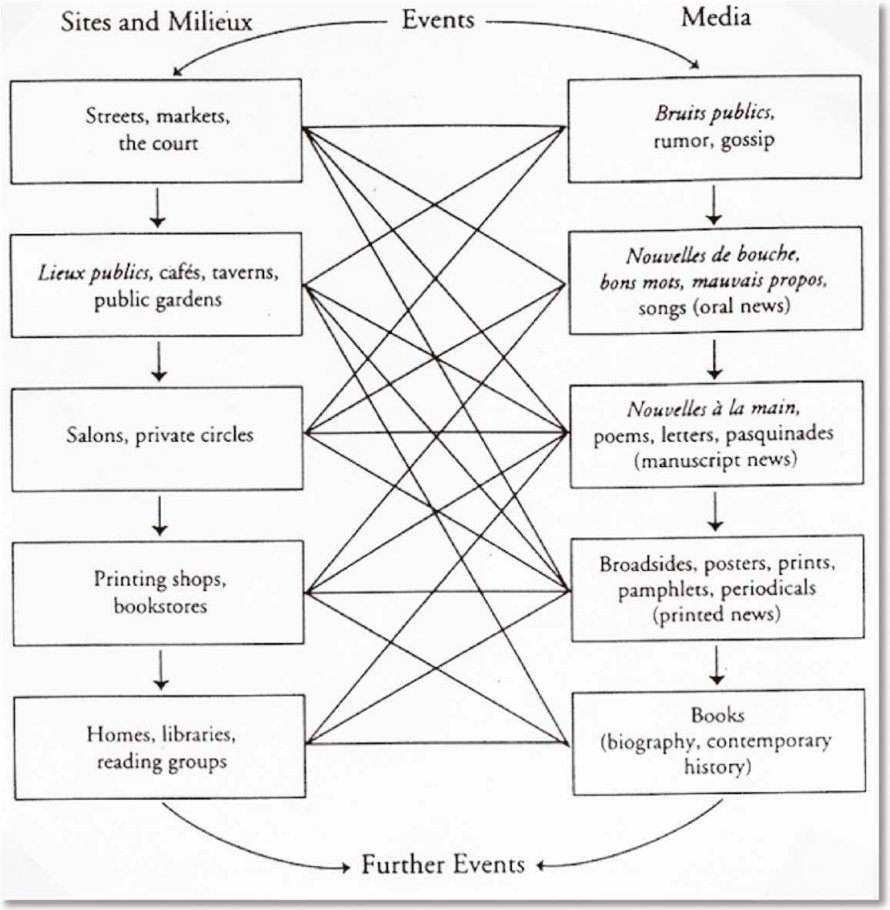


Figure 1 A schematic model of the 'Communication circuit' in 18th-century Paris (Darnton 1996, p.189). Reproduced by kind permission of Robert Darnton.

and 1773, from which Voltaire takes snippets and sends them to D'Alembert, commenting on their quality (D18450, see Voltaire 1975, p.42–43).

> 3 de juillet
> Voici, mon cher et grand philosophe, ma réponse à l'abbé philosophe.
> N'êtes vous pas bien content de ces petits mots d'Helvétius; tome I, page 107?
> 'Nous sommes étonnés de l'absurdité de la religion païenne; celle de la religion papiste étonnera bien davantage la postérité'.
> Et page 102: 'Pourquoi faire de dieu un tyran oriental? pourquoi mettre ainsi le nom de la divinité au bas du portrait du diable? Ce sont les méchants qui peignent dieu méchant. Qu'est ce que leur dévotion? Un voile à leurs crimes'.

> C'est dommage que ce ne soit pas un bon livre; mais il y a de très bonnes choses: c'est une arme qui tiendra son rang dans l'arsenal où nous avons déjà tant de canons qui menacent le fanatisme. Il est vrai que les ennemis ont aussi leurs armes: elles sont d'une autre espèce, elles ont tué le chevalier de la Barre, elles ont blessé à mort Helvétius; mais le sang de nos martyrs fait des prosélytes. Le troupeau des sages grossit à la sourdine.
> Bonsoir, mon sage; bonsoir, mon cher Bertrand; il ne me reste plus qu'un doigt pour tirer les marrons du feu, mais il est à votre service.

Voltaire remarks on the brilliance of certain passages despite his overall low opinion of the work, calling it a bad book but noting the several 'très bonnes choses' he nonetheless found within it. He sends similar excerpts of the same work to Madame Du Deffand, who was blind (D18607, Voltaire 1975, p.164).

> à Ferney 1er 9bre 1773
> Eh bien, Madame, je commence par les diamants brillants. Page 102, tome 1er. *Pourquoi faire de Dieu un tiran oriental? pourquoi lui faire punir des fautes légères par des châtiments éternels? pourquoi mettre le nom de la divinité au bas du portrait du diable?*
> Page 107. *Nous sommes étonnés de l'absurdité de la religion paienne, celle de la religion papiste étonnera bien d'avantage la postérité.*
> Page 121. *Pour être philosophe, dit Mallebranche, il faut voir évidemment, et pour être fidèle il faut croire aveuglément. Mallebranche ne s'aperçoit pas que de son fidèle il en fait un sot.*
> Page 321. *Pourquoi tout moine qui déffend avec un emportement ridicule les faux miracles de son fondateur, se moque t-il de l'existence des vampires? C'est qu'il n'a point d'intérêt à la croire. Otez l'intérêt, reste la raison; et la raison n'est pas crédule.*
> Je prends ces petits diamants au hazard, Madame. Il y en a mille dans ce goût dont l'éclat m'a frappé. Cela n'empêche pas que le livre ne soit très mauvais. Je passe ma vie à chercher des pierres précieuses dans du fumier, et quand j'en rencontre je les mets à part et j'en fais mon profit. C'est par là que les mauvais livres sont quelquefois très utiles.

With this last paragraph in particular, Voltaire sums up the valuable service he has provided to his correspondent by sending her only the best parts of a book, thereby sparing her the effort of reading (or, in the case of Mme Du Deffand, having her secretary read to her) what he considered subpar content.

This anecdote captures another striking form of the poetics of text reuse in the 18th century: the process of accruing knowledge through the dissemination of excerpts. Voltaire's method involved not just extracting valuable passages from texts but also sharing them widely, thereby facilitating their circulation and, as we saw

Marmontel arguing above, safeguarding them from future oblivion. There is thus a deep sociability inherent to 18th-century text-reuse practices, contingent upon someone actively reusing and sharing content through early-modern social networks.

5. Intertextual networks

Finally, moving on to a more recent project exploring practices of text reuse, a team of researchers at the Sorbonne are aiming to identify and analyse these 18th-century intertextual networks on a vast scale. The project, generously funded by the European Research Commission and titled ModERN (Modelling Enlightenment: reassembling networks of modernity using data-driven research), participates in the so-called 'network turn' of digitally-inflected humanities research that has been on the rise over the past ten years.[15] Social network analysis (SNA) – borrowed from the computational social sciences – has proven particularly useful for understanding cultural exchanges across various domains and linguistic traditions. Here, Ruth and Sebastian Ahnert have been instrumental in promoting this approach for historical and literary studies using correspondence networks (Ahnert and Ahnert 2015; Ryan and Ahnert 2021; Ahnert and Ahnert 2024). Further applications of SNA in the realm of Enlightenment studies have also been highlighted in a recent volume edited by Chloe Edmondson and Dan Edelstein (2019). Additionally, the 2021 Cambridge Element, *The Network Turn*, co-authored by the Ahnerts with Nicole Coleman and Scott Weingart, serves as an excellent primer on this methodology and its potential applications across the humanistic disciplines (Ahnert et al. 2021).

Beyond SNA as a guiding methodology, ModERN also aims to incorporate Bruno Latour's Actor-Network Theory (ANT) as an underlying theoretical framework (Latour 2007). ANT has two main advantages: first, it avoids presuppositions and *a priori* assumptions when considering networks and their behaviour, aligning well with a 'data-driven' research philosophy. Secondly, ANT considers any entity (human or non-human) that exerts an influence on network behaviour as an actor; so, authors and readers, certainly, but also books themselves, letters, pamphlets, extracts, cafés, salons, *ad infinitum*. Another key insight from ANT is that mediators – those actors that facilitate communication and exchange – can often be the most critical nodes in a network, moving focus away from the originators or receptors of the information flow. This sets up the 'mediator' as a key actor in our intertextual networks, one that navigates the waters between originality and imitation outlined above.

To build these networks we obviously need extensive data. Thus, the first years of the project have been focused on acquiring as many 18th-century sources as possible, with the sole caveat that the texts must

[15] Research funded by the European Union (ERC Consolidator Grant 101043369). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

already be digitised in some form or another. We rely on a mixture of corrected or transcribed works, which we call the *Canon* sub-corpus, and uncorrected texts that have been digitised using OCR technologies, which we call the *Archive*. Our main combined database (*Canon* + *Archive*) includes just over 13 000 documents (manly printed books), which we will compare iteratively with other secondary corpora of 18th-century pamphlets, letters, dictionaries and newspapers.[16] The variety of our collections speaks to our goal of considering Darnton's communication network in its totality, recognising that books are just one part of the 18th-century information flow. As we have seen above, dictionaries and private letters both play crucial, albeit distinct, roles in the formation and exchange of intertextual networks, and we expect pamphlets and the press to do the same.

Our overall goal for the project is twofold: first, our database of 18th-century text reuses will act as a supplement to traditional 'close-reading' methods, and, ideally, offer new ways of performing historical and literary analyses with significantly greater speed and breadth. By leveraging the power of computational tools such as Text-PAIR to analyse a massive textual corpus, we can begin to tackle essential questions about the literary history of the 18th century at an unprecedented scale. These tools will thus not only enrich our understanding of the past but also demonstrate the dynamic interplay between texts and their historical contexts, offering a more nuanced and comprehensive picture of French literary history as it was being debated and codified. Secondly, and in a more 'revolutionary' spirit, the ModERN project aims to deploy new 'distant-reading' methodologies in order to challenge received ideas about the Enlightenment and its main actors.[17] A greater focus on intertextual networks will enable us to remap the literary history of the 18th century using SNA and other computational approaches to trace and categorise how texts and ideas move through different contexts. This approach embraces the more radical version of intertextuality espoused by Roland Barthes, one in which every text is a potential (or real) intertext (Barthes 1973):

> Épistémologiquement, le concept d'intertexte est ce qui apporte à la théorie du texte le volume de la socialité: c'est tout le langage antérieur et contemporain qui vient au texte, non selon la voie d'une filiation repérable, d'une imitation volontaire, mais selon celle d'une dissémination – image qui assure au texte le statut non d'une *reproduction*, mais d'une *productivité*.

As Barthes suggests, intertextuality is not merely about the reproduction of some other text, but rather a latent productivity that informs all texts. An intertext or cluster of intertexts thus becomes a productive

[16] These collections come mainly from our institutional partners: the ARTFL Project, University of Chicago; the Voltaire Foundation, University of Oxford; the Newberry Library; Gale Primary Sources; and the BnF DataLab at the Bibliothèque nationale de France. We would like to take the opportunity here to thank them again for their generous data-sharing policies. For a fuller description of the project and its corpora, see Fedchenko, Nicolosi and Roe (2024).

[17] For an excellent overview of 'distant-reading' methods, past and present, see Underwood (2017).

force, often independent of any authorial intent.[18] These reuses traverse texts and contexts, sometimes attached to author-names, but often not, a fact that invites a re-evaluation of authorship in our intertextual model of literary dissemination. For us, an author (named or unnamed) is but one attribute of a text-object; the model's primary focus is on the text itself and its productive circulation.

When we do generate author (or ego) networks, we understand an author's importance not from the perspective of traditional literary history but from their position within the network and their author-function as text reusers. Our model thus defines an author-object as the result of the combination of all the characteristics attributed to it by the relations between text-objects; its 'influence' is measured by way of its position within the intertextual network. Fénelon, for example, while certainly an important 'authority' in terms of text reuse, is more influential in our model as a 'mediator', i.e. a node that bridges several different communities that are normally separate, such as biblical exegetes and the *philosophes*, thereby underscoring the fluidity and interconnectedness of 18th-century literary culture.

This methodology also prompts us to reconsider what constitutes a 'text-object' in our model. Prosaically, we define a text-object as a co-occurrence of at least four similar trigrams within a defined interval that has been identified in at least two different documents. This definition raises several issues, especially since not all reuses are equal. In Figure 2 we see two pairs of identified 'reuses' from our database. In the first, we have the curious example of Marat, in the midst of the revolutionary furore, reusing – whether consciously or not we cannot say – a passage from Fénelon's *Télémaque* in his *Chaînes de l'esclavage*. The passages are nearly identical, though Marat replaces 'lion de Numidie' with 'tigre' and omits three words, 'de faibles brebis', but the meaning of the passage is completely transformed in its revolutionary context. This example is exactly the kind of productive reuse we seek to identify. The second example, however, would seem superficially important, given its length, near-exact replication, and the similar book titles in which it occurs. But, on examination, this match might fall into the category of 'noise': a formula which is so much of a commonplace that it has little or no significant intertextual value. These sorts of decisions – what constitutes a 'reuse' in our database – along with establishing a typology of all reuses, will be one of the main research axes for the project going forward.

These examples bring us back to Barthes' concept of intertextuality. For Barthes, 'tout texte est un intertexte', and, as such, 'intertextualité, condition de tout texte, quel qu'il soit, ne se réduit évidemment pas à un problème de sources ou d'influences; l'intertexte est un champ général de formules anonymes, dont l'origine est rarement repérable, de citations inconscientes ou automatiques, données sans guillemets' (Barthes 1973). The challenge for us thus lies in identifying and classifying these different sorts of intertexts, from direct quotations, indirect borrowings, loose paraphrases and allusions, down to the aforementioned 'anonymous formulas', all the while bearing in mind the various author-functions (anonymity, pseudonymity, collaborative authorship, etc.) attached to them.

[18] In this way, intertextuality as a concept challenges the dominance of authorial intentionality in literary theory in much the same way as Barthes' 'mort de l'auteur' (1968) and Foucault's 'fonction-auteur' (1969).

Fénelon, *Les aventures de Télémaque*, 1699

La mort courait de rang en rang partout sous ses coups [de Mentor]. **Semblable à un lion de Numidie que la cruelle faim dévore, et qui entre dans un troupeau** de faibles brebis : **il déchire, il égorge, il nage dans le sang,** et les bergers, loin de secourir le troupeau, fuient tremblants, pour se dérober à sa fureur.

Marat, *Les chaînes de l'esclavage*, 1792

Ainsi, la vie des citoyens est sans cesse sacrifiée à la prétendue paix de l'état : la mort court partout de rang en rang, sous ses pas ; **semblable à un tigre que la cruelle faim dévore, et qui entre dans un troupeau, il déchire, il égorge, il nage dans le sang.**

Daniel, *Voyage du Monde de Descartes*, 1690

L'étendue de chacun est aussi proportionnée à l'excellence de sa nature : ils ont partagé, comme frères, les quatre qualités : ils en ont chacun deux, dont ils en possèdent une dans le souverain degré. **La terre est froide et sèche ; l'eau est froide et humide ; l'air est chaud et humide ; et le feu est chaud et sec.**

Barthélemy, *Voyage du jeune Anacharsis*, 1788

Aux quatre éléments sont attachées quatre propriétés essentielles : froideur, chaleur, sécheresse et humidité. Les deux premières sont actives, les deux secondes passives ; chaque élément en possède deux : **la terre est froide et sèche ; l'eau, froide et humide ; l'air, chaud et humide ; le feu, sec et chaud.**

Figure 2 Two pairs of reuses from the ModERN database.

6. In search of intertextual hubs

To identify and classify the most highly intertextual texts or 'hubs' in our data, we rely on 'graph profiles' that combine several well-known metrics for network centrality: in- and out-degree centrality (Freeman 1978), betweenness centrality (Barthelemy 2004), and PageRank (Page et al. 1999), among others. While these SNA measures are in need of proper assessment concerning their applicability for networks constructed out of text-reuse data, we are nonetheless encouraged by our preliminary results. As text-reuse networks are not 'natural' networks in the sense that (some) social networks are, using standard SNA measures on them requires rigorous reflection on what is actually being measured and analysed. While these assessments are outside the scope of this paper, our initial deployment of these graph profiles helps us identify and classify a large variety of intertexts, i.e. those that are reused the most (authorities), those that reuse the most (observers), those with high betweenness that bridge communities (mediators) and those whose reuses are the most influential based on their PageRank score (influencers). A work's graph profile can thus give us an indication of how it behaves in the network and, ideally, bring lesser-known works to light.

To give a salient example, a text previously unknown to us that repeatedly emerges in our graph analyses is a certain anonymous volume entitled *Pensées républicaines* (1794), which was part of the *Bibliothèque bleue de Troyes* collection.[19] Initially, we noticed its relatively high ranking in several network measures without fully

[19] *Pensées républicaines pour tous les jours de l'année; à l'usage, sur-tout, des enfans. Par l'auteur du Catéchisme moral et républicain*. On the *Bibliothèque bleue de Troyes* collection, see https://artfl-project.uchicago.edu/bibliotheque-bleue.

Number of identified passages	Author	Titles
11	Jean-Jacques Rousseau	*Discours sur l'économie politique, Émile, La Nouvelle Héloïse, Les Confessions*
10	Joseph de Guignes	*Mémoires concernant l'histoire, les sciences, les arts, les mœurs, les usages, des Chinois*
8	Charles Duclos	*Histoire de Louis XI, Considérations sur les mœurs*
6	Benjamin Franklin	*La Science du Bonhomme Richard, ou moyen facile de payer les impôts*
6	Jean de La Bruyère	*Les Caractères ou les mœurs de ce siècle*
4	Fontette de Sommery	*Doutes sur différentes opinions reçues dans la société*
3	Gabriel de Mably	*Entretiens de Phocion*
3	Joseph-Michel-Antoine Servan	*Discours sur l'administration de la justice criminelle, par M. S. M. Servan*
3	Anthony Ashley Cooper Shaftesbury	*Philosophie morale réduite à ses principes, ou Essai de M. S*** sur le mérite et la vertu*.
3	Josué Le Marchant	*Recueil de maximes, de pensées et de réflexions. Par Josué Le Marchant, écuyer*
2	William Shakespeare, Voltaire, François Andrieux, Charles-François-Nicolas Le Maître de Claville, Edward Young, Paul Henri Dietrich baron d'Holbach, Alexander Pope, Martin Sherlock…	

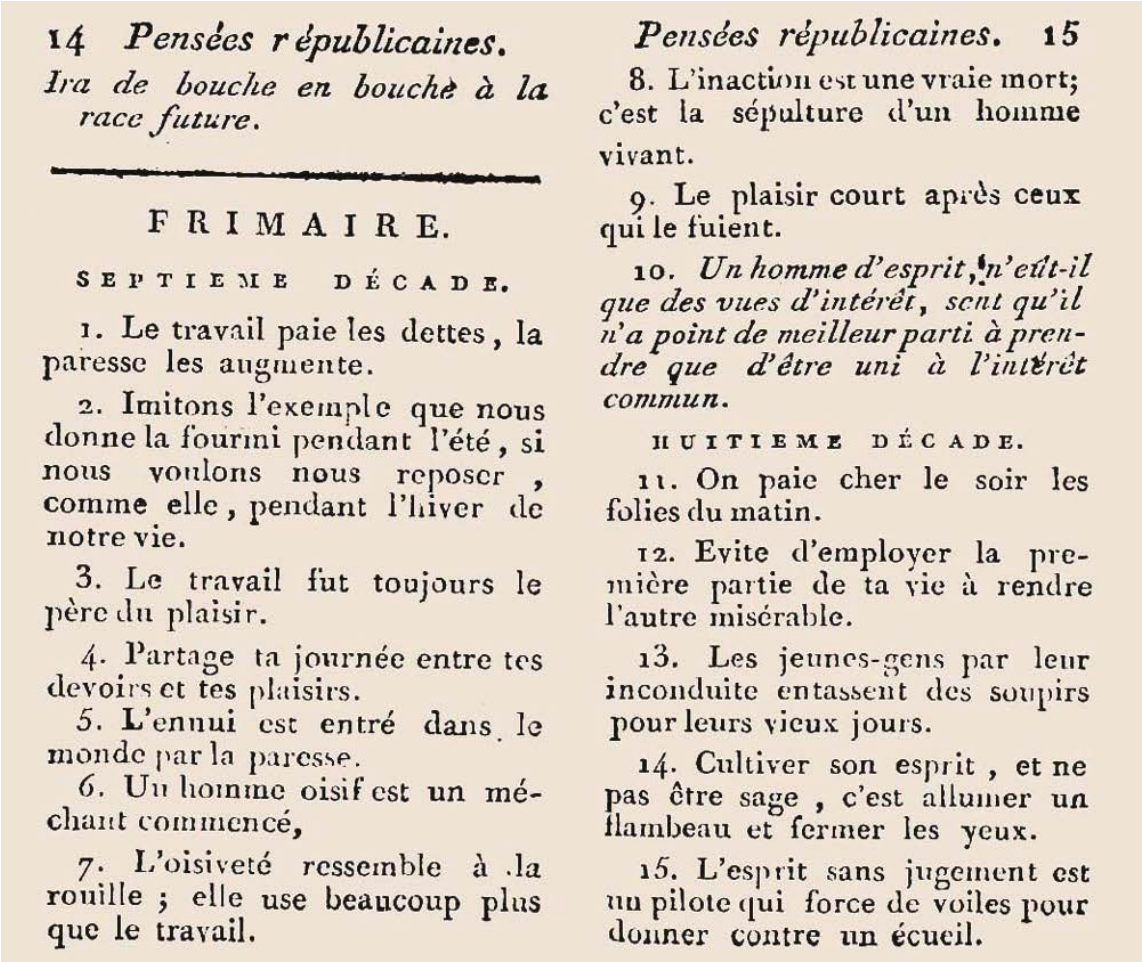
Table 1 Number of identified passages by authors and titles in the *Pensées républicaines*

understanding its significance.[20] Upon closer inspection, it became clear why it scored so highly: it is essentially a text-reuse machine. Written at the beginning of the Revolution, *Pensées républicaines* is a collection of maxims and aphorisms intended for children, with one for each day of the Republican calendar. None of the maxims are attributed in the original text, and our general alignment only found around 100 of the 360 odd entries in our main corpus. Nonetheless, the sources we did identify attest to its syncretic nature, drawing on a wide range of sources, including Rousseau, La Bruyère, Benjamin Franklin and a history of China, among others (see Table 1).

If we examine just one 'decade', the seventh of Frimaire, we see a fascinating intertextual mosaic emerge (see Figure 3).

Benjamin Franklin is reused twice (nos 1 and 7), along with Voltaire (no. 3), La Bruyère (no. 5), Seneca (no. 8), Duclos' *Histoire de Louis XI* (no. 10), and Servan's *Discours sur l'administration de la justice criminelle* (no. 6), with each maxim loosely related thematically (work/idleness/pleasure). What is especially notable

[20] The *Pensées républicaines* ranked 134th for in-degree, 18th for out-degree, 11th in betweenness and 134th for PageRank. The disproportion between out- and in-degree tells us that this work reuses a great deal but is very rarely reused; such a low PageRank also suggests that this text is probably late (it was published in 1794). A betweenness this high suggests that this work cites hubs, which makes it 'central' in the network. At the same time, it quotes 'different' authors and not necessarily those only from the same group or cluster.



14 *Pensées républicaines.*

Ira de bouche en bouche à la race future.

FRIMAIRE.

SEPTIEME DÉCADE.

1. Le travail paie les dettes, la paresse les augmente.

2. Imitons l'exemple que nous donne la fourmi pendant l'été, si nous voulons nous reposer, comme elle, pendant l'hiver de notre vie.

3. Le travail fut toujours le père du plaisir.

4. Partage ta journée entre tes devoirs et tes plaisirs.

5. L'ennui est entré dans le monde par la paresse.

6. Un homme oisif est un méchant commencé,

7. L'oisiveté ressemble à la rouille ; elle use beaucoup plus que le travail.

Pensées républicaines. 15

8. L'inaction est une vraie mort; c'est la sépulture d'un homme vivant.

9. Le plaisir court après ceux qui le fuient.

10. *Un homme d'esprit, n'eût-il que des vues d'intérêt, sent qu'il n'a point de meilleur parti à prendre que d'être uni à l'intérêt commun.*

HUITIEME DÉCADE.

11. On paie cher le soir les folies du matin.

12. Evite d'employer la première partie de ta vie à rendre l'autre misérable.

13. Les jeunes-gens par leur inconduite entassent des soupirs pour leurs vieux jours.

14. Cultiver son esprit, et ne pas être sage, c'est allumer un flambeau et fermer les yeux.

15. L'esprit sans jugement est un pilote qui force de voiles pour donner contre un écueil.

Figure 3 *Pensées républicaines* 1794, pp.14–15 ('Frimaire. Septième décade').

in the text is the total absence of all authorial authority; the maxims stand on their own, disconnected from any origin or source.

Moving from quantitative to qualitative analysis, or from distant- to close-reading modes, there are numerous deformations to identified maxims that are altogether remarkable. For instance, the passage drawn from Duclos' *History of Louis XI* is actually altered, changing its meaning entirely (differences in bold):

Duclos, *Histoire de Louis XI*, Préface:

> Ce que j'avance au sujet d'une nation peut s'appliquer aux particuliers. Les hommes privés de lumières sont toujours dans l'occasion du crime; au lieu *qu'un homme d'esprit, n'eût-il que des vues d'intérêt, sent qu'il n'a point de meilleur parti à prendre que d'être* **honnête homme**. On est bien près de suivre la vertu, quand on est obligé de rougir du vice.[21]

[21] Duclos (1745), vol.1, pp.xxvi–xxix. See https://artflsrv04.uchicago.edu/philologic4.7/frantext0822/navigate/831/1.

Pensées républicaines, Frimaire, septième décade, pensée 10:

> *Un homme d'esprit, n'eût-il que des vues d'intérêt, sent qu'il n'a point de meilleur parti à prendre que d'être* **uni à l'intérêt commun**.[22]

One need not be a historian of the Revolution to understand the importance of placing the common good above the individual interest of the 'honnête homme'.

The text becomes even more intriguing when considering examples from elsewhere in the calendar. In the sixth decade of Brumaire, we find the following *pensée* (*Pensées républicaines* 1794, pp.13–14):

> 30. Mourir pour son pays n'est pas un triste sort,
> C'est s'immortaliser par une belle mort.
> De BARRA, jeune encore, l'étonnante aventure
> Ira de bouche en bouche à la race future.[23]

This maxim is actually drawn from two different sources, and then slightly altered (the relevant text is shown in bold in each case):

Corneille, *Le Cid*, IV.v:

> Qu'il meure pour mon père, et non pour la patrie,
> Que son nom soit taché, sa mémoire flétrie;
> **Mourir pour le pays n'est pas un triste sort;**
> **C'est s'immortaliser par une belle mort.**
> J'aime donc sa victoire, et je le puis sans crime,
> Elle assure l'État, et me rend ma victime,[24]

Voltaire, *La Henriade*, Chant 2:

> Quelques-uns, il est vrai, dans la foule des morts,
> Du fer des assassins trompèrent les efforts.
> **De Caumont, un enfant, l'étonnante aventure**
> **Ira de bouche en bouche à la race future.**
> Son vieux père, accablé sous le fardeau des ans,
> Se livrait au sommeil entre ses deux enfants;
> Un lit seul enfermait et les fils et le père.[25]

What are we to make of this amalgamation? Does the *pensée* belong to Voltaire or Corneille? Or neither? This unexpected juxtaposition – not to mention the insertion of Barras to replace Caumont – again

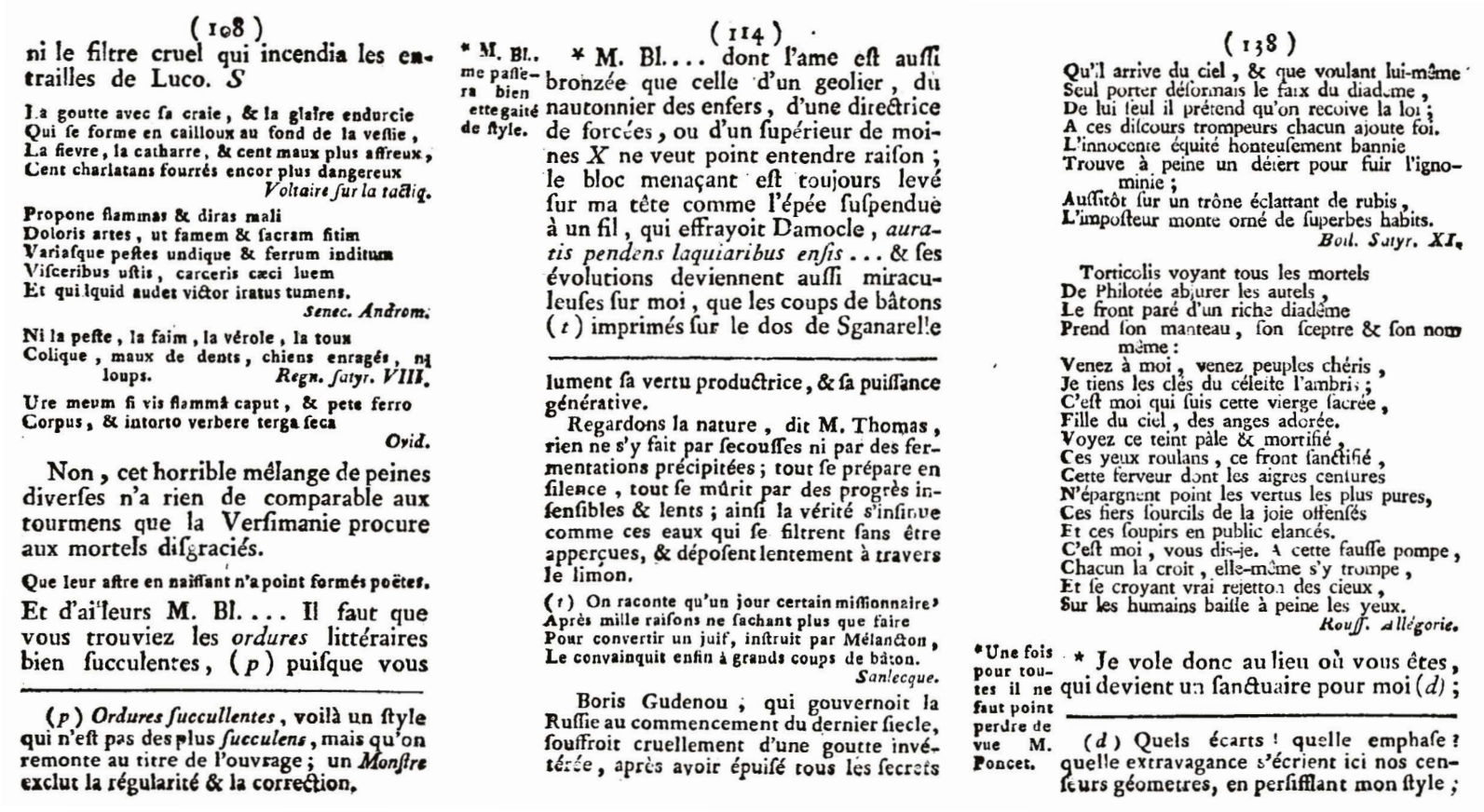
[22] *Pensées républicaines* (1794), p.15. See https://artflsrv04.uchicago.edu/philologic4.7/frantext0822/navigate/1379/4/2.
[23] See https://artflsrv04.uchicago.edu/philologic4.7/frantext0822/navigate/1379/3/4.
[24] Corneille (2017), p.56. See https://theatre-classique.fr/pages/programmes/edition.php?t=../documents/CORNEILLEP_CID.xml.
[25] Voltaire (1970), p.406. See https://artflsrv03.uchicago.edu/philologic4/toutvoltaire/navigate/44/1/19/.

raises questions about the validity of our understanding of intentionality when viewed through the prism of intertextuality. These subtle, yet significant changes, also highlight the limitations of computational analysis; while the computer can identify these connections, understanding their significance requires deeper interpretive work. The hybridity of the *Pensées républicaines* is a testament to the complexity of intertextuality in the 18th century, where texts constantly intersect and transform, challenging our modern notions of authorship and originality along the way.

Finally, another text that has surfaced repeatedly in our analyses is a fascinating work by a certain Jean-Marie Chassaignon (1779), whose title is 'tout un programme': *Cataractes de l'imagination, déluge de la scribomanie, vomissement littéraire, hémorragie encyclopédique, monstre des monstres, par Epiménide l'Inspiré… dans l'antre de Trophonius, au pays des visions*. Chassaignon was a poet from Lyon who was highly active in the illuminist and *antiphilosophe* circles assembled around Louis-Claude de Saint-Martin (Poirson 2017). Chassaignon's 'frenetic' and 'excessive' writing style was indicative of a certain stylistic 'énergie' that pervades the late Enlightenment according to Michel Delon (1988). His four-volume *magnum opus*, the *Cataractes*, published in 1779 and never re-edited, is a strange and fascinating work; taking Diderot's notion of eclecticism to a rare extreme, it mixes genres from the poetic to the philosophical, mystical and theosophical, and enacts the productive poetics of text reuse we have been discussing in this essay.

Chassaignon begins the book with a preface 'qui n'en est pas une', in which he immediately denigrates the very idea of any prefatory introduction: 'On ne lit plus les préfaces: elles ne sont pour l'ordinaire qu'un surcroît d'ennui pour le lecteur à qui elles n'apprennent rien' (1779, vol.1, p.5). And yet, he then proceeds over the next hundred pages or so to introduce his work in painstaking detail: 'Mais, moi qui crie contre les préfaces, je m'en permets cependant une, & me laisse surprendre en flagrant délit; oui, parce que mon opuscule ne pourroit s'en passer. J'écris dans un genre inconnu à ce siécle & il faut que je donne la clé de l'énigme, si je veux être compris' (pp.6–7).

Though it was presented as being in a wholly new genre, Chassaignon's work was not without predecessors: 'Plutarque, [La Moth] Lévayer, Bayle & Montagne qui ont entassé pêle & mêle sur le papier tout ce qui leur venoit dans la pensée & ont submergé leurs écrits d'annotation, m'ont communiqué le goût de leur genre. Le dernier sur-tout m'a enhardi à suivre ses traces; on ne peut résister à sa maniere naïve & entraînante' (p.18). As we have seen, Montaigne is key figure in the history of text reuse, though Chassaignon admires him most, it seems, for writing chapters that diverge entirely from their titles, following his disparate thoughts wherever they may lead, and including as many citations as needed to demonstrate his way of thinking. In Pierre Bayle – avowed predecessor to the *Encyclopédie* and the 'arsenal of Enlightenment' according to Cassirer (1951, p.167) – Chassaignon finds structural inspiration for his work, and particularly Bayle's extensive system of notation, weaving text, remarks, footnotes, marginal notes, references, etc. into the fabric of his *Dictionnaire historique et critique* (Van der Lugt 2016). Like Bayle, Chassaignon's notes can range from a single phrase to several pages of self-commentary, often running longer than the main text, which is punctuated by note calls, both foot and marginal, and discursive asides marked with an asterisk – perhaps recalling Diderot's authorial mark in the *Encyclopédie* (see Figure 4).



(108)

ni le filtre cruel qui incendia les entrailles de Luco. *S*

La goutte avec ſa craie, & la glaire endurcie
Qui ſe forme en cailloux au fond de la veſſie,
La fievre, la catharre, & cent maux plus affreux,
Cent charlatans fourrés encor plus dangereux
Voltaire ſur la tactiq.

Propone flammas & diras mali
Doloris artes, ut famem & ſacram ſitim
Variaſque peſtes undique & ferrum inditum
Viſceribus uſtis, carceris cæci luem
Et qui liquid audet victor iratus tumens.
Senec. Androm.

Ni la peſte, la faim, la vérole, la toux
Colique, maux de dents, chiens enragés, ni loups.
Regn. ſatyr. VIII.

Ure meum ſi vis flammâ caput, & pete ferro
Corpus, & intorto verbere terga ſeca
Ovid.

Non, cet horrible mélange de peines diverſes n'a rien de comparable aux tourmens que la Verſimanie procure aux mortels diſgraciés.

Que leur aſtre en naiſſant n'a point formés poëtes.

Et d'ailleurs M. Bl.... Il faut que vous trouviez les *ordures* littéraires bien ſucculentes, (*p*) puiſque vous

(*p*) *Ordures ſuccullentes*, voilà un ſtyle qui n'eſt pas des plus *ſucculens*, mais qu'on remonte au titre de l'ouvrage; un *Monſtre* exclut la régularité & la correction.

(114)

* M. Bl.. me paſſera bien cette gaité de ſtyle.

* M. Bl.... dont l'ame eſt auſſi bronzée que celle d'un geolier, du nautonnier des enfers, d'une directrice de forcées, ou d'un ſupérieur de moines *X* ne veut point entendre raiſon; le bloc menaçant eſt toujours levé ſur ma tête comme l'épée ſuſpendue à un fil, qui effrayoit Damocle, *auratis pendens laquiaribus enſis*... & ſes évolutions deviennent auſſi miraculeuſes ſur moi, que les coups de bâtons (*t*) imprimés ſur le dos de Sganarelle

lument ſa vertu productrice, & ſa puiſſance générative.

Regardons la nature, dit M. Thomas, rien ne s'y fait par ſecouſſes ni par des fermentations précipitées; tout ſe prépare en ſilence, tout ſe mûrit par des progrès inſenſibles & lents; ainſi la vérité s'inſinue comme ces eaux qui ſe filtrent ſans être apperçues, & dépoſent lentement à travers le limon.

(*t*) On raconte qu'un jour certain miſſionnaire
Après mille raiſons ne ſachant plus que faire
Pour convertir un juif, inſtruit par Mélancton,
Le convainquit enfin à grands coups de bâton.
Sanlecque.

Boris Gudenou, qui gouvernoit la Ruſſie au commencement du dernier ſiecle, ſouffroit cruellement d'une goutte invétérée, après avoir épuiſé tous les ſecrets

(138)

Qu'il arrive du ciel, & que voulant lui-même
Seul porter déſormais le faix du diadême,
De lui ſeul il prétend qu'on reçoive la loi;
A ces diſcours trompeurs chacun ajoute foi.
L'innocente équité honteuſement bannie
Trouve à peine un déſert pour fuir l'ignominie;
Auſſitôt ſur un trône éclattant de rubis,
L'impoſteur monte orné de ſuperbes habits.
Boil. Satyr. XI.

Torticolis voyant tous les mortels
De Philotée abjurer les autels,
Le front paré d'un riche diadême
Prend ſon manteau, ſon ſceptre & ſon nom même:
Venez à moi, venez peuples chéris,
Je tiens les clés du céleſte l'ambri;
C'eſt moi qui ſuis cette vierge ſacrée,
Fille du ciel, des anges adorée.
Voyez ce teint pâle & mortifié,
Ces yeux roulans, ce front ſanctifié,
Cette ferveur dont les aigres centures
N'épargnent point les vertus les plus pures,
Ces fiers ſourcils de la joie offenſés
Et ces ſoupirs en public elancés.
C'eſt moi, vous dis-je. A cette fauſſe pompe,
Chacun la croit, elle-même s'y trompe,
Et ſe croyant vrai rejetton des cieux,
Sur les humains baiſſe à peine les yeux.
Rouſſ. Allégorie.

*Une fois pour toutes il ne faut point perdre de vue M. Poncet.

* Je vole donc au lieu où vous êtes, qui devient un ſanctuaire pour moi (*d*);

(*d*) Quels écarts! quelle emphaſe? quelle extravagance s'écrient ici nos cenſeurs géometres, en perſifflant mon ſtyle;

Figure 4 Three pages from Chassaignon's *Cataractes de l'imagination* (1779), vol.1.

Further on in his preface, Chassaignon outlines his own theory, philosophy and practice of text reuse, all in one go (1779, vol.1, pp.34–36):

> Je noue de la prose aux vers [...]; j'appelle toutes les annales à mon secours: je mets à contribution tous les auteurs, je les passe tous en revue; nouveau dictateur de la république littéraire, je gourmande les factieux, qui font siffler au sein de leur patrie, les serpens de la discorde.
>
> *Iam caelum terramque meo sine numine, venti*
> *Miscere, & tantas audetis tollere moles?*
> *Quas ego...*
>
> Je m'égare dans des digressions, je quitte vingt fois mon objet, & j'y reviens après de longs écarts. […] Mon opuscule ainsi étoffé de traits historiques, gonflé d'incartades de cerveau, chamarré de citations de toutes les especes, devient une *mosaïque littéraire*, un nouveau *moyen de parvenir* que pourront étudier les jeunes amateurs, & que les érudits parcourront avec plaisir.
>
> *Quae pueri discant, & ament meminisse periti.*

The two Latin quotations, included without attribution, deftly demonstrate Chassaignon's poetics of reuse. The first comes from Virgil's *Aeneid* (Book 1, lines 133–35), and serves as an illustration of Chassaignon's eclectic project of mixing genres and elements: 'Do you now dare, winds, without command of mine, to mingle earth and sky, and raise confusion thus?'[26] This passage would certainly have been recognisable by most 18th-century readers and perhaps needed no attribution. The second citation, however, is a bit murkier. It would seem to be a deformation of a verse written by Charles-Jean-François 'le Président' Hénault and published as an epigraph to the first edition of his *Nouvel abrégé chronologique* (1744): 'Indocti discant et ament meminisse periti.' It was initially thought to be from Horace, a misattribution that persisted well into the 19th century (Longpérier 1875);[27] although Hénault would eventually disclose the origin of the verse in the third edition of his work: namely his own translation of lines 741–42 of Pope's *Essay on Criticism* (1713):

> Content if hence the unlearned their wants may view,
> The learned reflect on what before they knew

Whatever its paternity, Chassaignon's replacement of 'unlearned' with 'children' makes the verse his own, all the while keeping its intertextual link with Hénault (not to mention Pope and Horace) for those in the know.

Chassaignon's eclectic virtuosity in reusing texts thus forms an integral part of his poetic enterprise (Moore 2010). It is through the interweaving of old and new, ancients and moderns, known and unknown, that new forms and meanings can emerge (Chassaignon 1779, vol.1, pp.33–34):

> Échauffé, encouragé, transporté, je m'abandonne au délire de mes idées […]. Le génie né altier & indépendant, qui ne connoît, ni sent, & n'apperçoit que lui, s'indignant des entraves qui gênent sa

[26] Translation by H. Rushton Fairclough, Virgil (1999).

[27] See also https://www.france-pittoresque.com/spip.php?article13736.

> fougue & repoussent son égoïsme; loin de se rétrecir à aucun genre, & d'avouer son impuissance pour un autre, doit débuter par une espece d'*Encyclopédie*, étonner par l'*universalité* de sa production & la multiplicité de ses talens, jetter dans un même ouvrage le sel de l'épigramme, les roses de l'adulation, les orties de la satyre…

It is an almost alchemical approach to literature, hoping that the right mixture of elements might transform literary lead into gold. The productive intertextuality mentioned early is thus very much operative in Chassaignon's work, which is another example of a text-reuse machine, weaving together a vast array of sources into a cohesive, albeit chaotic, whole. Like the *Encyclopédie* before him, which generated new knowledge through the critical reuse of sources, Chassaignon's project is to generate enlightenment itself, the deep illumination that can only come from literary creation. It is not an easy road to travel, and one can easily get lost, but given the right tools and enough time, anything is possible: 'un bisarre débordement de pensées errantes, & de passages dérobés n'exige pas un effort bien étrange de l'*imaginative*; avec des dictionnaires, des recueils, de la patience, des copistes & du temps, le scribe le plus borné est en état de peupler l'univers littéraire de monstres encyclopédiques' (p.36). This conception of creation through reuse should resonate with our times, where the right (digital) tools, such as ChatGPT and other LLMs, can also create poetic 'monsters' of text reuse.

Diderot, of course, already foresaw these sorts of monstrosities in the making when he took stock of the *encyclopédistes*' 'universalising' enterprise in 1755. Far from the perfect system of human understanding proposed by the mathematician D'Alembert, what they had created in reality was uneven, unwieldy and ultimately uncontrollable (Diderot and D'Alembert 2022, vol.5, pp.641–42):

> La preuve en subsiste en cent endroits de cet Ouvrage. Ici nous sommes boursouflés & d'un volume exorbitant; là maigres, petits, mesquins, secs & décharnés. Dans un endroit, nous ressemblons à des squeletes; dans un autre, nous avons un air hydropique; nous sommes alternativement nains & géants, colosses & pigmées; droits, bienfaits & proportionnés; bossus, boiteux & contrefaits. Ajoûtez à toutes ces bisarreries celle d'un discours tantôt abstrait, obscur ou recherché, plus souvent négligé, traînant & lâche; & vous comparerez l'ouvrage entier au monstre de l'art poétique, ou même à quelque chose de plus hideux.

Despite these reservations, the *Encyclopédie* became an intertextual dynamo, driven forward by the potentiality of critical text reuse and the creative productivity of its inherent intertextuality. In this way Diderot's dictionary can rightly be seen as predecessor not only to modern encyclopaedic projects like Wikipedia and the internet (Bianco 2002; Fernandez 2019), but also, and perhaps more significantly, it shares a direct lineage with the imaginary information machines of Charles Babbage, Ada Lovelace, Alan Turing and Vannevar Bush. The poetics of reuse outlined in the *Encyclopédie* and put into practice throughout the 18th century is thus a resolutely modern, if not contemporary, phenomenon. In our age of AIs and digital

information machines, recognising both the historicity and contemporaneity of these intertextual practices is crucial for our understanding of the legacy of the Enlightenment as well as our place in it. In some ways, one could go so far as to argue that works like the *Encyclopédie*, Voltaire's correspondence, the *Pensées républicaines* and Chassaignon's *Cataractes* – literary monsters one and all – are only fully actualised in and by the intertextual affordances of the digital.

Returning to our illuminist poet, finding more information about Chassaignon has proven challenging. A 19th-century biographical entry in Quérard's *La France littéraire* simply states his birth and death dates and lists his published works, including the *Cataractes*. Curiously, the entry ends with an editorial note that claims that Chassaignon 'avait laissé beaucoup de manuscrits que son frère, épicier à Lyon, a employés pour envelopper les drogues de sa boutique' (Quérard 1828, p.145). Less conventional, and certainly far from computationally tractable, this tangible example of text reuse serves as a final reminder of the enduring relevance and diversity of text reuse as a cultural practice.

7. Conclusions and perspectives

Our goal with this article has been to present a thorough, but by no means exhaustive, investigation into the practice of text reuse in the 18th century, demonstrating its significance not only as a literary phenomenon but also as a cultural mechanism that shaped intellectual exchange. By examining both historical and computational perspectives, we highlight how the boundaries between plagiarism, citation and innovation were fluid during this period. Our analysis reveals that text reuse was not merely a form of borrowing, or worse, plagiarism, but rather a means of contributing to the dialogic process of early-modern knowledge production, reflecting the broader socio-political and intellectual currents of the time.

As we have seen, automatic text-reuse detection tools such as Text-PAIR, combined with large-scale digital collections, can offer new insights into how intertextuality functioned historically, revealing hidden connections between texts and authors that traditional close reading might overlook. These methods enable scholars to trace the intricate ways in which ideas circulated and were reshaped within 'intertextual hubs' such as the *Encyclopédie* and Voltaire's prodigious correspondence. In this context, the blending of voices, particularly through the identification of unacknowledged contributions of figures such as Émilie Du Châtelet, underscores the complexity of authorship practices and the aesthetics of eclecticism prevalent in the 18th century.

Ultimately, this article not only provides a deeper understanding of historical text reuse but also demonstrates the potential of the digital humanities to reconfigure our traditional approaches to literary and cultural history. Through ongoing projects that aim to map these intertextual networks, the study offers a nuanced view of the Enlightenment as a period of rich intellectual exchange, driven by both overt and covert forms of collaboration. This approach reaffirms the importance of considering both humanistic and technological perspectives when examining the literary past, thus opening up new avenues for future research.

We expect text-reuse approaches to continue to evolve rapidly as machine learning and AI models become more prevalent in the investigation and exploitation of our cultural collections. LLMs will certainly have their place in this evolution and might in fact guide both technological and theoretical reflection about intertextuality moving forward. As suggested above, the 'poetic monsters' of generative AI systems such as ChatGPT function primarily as massive intertextual machines; those that combine and remix past texts probabilistically in order to generate 'new' textual artefacts delivered through dialogical interfaces. From a theoretical perspective we are thus closer today than ever to Roland Barthes' adage that 'tout texte est un intertexte', and these same LLMs will no doubt prove helpful not only from a generative perspective, but also analytically, allowing us to detect broader, more nuanced forms of text reuse that go beyond the current sequence-based models (Kanerva et al. 2024).

References

Ahnert R. and Ahnert S. 2015. 'Protestant letter networks in the reign of Mary I: a quantitative approach'. In: *ELH* 82:1, 1–33. https://www.jstor.org/stable/24477807.

Ahnert R. and Ahnert S. 2024. *Tudor Networks of Power*. Oxford: Oxford University Press.

Ahnert R., Ahnert S., Coleman C. and Weingart S. 2021. *The Network Turn: Changing Perspectives in the Humanities*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108866804.

Albertan C. 1992. 'Les journalistes de Trévoux lecteurs de l'*Encyclopédie*'. In: *Recherches sur Diderot et sur l'Encyclopédie* 13, 107–16. https://doi.org/10.3406/rde.1992.1188.

Allen T., Cooney C., Douard S., Horton R., Morrissey R., Olsen M., Roe G. and Voyer R. 2010. 'Plundering philosophers: identifying sources of the *Encyclopédie*'. In: *Journal of the Association for History and Computing* 13:1. http://hdl.handle.net/2027/spo.3310410.0013.107.

Altschul S., Gish W., Miller W., Myers E. and Lipman D. 1990. 'Basic local alignment search tool'. In: *Journal of Molecular Biology* 215:3, 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Antoine-Mahut D. 2024. 'Eclecticism and its discontents'. In: Sinclair M. and Whistler D. (eds) *The Oxford Handbook of Modern French Philosophy*. Oxford: Oxford University Press, 43–54. https://doi.org/10.1093/oxfordhb/9780198841869.013.2.

Barthélemy J.-J. 1788. *Voyage du jeune Anacharsis*. Paris: de Bure. https://artflsrv04.uchicago.edu/philologic4.7/frantext0822/navigate/1308/table-of-contents.

Barthelemy M. 2004. 'Betweenness centrality in large complex networks'. In: *The European Physical Journal* 38:2, 163–68. https://doi.org/10.1140/epjb/e2004-00111-4.

Barthes R. 1968. 'La mort de l'auteur'. In: *Le Bruissement de la langue. Essais critiques IV*. Paris: Seuil, 63–70.

Barthes R. 1973. 'Texte (théorie du)'. In: *Encyclopædia Universalis* 15, 1013–20.

Bayle P. 1697. *Dictionnaire historique et critique*. Amsterdam: Reinier Leers.

Bergroth L., Hakonen H. and Raita T. 2000. 'A survey of longest common subsequence algorithms'. In: *Proceedings Seventh International Symposium on String Processing and Information Retrieval*. A Coruña: IEEE Computer Society, 39–48. https://doi.org/10.1109/spire.2000.878178.

Bianco J.-F. 2002. 'Diderot avait-il inventé le Web?' In: *Recherches sur Diderot et sur l'Encyclopédie* 31–32, 17–25. https://doi.org/10.4000/rde.13.

Blischak J., Davenport E. and Wilson G. 2016. 'A quick introduction to version control with Git and GitHub'. In: *PLoS Computational Biology* 12:1. https://doi.org/10.1371/journal.pcbi.1004668.

Bloom H. 1973. *The Anxiety of Influence: A Theory of Poetry*. New York: Oxford University Press.

Büchler M., Burns P., Müller M., Franzini E. and Franzini G. 2014. 'Towards a historical text re-use detection'. In: Biemann C. and Mehler A. (eds) *Text Mining. Theory and Applications of Natural Language Processing*. Cham: Springer, 221–38. https://doi.org/10.1007/978-3-319-12655-5_11.

Burrows S. and Roe G. (eds) 2020. *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies.* Oxford University Studies in the Enlightenment. Liverpool: Liverpool University Press.

Cassirer E. 1951. *The Philosophy of the Enlightenment*. Koelln C. A. and Pettegrove J. P. (trans.). Princeton NJ: Princeton University Press.

Chassaignon J.-M. 1779. *Cataractes de l'imagination, déluge de la scribomanie, vomissement littéraire, hémorragie encyclopédique, monstre des monstres, par Epiménide l'Inspiré*. 4 vol. n.p. [Dans l'antre de Trophonius, au pays des visions]: n.pub.

Clough P., Gaizauskas R., Piao S. and Wilks Y. 2002. 'METER: MEasuring TExt Reuse'. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Philadelphia PA: Association for Computational Linguistics, 152–59. https://doi.org/10.3115/1073083.1073110.

Coffee N., Koenig J.-P., Poornima S., Forstall C., Ossewaarde R. and Jacobson S. 2013. 'The Tesserae Project: intertextual analysis of Latin poetry'. In: *Literary and Linguistic Computing* 28:2, 221–28. https://doi.org/10.1093/llc/fqs033.

Compagnon A. 1979. *La Seconde Main ou le travail de la citation*. Paris: Seuil.

Compagnon A. 1980. *Nous, Michel de Montaigne*. Paris: Seuil.

Corneille P. 2017. *Le Cid*. Fièvre E. and Fièvre P. (eds). In: *Théâtre Classique*. https://theatre-classique.fr/pages/pdf/CORNEILLEP_CID.pdf.

Cronk N. 2020. 'Electronic Enlightenment: recreating the Republic of Letters'. In: Burrows S. and Roe G. (eds) *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies.* Oxford University Studies in the Enlightenment. Liverpool: Liverpool University Press, 55–72.

Cronk N. and Roe G. 2020. *Voltaire's Correspondence: Digital Readings*. Cambridge: Cambridge University Press.

Daniel G. 1690. *Voiage du monde de Descartes*. Paris: Veuve S. Bernard. https://artflsrv04.uchicago.edu/philologic4.7/frantext0822/navigate/497/table-of-contents.

Darnton R. 1996. *The Forbidden Best-Sellers of Pre-Revolutionary France*. New York: Norton.

Delon M. 1988. *L'Idée d'énergie au tournant des Lumières (1770–1820)*. Paris: Presses universitaires de France.

Dictionnaire de l'Académie française. 1762. 2 vol. 4th edn. Paris: Brunet. https://gallica.bnf.fr/ark:/12148/bpt6k12803850.

Dictionnaire de l'Académie française. 1835. 2 vol. 6th edn. Paris: Firmin Didot frères. https://gallica.bnf.fr/ark:/12148/bpt6k50407h.

Dictionnaire universel françois et latin. 1743. 6 vol. Paris: Delaulne. https://gallica.bnf.fr/ark:/12148/bpt6k15068037.

Diderot D. 1950. 'Lettre à Sophie Volland, le 10 mai 1759'. In: Babelon A. (ed.) *Lettres à Sophie Volland*. Paris: Gallimard, 27–29.

Diderot D. and D'Alembert J. (eds) 2022. *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers.* Morrissey R. and Roe G. (eds). 28 vol. Chicago IL: ARTFL Encyclopédie Project. https://encyclopedie.uchicago.edu.

Dobson J. 2019. *Critical Digital Humanities: The Search for a Methodology*. Urbana IL: University of Illinois Press.

Dobson J. 2023. 'On reading and interpreting black box deep neural networks'. In: *International Journal of Digital Humanities* 5, 1–19. https://doi.org/10.1007/s42803-023-00075-w.

Duclos C. 1745. *Histoire de Louis XI*. 3 vol. Paris: Guérin & Prault. https://artflsrv04.uchicago.edu/philologic4.7/frantext0822/navigate/831/table-of-contents.

Edelstein D., Morrissey R. and Roe G. 2013. 'To quote or not to quote: citation strategies in the *Encyclopédie*'. In: *Journal of the History of Ideas* 74:2, 213–36. http://www.jstor.org/stable/43291299.

Edmondson C. and Edelstein D. (eds) 2019. *Networks of Enlightenment: Digital Approaches to the Republic of Letters*. Oxford University Studies in the Enlightenment. Liverpool: Liverpool University Press.

Fedchenko V., Nicolosi D. M. and Roe G. 2024. 'À la recherche des réseaux intertextuels: défis de la recherche littéraire à grande échelle'. In: *Humanités numériques* 9. https://doi.org/10.4000/11wmw.

Fénelon F. de Salignac de La Mothe- 1920. *Les Aventures de Télémaque*. Paris: Hachette. https://artflsrv04.uchicago.edu/philologic4.7/frantext0822/navigate/567/table-of-contents.

Fernandez M. 2019. 'Wikipédia, le rêve de Diderot?' In: *Philonomist* 30 April. https://www.philonomist.com/fr/article/wikipedia-le-reve-de-diderot.

Foucault M. 1969. 'Qu'est-ce qu'un auteur?'. In: *Bulletin de la Société Française de Philosophie* 63:3, 73–104.

Franzini G., Passarotti M., Moritz M. and Büchler M. 2019. 'Using and evaluating TRACER for an *Index fontium computatus* of the *Summa contra Gentiles* of Thomas Aquinas'. Fifth Italian Conference on Computational Linguistics (CLiC-it 2018). Turin: Zenodo. https://doi.org/10.5281/zenodo.3362130.

Freeman L. 1978. 'Centrality in social networks: conceptual clarification'. In: *Social Networks* 1, 215–39.

Genette G. 1982. *Palimpsestes. La littérature au second degré*. Paris: Seuil.

Gevrey F. 1995. 'Présentation'. In: *La Réecriture au XVII^e siècle. XVII^e Siècle* 186, 5–7. https://gallica.bnf.fr/ark:/12148/bpt6k9736123j/f1.item.

Gladstone C. and Cooney C. 2020. 'Opening new paths for scholarship: algorithms to track text reuse in Eighteenth-Century Collections Online'. In: Burrows S. and Roe G. (eds) *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*. Oxford University Studies in the Enlightenment. Liverpool: Liverpool University Press, 353–71.

Hénault C.-J.-F. 1744. *Nouvel Abrégé chronologique de l'histoire de France jusqu'à la mort de Louis XIV*. Paris: Prault.

Hunt J. and McIlroy D. 1976. 'An algorithm for differential file comparison'. In: *Computing Science Technical Report* 41. Murray Hill NJ: Bell Laboratories, 1–9.

Kanerva J., Kitti H., Chang L., Vahtola T., Creutz M. and Ginter F. 2024. 'Semantic search as extractive paraphrase span detection'. In: *Language Resources & Evaluation*. https://doi.org/10.1007/s10579-023-09715-7.

Kristeva J. 1969. *Sēmeiōtikē. Recherches pour une sémanalyse*. Paris: Seuil.

Latour B. 2007. *Reassembling the Social. An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Latournerie A. 2004. 'Droits d'auteur, droits du public: une approche historique'. In: *L'Économie politique* 22:2, 21–33. https://doi.org/10.3917/leco.022.0021.

Li Y. and Mullen L. 2020. *textreuse: Detect Text Reuse and Document Similarity*. https://docs.ropensci.org/textreuse.

Longpérier A. 1875. 'L'origine du vers *Indocti discant et ament meminisse periti* attribuée au président Hénault'. In: *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 19:3, 203–204.

Mallon T. 1989. *Stolen Words: Forays into the Origins and Ravages of Plagiarism*. New York: Ticknor & Fields.

Marat J.-P. 1972. *Les Chaînes de l'esclavage*. Paris: Union générale d'Éditions. https://artflsrv04.uchicago.edu/philologic4.7/frantext0822/navigate/1351/table-of-contents.

Marmontel J.-F. 1777. 'Plagiat'. In: Robinet J.-B. (ed.) *Supplément à l'Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers*. 5 vol. Paris: Panckoucke, vol.4, 388–90. https://encyclopedie.uchicago.edu/node/137.

Melançon B. 2004. 'Sommes-nous les premiers lecteurs de l'*Encyclopédie*?' In: Salaün J.-M. and Vandendorpe C. (eds) *Les Défis de la publication sur le web: hyperlectures, cybertextes et méta-éditions*. Lyon: Presses de l'ENSSIB, 145–65.

Montaigne M. de 2004. *Les Essais*. Villey P., Saulnier V. L. and Conche M. (eds). Paris: Presses universitaires de France. https://www.lib.uchicago.edu/efts/ARTFL/projects/montaigne/.

Moore F. 2010. 'The poetry of the Super-Enlightenment: the theories and practices of Cazotte, Chassaignon, Mercier, Saint-Martin and Bonneville'. In: Edelstein D. (ed.) *The Super-Enlightenment: Daring to Know Too Much*. *SVEC* 2010:01. Oxford: Voltaire Foundation, 137–66.

Mortier R. 1981. *L'Orginalité. Une nouvelle catégorie esthétique au siècle des Lumières*. Geneva: Droz.

Mulsow M. 1997. 'Eclecticism or skepticism? A problem of the early Enlightenment'. In: *Journal of the History of Ideas* 58:3, 465–77. https://dx.doi.org/10.1353/jhi.1997.0028.

Norman L. 2015. 'La Querelle des Anciens et des Modernes, ou la métamorphose de la critique'. In: *Littératures classiques* 86, 95–114. https://doi.org/10.3917/licla1.086.0095.

Olsen M., Horton R. and Roe G. 2011. 'Something borrowed: sequence alignment and the identification of similar passages in large text collections'. In: *Digital Studies/Le Champ numérique* 2:1. http://doi.org/10.16995/dscn.258.

Page L., Brin S., Motwani R. and Winograd T. 1999. 'The PageRank citation ranking: bringing order to the web'. *Technical Report*. Stanford CA: Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/.

Pensées républicaines, pour tous les jours de l'année, à l'usage des enfans. Par l'Auteur du Catéchisme moral et républicain. 1794. Troyes: Sainton. https://artflsrv04.uchicago.edu/philologic4.7/bibbleue/navigate/203/.

Piles R. de 1699. *Abregé de la vie des peintres, avec des reflexions sur leurs ouvrages, et un traité du peintre parfait, de la connoissance des desseins, & de l'utilité des estampes*. Paris: Muguet. https://gallica.bnf.fr/ark:/12148/bpt6k6568782t.

Poirson M. 2017. 'Illuminismes'. In: Masseau D. (ed.) *Dictionnaire des anti-Lumières et des antiphilosophes (1715–1815)*. 2 vol. Paris: Honoré Champion, 726–31.

Pope A. 1713. *An Essay on Criticism: Written by Mr. Pope*. London: W. Lewis.

Quérard J.-M. 1828. *La France littéraire ou dictionnaire bibliographique des savants, historiens et gens de lettres de la France, ainsi que des littérateurs étrangers qui ont écrit en français, plus particulièrement pendant les XVIII^e^ et XIXe siècles*. 10 vol. Paris: Firmin Didot, vol. 2.

Racine J. 2015. *Britannicus. Tragédie*. Fièvre G., Fièvre E. and Fièvre P. (eds). In: *Théâtre Classique*. https://www.theatre-classique.fr/pages/pdf/RACINE_BRITANNICUS.pdf.

Roe G. 2012. 'Intertextuality and influence in the age of Enlightenment: sequence alignment applications for humanities research'. In: *Digital Humanities 2012*. Hamburg: Alliance of Digital Humanities Organizations. https://dh-abstracts.library.virginia.edu/works/1467.

Roe G. 2018. 'A Sheep in Wolff's clothing: Émilie Du Châtelet and the *Encyclopédie*'. In: *Eighteenth-Century Studies* 51:2, 179–96. http://doi.org/10.1353/ecs.2017.0059.

Rosson D., Mäkelä E., Vaara V., Mahadevan A., Ryan Y. and Tolonen M. 2023. 'Reception Reader: exploring text reuse in early modern British publications'. In: *Journal of Open Humanities Data* 9. https://openhumanitiesdata.metajnl.com/articles/10.5334/johd.101.

Rousseau J.-J. 2022. *Les Confessions.* L'Heuillet H. (ed.). Paris: Éditions Points.

Ryan Y. and Ahnert S. 2021. 'The Measure of the archive: the robustness of network analysis in early modern correspondence'. In: *Journal of Cultural Analytics* 6:3. https://doi.org/10.22148/001c.25943.

Ryan Y., Mahadevan A. and Tolonen M. 2023. 'A Comparative text similarity analysis of the works of Bernard Mandeville'. In: *Digital Enlightenment Studies* 1, 28–58. https://doi.org/10.61147/des.6.

Salmi H., Paju P., Rantala H., Nivala A., Vesanto A. and Ginter F. 2020. 'The Reuse of texts in Finnish newspapers and journals, 1771–1920: a digital humanities perspective'. In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 54:1, 14–28. https://doi.org/10.1080/01615440.2020.1803166.

Samoyault T. 2005. *L'Intertextualité. Mémoire de la littérature*. Paris: Armand Colin.

Schöch C. 2011. 'Le Temps du génie: attributs temporels du génie créateur et idées sur la temporalité au XVIII[e] siècle français'. In: *Revue des sciences humaines* 303, 137–55. https://hal.science/hal-00934410.

Smith D., Cordell R. and Mullen A. 2015. 'Computational methods for uncovering reprinted texts in antebellum newspapers'. In: *American Literary History* 27:3, E1–E15. https://doi.org/10.1093/alh/ajv029.

Starobinski J. 1995. 'Diderot et l'art de la démonstration'. In: *Recherches sur Diderot et sur l'Encyclopédie* 18–19, 171–90.

Suen C. Y. 1979. 'n-Gram statistics for natural language understanding and text processing'. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1:2, 164–72. https://doi.org/10.1109/TPAMI.1979.4766902.

Underwood T. 2017. 'A Genealogy of distant reading'. In: *Digital Humanities Quarterly* 11:2. https://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html.

Van der Lugt M. 2016. 'Bayle's labyrinth: the genesis, style, and structure of the *Dictionnaire* (1689–96)'. In: Van der Lugt M. (ed.) *Bayle, Jurieu, and the Dictionnaire Historique et Critique*. Oxford: Oxford University Press, 15–69. https://doi.org/10.1093/acprof:oso/9780198769262.003.0002.

Vesanto A., Nivala A., Rantala H., Salakoski T., Salmi H. and Ginter F. 2017. 'Applying BLAST to text reuse detection in Finnish newspapers and journals, 1771–1910'. In: Bouma G. and Adesam Y. (eds) *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*. Gothenburg: Linköping University Electronic Press, 54–58. https://aclanthology.org/W17-0510.

Virgil 1999. *Eclogues. Georgics. Aeneid, Books 1–6*. Fairclough H. R. (trans.), Goold G. P. (ed.). Loeb Classical Library 63. Cambridge MA: Harvard University Press.

Voltaire 1970. *La Henriade*. Taylor O. R. (ed.). In: *Œuvres complètes de Voltaire*. 205 vol. Oxford: Voltaire Foundation, vol.2.

Voltaire 1975. *Correspondence and Related Documents*. Besterman T. (ed.). In: *Œuvres complètes de Voltaire*. 205 vol. Oxford: Voltaire Foundation, vol.124.

Voltaire 2012. *Questions sur l'Encyclopédie* (VII). Cronk N. and Mervaud C. (eds). In: *Œuvres complètes de Voltaire*. 205 vol. Oxford: Voltaire Foundation, vol.42B.

Zinsser J. 2006. *Émilie Du Châtelet: Daring Genius of the Enlightenment*. New York: Penguin.