



HAL
open science

Modelling enlightenment: reassembling intertextual networks through data-driven research (ModERN)

Dario Maria Nicolosi, Glenn Roe

► **To cite this version:**

Dario Maria Nicolosi, Glenn Roe. Modelling enlightenment: reassembling intertextual networks through data-driven research (ModERN). Digital Enlightenment Studies, 2024, 2, pp.62-84. 10.61147/des.22 . hal-04828110

HAL Id: hal-04828110

<https://hal.science/hal-04828110v1>

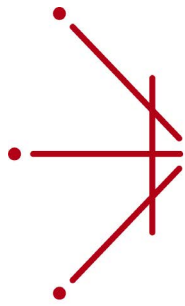
Submitted on 9 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



DIGITAL
Enlightenment
STUDIES

VOLTAIRE FOUNDATION
UNIVERSITY OF OXFORD

Modelling Enlightenment: reassembling intertextual networks through data-driven research (ModERN)

Dario Maria Nicolosi, Sorbonne Université, dario.nicolosi.92@gmail.com

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

Nicolosi D. M. and Roe G. 2024. 'Modelling Enlightenment: reassembling intertextual networks through data-driven research (ModERN)'. In: *Digital Enlightenment Studies* 2, 62–84.

DOI: [10.61147/des.22](https://doi.org/10.61147/des.22)

The European Research Council's ModERN project (Modelling Enlightenment: reassembling networks of modernity through data-driven research) is a pioneering five-year research initiative. This programme seeks to redefine the conventional understanding of 18th-century literary history by employing advanced data-modelling and analysis techniques. By developing a comprehensive corpus of 18th-century French texts and leveraging a range of data-science methodologies such as text-reuse detection and network analysis, the project aims to uncover novel research avenues and provide fresh insights into early-modern French print culture and its intertextual dynamics.

In this report, we discuss some theoretical points underlying our research; we explain the choices made in constructing our corpus and their implications; and we present some case studies to show the potential of our research and the most prudent methodologies to adopt.

Keywords: intertextuality, network analysis, text reuse, SNA



1. Introduction and scope

The European Research Council project ModERN (Modelling Enlightenment: reassembling networks of modernity through data-driven research) is a five-year research programme that aims to challenge received notions of 18th-century literary history through large-scale data modelling and analysis. By way of the creation of a new, extensive corpus of 18th-century French texts and mobilising various data-science methods like text-reuse detection and network analysis, we intend to open up new lines of research in early-modern French print culture and its intertextual connections.

Intellectually, our project aims to align itself with previous work that has highlighted the highly interconnected nature of 18th-century culture, analysing the social networks of intellectual figures (e.g. Brockliss 2002), the correspondence networks between members of the Republic of Letters (Comsa et al. 2016; Edmondson and Edelstein 2019) or focusing on the diffusion and reception of literary and philosophical works (Burrows 2018; Burrows 2020; Darnton 1982; Darnton 2021). However, while these important and stimulating studies have emphasised the circulation of people and ideas that contributed to defining the main ideological axes and the European scope of the Enlightenment, they analyse these exchanges from a purely material perspective (letters, books, sales records, etc.), reducing the textual content of these objects to data points. Our project will instead employ new techniques for large-scale text analysis to identify and analyse conceptual and intertextual networks over an unprecedented collection of 18th-century texts. As we know, 18th-century authors demonstrated great agility in their appropriation and reappropriation of both ancient and modern sources, relying on the shared cultural knowledge of their readers to identify these borrowings (Edelstein, Morrissey and Roe 2013). Today, however, given that most of these references remain hidden to contemporary readers, the identification of these intertextual relationships can provide new insights into the reciprocal influences, models and authorities that shaped the evolution of various ideological, political and aesthetic discourses of the period.

At its core, the project seeks to understand how the modern constellation of Enlightenment authors we have inherited today came into being; to uncover the cultural and ideological processes by which these writers, and by extension the texts and concepts they helped disseminate, became so indissolubly linked to the Enlightenment as an idea, while others – mostly forgotten today – were gradually excluded from these same assemblages. In order to re-establish these lost voices, or, to put it another way, to reassemble these lost networks, we need to drastically expand the corpus of texts on which we have traditionally drawn to understand the French Enlightenment and its reception.¹ Thankfully, this process of expansion is already underway, as the 18th century has benefited greatly, perhaps more than any other historical period, from the past two decades of digital transformation and the subsequent rise of the digital humanities (Burrows and Roe 2020; Paige 2021).

¹ For further insights into the implications of the large-scale analysis of ‘non-canonical’ texts, on their heuristic potential as well as the risks behind such broad analyses, see Moretti (2017).



Digital projects, databases and collections in 18th-century studies are now reaching a point of maturity in their elaboration, as well as a critical mass in number, such that we can now begin to think in terms of the literary or cultural systems, rather than individual works or authors, that inhabit our growing digital archives, at least in the francophone context. The main ModERN corpora have thus been built including not only canonical works and printed books that have been progressively digitised over the past two decades, but also more ephemeral texts such as private correspondence, pamphlets, newspapers and journals: the breadth of these collections provides a unique opportunity to trace a much broader range of intertextual practices than has traditionally been conceivable (Kristeva 1969; Barthes 1984; Genette 1992). By identifying and scrutinising a large swathe of these sorts of intertextual practices – from borrowings, citations, mentions and references to paraphrases and allusions, etc. – we gain a deeper understanding of the intricate web of influences that shaped literary and philosophical works but which also enriches our comprehension of the historical context in which these texts were produced.

Which 18th-century texts were most frequently ‘cited’, in what form and why? Which authors prove to be the most ‘influential’ – their words resonating and circulating the most within other texts of the same period? What verses, maxims and concepts seemed to gain (or lose) popularity over the course of the long 18th century? Do specific communities emerge, centred around an authority or a foundational text, or are textual exchanges rather to be found transversally, cutting across numerous literary and cultural fields? Similar reflections can be made regarding the reception of Antiquity. As we know, Greek and Latin literature form the cultural and educational foundations of the time, and will, starting with the *Querelle des Anciens et des Modernes*, eventually become an instrument for discussing and proposing new philosophical and aesthetic theories (Grell 1995; Norman 2011). Which ancient authors are most frequently cited, and in what context? In the original language or in translation? While the influence of the most important authors (e.g. Aristotle, Cicero, Plutarch) is well known, is it possible to unearth ‘secondary’ figures whose reception is nevertheless crucial to understanding strategies of reusing Antiquity during the Enlightenment?

From a technical and methodological standpoint, we ground our understanding of these intertextual exchanges as a specific instance and application of network modelling and analysis methodologies, particularly social network analysis (SNA) for literary-historical studies. Over the past decade or so, several research groups have begun to exploit the potential of applying the heuristic tools offered by SNA, developed mainly in the social and data science fields, to humanities projects.² But what are the implications behind applying this model to a phenomenon like intertextuality? What kind of information can be extracted, and how is it influenced by our choices of formalisation and representation? In general, it is always important to remember that any modelling attempt is, by definition, constructed with inherent biases related to the dataset in question (the extent and quality of the sources it is derived from), the tools used to create it

² On the ‘network turn’ in the humanities, see Ahnert et al. (2021). On the use of SNA in 18th-century studies, see Edmondson and Edelstein (2019).



(computational techniques and their performance), and the specific research objectives (which questions are posed, and how the results are interpreted). However, these potential obstacles remain surmountable if it is clear to the researcher that each model is more of a heuristic tool for analysis than a repository of immutable truth: models serve less to find definitive answers than to propose new questions. The ability to engage with an unprecedented amount of data offers specialists a new way to look at phenomena, in our case, intertextuality and influence in the 18th century.

We are confident that such large-scale projects are worthwhile, and that avenues of research opened up by such programmes can lead to new knowledge and, eventually, new historical paradigms (Moretti 2008; McCarty 2018). We will thus present the corpus construction and alignment methodologies that inform our models, and then conclude with some practical use-cases for exploring and analysing intertextual networks at scale.

2. Corpus

Thanks to institutional agreements with the University of Chicago, Gale Primary Sources, the University of Oxford and the Bibliothèque nationale de France (BnF), our primary corpora are drawn from four main sources: transcribed and curated data holdings from the ARTFL Project and the Voltaire Foundation;³ Paul Fièvre's Théâtre Classique database (transcriptions) of French theatre;⁴ digitised texts in French (via OCR) taken from the Gale Eighteenth Century Collections Online (ECCO) and The Goldsmiths' -Kress Library of Economic Literature;⁵ and texts drawn from the Gallica digital library housed at the BnF (OCR).⁶ Overall, our ingestion policy included the following criteria: digitised texts in French published roughly between 1685 and 1800 and, in the case of multiple editions of the same text, the earliest edition available across collections. Preference was given to transcribed versions of texts regardless of publication date.

Our main corpus of texts is thus the result of an amalgamation of several independent collections derived from distinct digitisation campaigns, the combination of which led to a series of inevitable problems that we had to address. From the form and content of the metadata to the text formats and TEI-XML file structures, each collection was unique and followed no real standard. Being conceived at different times and in response to specific research objectives, each digitisation project adopts its own encoding and classification logic, which introduces significant variability into the combined metadata of any large-scale corpus.

³ ARTFL holdings include relevant texts from the ARTFL-Frantext database as well as other open-access and subscription-based collections, including the *Bibliothèque bleue de Troyes*, the ARTFL *Encyclopédie* and *Dictionnaires d'autrefois*, see <https://artfl-project.uchicago.edu/>; in collaboration with ARTFL, the Voltaire Foundation has developed the TOUT Voltaire dataset, which was made available to our project, see <https://www.voltaire.ox.ac.uk/voltaire-lab/tout-voltaire/>.

⁴ See <https://www.theatre-classique.fr/>.

⁵ See <https://www.gale.com/c/eighteenth-century-collections-online-part-i> and <https://www.gale.com/c/making-of-the-modern-world-part-i>.

⁶ See <https://gallica.bnf.fr/>.



This variability comes also from the editorial, literary and historical specificity of each corpus, which influences the choices of the researchers who compiled and encoded our corpora. For example, the issue of paratextual elements takes on greater significance depending on the historical or linguistic nature of the corpus analysis one seeks to enact, begging the question of whether they should be included or excluded from digital editions. The ARTFL-Frantext database, for instance, has removed all non-authorial elements from its texts, in an effort to better represent the linguistic context in which they were produced.⁷ Other collections, such as ECCO, reproduce texts in their entirety, including any and all paratextual elements whether originating from the author or not.

These tensions are indicative of larger debates around digitisation protocols and, more specifically, digital scholarly editions and the TEI-XML encoding standard: how much to encode, and at what levels of granularity, are decisions often made by previous editors whose justifications may no longer be legible at the time of corpus construction, leading to editorial artefacts that can skew results downstream if one is not careful. And yet, the first (and perhaps only) necessary condition behind the construction of an analytical model is that the data within it are consistent with each other: if every choice and selection is justifiable in itself, a model is valid only if its elements are homogeneous and functional to the type of analysis envisioned.

Given the diversity of text-encoding options, it thus becomes necessary to seek automated or semi-automated methods to harmonise corpus metadata, and in particular titles and author names. The approach we adopted combines three digital methods that assess the similarity between strings of characters. In a first instance, we deployed two well-established methods developed in natural language processing (NLP) – Levenshtein distance and cosine similarity – in order to score the lexical similarity of author labels.⁸ The results were fairly promising, allowing us to ascertain that ‘CARMONTELLE, Louis Carrogis, dit Louis de Carmontelle (1717–1806)’ and ‘CARMONTELLE, Louis Carrogis de (1717–1806)’ were indeed the same author. This may seem intuitive to the human eye, and indeed it is, but for the computer they are two very distinct strings that inhabit the same XML field and are therefore treated as separate entities. More uncertain cases – author names that were too lexically dissimilar to be caught – ‘Anne-Claude-Philippe de Tubières, comte de Caylus’ and ‘Comte de Caylus (1692–1765)’, for instance – required the use of more direct techniques, such as comparing the two longest words in a string or systematically removing all dates before comparison. Thanks to these approaches, we were able to disambiguate our author names, which were then standardised across our corpus. The same semi-automated standardisation process was employed for the titles of texts, grouping the volumes of the same work under a single label, often distinguished by the mention of the volume number.

⁷ ARTFL-Frantext, like its French counterpart Frantext, was originally constructed by lexicographers compiling the *Trésor de la langue française* dictionary in the 1970s. See <https://artfl-project.uchicago.edu/content/artfl-frantext>.

⁸ On these two measures of textual ‘similarity’, see Buscaldi et al. (2020).



Here we encountered the thorny issue of likely duplicates, i.e. those texts with similar, but not identical, titles that in fact represent two (or more) versions of the same text. In order to make our dataset conform to the analytical model we have chosen – to model intertextual exchanges using graphs and SNA tools – the presence of a duplicate text can significantly alter the results. Two or more copies of the same text, identical or very similar, will tend to generate a very high number of co-occurrences among themselves (or just one, but concerning the entire text), strongly influencing the use of quantitative metrics derived from graph analysis; and the same happens for every single text that cites (or is cited by) this same multiple source, creating double, triple, etc., intertextual links. For the reasons previously mentioned, simple comparison of metadata is not effective, given the wide range of variables in the indication of authors, titles, dates, etc. The solution we found consists of exploiting precisely these two characteristics of duplicates: if the automatically detected co-occurrence is extremely long, or if it coincides with the near entirety of a document, it is highly likely that it is a duplicate; the same applies if two texts present an anomalous number of co-occurrences, which require inspection to explain.

This last case gives us cause to recall and insist on two key points. First, any automatic process must always be conducted in a supervised manner (hence our use of the term ‘semi-automated’): ambiguous cases are always numerous, given that the literary, editorial, and cultural reality of any period is always more complex and elusive than what can be strictly formalised. In the case of metadata, cases of homonymy can occur: often, different works that are interested in the same theme may have similar titles (*Essai sur la poésie épique, Essai sur les intérêts du commerce maritime...*). In the case of duplicate detection, some editorial forms, such as anthologies or encyclopaedic collections, exist precisely because they serve as repositories of long intertextual excerpts, which quantitative analysis alone would tend to indicate as duplicates. Therefore, each algorithmic intervention must be evaluated by a domain specialist.

Secondly, there are always margins of error, independent of the researcher’s intentions, which cannot be avoided because they are inherent to the nature of the data. For example, our digitised texts are fundamentally different in terms of the underlying quality of the textual data, transcribed to near 100 per cent accuracy in some cases, while others are the result of an automatic OCR process whose accuracy can vary greatly depending on the source and digitisation campaign. Automatic correction often introduces more errors than it corrects, and as a result, most of these texts retain high levels of OCR errors that can affect the performance of even the most robust text-reuse detection system. But, beyond the incompleteness of results (a common problem in both ‘traditional’ or digital research), it is the introduction of a non-homogeneous dimension (corrected texts versus OCR) that can skew data analysis: the co-occurrences of a corrected text will be much more easily retrievable than those of an OCR-generated text, and this can create significant variations in the results. However, this should not be considered an impediment to proceeding, but rather a warning to exercise caution across the entire data-processing pipeline. Understanding one’s dataset and its limitations is a first, necessary, step to ensuring that downstream tasks and results are not influenced by outside factors.



With the above observations in mind, we began the corpus construction phase of our project: once duplicates had been removed and the metadata standardised, we were left with a main research corpus of 13 133 documents (mainly books) totalling over 511 million words in total. Of these 13 133 documents, 3385 came from curated or transcribed sources while the remaining 9748 were the result of automatic OCR. A rough distribution of text genres in the corpus can be seen in Figure 1. These classifications were applied by the project team using a simplified version of the duc de La Vallière's 18th-century classification scheme for his private library, one of the most extensive in late Enlightenment France (see Van Praet 1783).

ModERN Corpus text genre

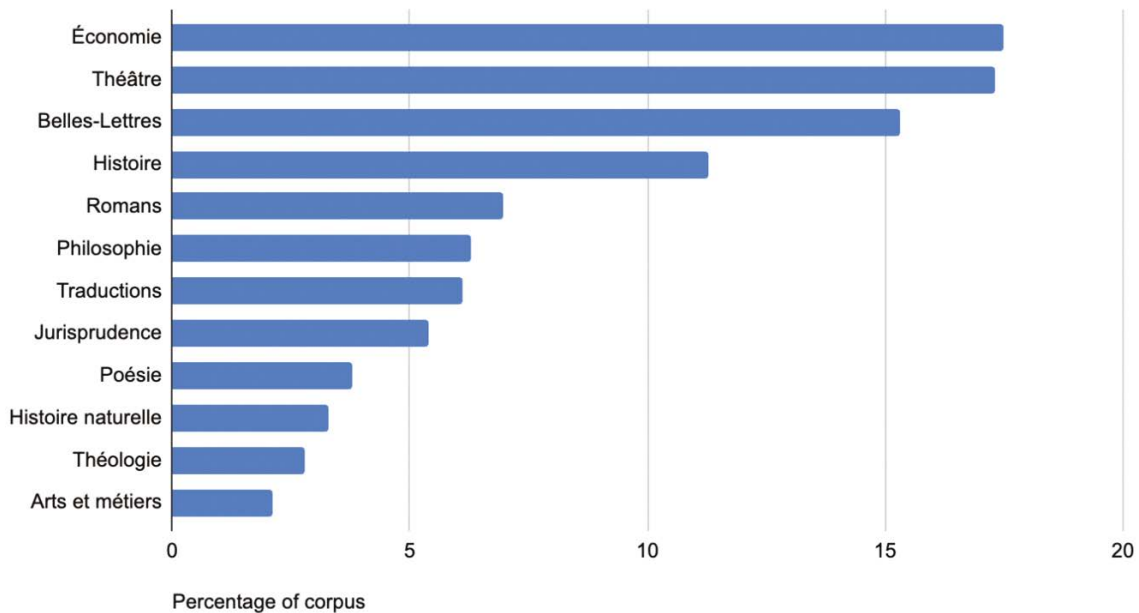


Figure 1 Distribution of text genre as percentage of total ModERN corpus.

The most prevalent authors in our corpus, i.e. those with 30 or more titles attributed to them, can be seen in Table 1.

A brief look at this list may raise some doubts, which need to be discussed, even if merely to draw some general methodological observations. At first glance, there seems to be a rather significant over-representation of Voltaire in our corpus: due to the editorial decisions of the Voltaire Foundation at the University of Oxford, the source for all of our Voltaire files, single poems, some just a few lines long, are considered individual works, alongside more substantial texts, such as the *Essai sur les mœurs* and *Candide*. In this case as well, therefore, this bias must be taken into account during interpretation: each individual poem must always be contextualised by taking into account the actual editorial history of the composition and its dissemination.



Author name	Number of works
Voltaire (1694–1778)	1057
Carmontelle (1717–1806)	76
Cicero (106–43 BCE)	70
Bernard de Fontenelle (1657–1757)	69
Horace (65–8 BCE)	58
Plutarch (c.46–c.120)	56
Denis Diderot (1713–1784)	56
Florent Carton Dancourt (1661–1725)	51
Henri-Louis Duhamel Du Monceau (1700–1782)	47
Jean-Jacques Rousseau (1712–1778)	44
Honoré-Gabriel Riquetti comte de Mirabeau (1749–1791)	44
Pierre de Marivaux (1688–1763)	41
Pierre Corneille (1606–1684)	39
Jacques Necker (1732–1804)	39
Étienne Clavière (1735–1793)	39
Louis-Sébastien Mercier (1740–1814)	38
Louis Petit de Bachaumont (1690–1771)	37
Claude-Louis-Michel de Sacy (1746–1794)	37
Molière (1622–1673)	36
Jean-Antoine-Nicolas de Caritat marquis de Condorcet (1743–1794)	35
Antoine François Prévost (1697–1763)	34
Olympe de Gouges (1748–1793)	34
Charles-Simon Favart (1710–1792)	33
Thomas Corneille (1625–1709)	32
Tacitus (c.55–c.120)	32
Pierre-Samuel Dupont de Nemours (1739–1817)	32
Nicolas-Edme Rétif de La Bretonne (1734–1806)	30

Table 1 Authors with more than 30 texts in the ModERN corpus

Also of note is the strong presence of classical authors; those that were constantly edited and re-edited in the 17th and 18th centuries. For the most part, these are translations that form a coherent sub-corpus of 527 texts, identified by our research team and mostly taken from Google Books as EPUB files which were then cleaned, corrected and transformed into TEI-XML files. The logic behind including these texts, even if they were not present in the initial corpora, is clear: the importance of the classical world in 18th-century culture (from the *Querelle des Anciens et des Modernes* to the French Revolution, passing through neoclassicism and the literary ‘retour à l’antique’ in the 1750s–1760s) is well known, and excluding them from our research



would have meant losing a large quantity of citations and references that contributed to shaping the political and aesthetic thought of the period.

Finally, we decided to include various canonical and indispensable texts composed before the 18th century. In this case as well, we must consider the specifics of our project: if, for example, Montaigne or Pascal were absent from our corpus, an incredibly high number of references to their texts, crucial for the 18th century, would be untraceable for us, and our representation of intertextual patterns would be severely compromised. In fact, it would not only be a serious lack of information but a structural problem of our network: if two texts cite Montaigne without his work being present, they would appear as citing each other when, in reality, they are both independently referring to a previous work. Thus, while the intertextual networks we aim to produce will be bounded chronologically between 1685 and 1800 as beginning- and end-dates, the corpus of texts used to generate the reuses must necessarily include works that fall outside these somewhat arbitrary markers. If earlier texts were available in our base collections, then we tried to include as many of them as possible.⁹

Having established our corpus and outlined the reasons and choices that underlie its creation, and after considering its limitations and their implications in formulating our working hypotheses for interpreting our results, we were ready to proceed to the identification of intertextual connections within this corpus and then to explore some possible research avenues and various types of analyses that can be marshalled even with preliminary results.

3. Alignment and use-cases

Today, multiple software applications are available for identifying text reuses in various datasets. Among the freely available tools for extracting textual reuses in large corpora, we considered those that use programming languages such as R (R *textreuse* package¹⁰), Java (TRACER¹¹), PHP/Perl (Tesseract¹²) and Python (Passim,¹³ BLAST,¹⁴ a tool designed for DNA sequence analysis; and Text-PAIR¹⁵). Although these tools offer similar functionalities, we ultimately opted for Text-PAIR as it was specifically designed to meet the needs of literary-historical research, scales well to large corpora and can be compiled as part of the

⁹ For English publications the choice of potentially interesting early modern texts that predate the 18th century is much easier: one can simply leverage the texts made available by the EEBO (Early English Books Online) project: <https://proquest.libguides.com/eebopqp>. No such resource yet exists for publications in French.

¹⁰ <https://github.com/ropensci/textreuse/>. See also Li and Mullen (2020).

¹¹ <https://www.etrapp.eu/research/tracer/>. See also Büchler et al. (2014) and Franzini et al. (2019).

¹² <https://github.com/tesseract/tesseract/>. See also Coffee et al. (2013).

¹³ <https://github.com/dasmiq/passim/>. See also Romanello and Hengchen (2021).

¹⁴ Basic Local Alignment Search Tool: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. See also Vesanto et al. (2017) and Salmi et al. (2020).

¹⁵ Pairwise Alignment for Intertextual Relations: <https://github.com/ARTFL-Project/text-pair/>. See also Olsen, Horton and Roe (2011).



PhiloLogic search and retrieval corpus analysis system.¹⁶ PhiloLogic creates full-text indices of corpora, leveraging metadata and other textual elements from TEI-XML files, and organises them into a database that can be easily queried. PhiloLogic word indices then subsequently form the basis on which Text-PAIR runs its sequence alignment matching algorithm. Additionally, Text-PAIR is easy to configure and relatively fast, which allowed us to experiment with several key matching parameters and pre-processing options, including lemmatisation and stemming.¹⁷ Finally, Text-PAIR is particularly well suited for extracting noisy reuses – i.e. those that because of OCR or other factors may include broken or highly dissimilar sequences.

Once our main corpus was built as a PhiloLogic instance, and after having settled on our text pre-processing and matching parameters, we compared the entire corpus to itself using Text-PAIR. This initial pass generated almost two million potential text reuses – i.e. similar passages that co-occur in at least two different texts. While impressive, these results should be taken with an appropriate measure of salt, as they include many, many ‘noisy’ alignments. That is, passages that are indeed similar but that do not necessarily constitute a ‘reuse’ in its fullest sense: formulaic expressions, legal boilerplate, publishing and print privileges, commonplaces, and so on. We are actively developing a semi-automatic alignment filter designed to eliminate much of this ‘noise’, although many cases will require human intervention at a finer-grained level of filtering. Based on initial estimates, around 80 per cent of the identified alignments will likely be eliminated as ‘noise’, leaving us with roughly 200 000 to evaluate further. We also plan to compare our main corpus with several secondary corpora, including dictionaries, private correspondences, printed pamphlets and the 18th-century press.¹⁸ In the meantime, we were eager to demonstrate the utility of our approach and present some preliminary results and possible use-cases leveraging these filtered alignments.

4. Plagiarism

The potential of such large-scale text reuse data is manifold, and the type of studies that can be conducted with it extremely varied. For instance, it is possible to identify obvious examples of plagiarism, cases in which an author takes advantage of their source’s relative obscurity to appropriate their text with impunity. Similarly, one can find references that are difficult to trace for contemporary researchers, drawn from works which are not considered canonical today but that circulated at the time and participated in the shared literary culture. The results of an alignment may therefore represent a real surprise for a researcher, who can (re)discover intertextual connections that would otherwise remain invisible.

¹⁶ <https://github.com/ARTFL-Project/PhiloLogic4/>. See also Tharsen and Gladstone (2020). Since 2015, Clovis Gladstone, associate director at the University of Chicago’s ARTFL project, has been the lead developer of both the PhiloLogic and Text-PAIR codebases. We are grateful for his invaluable support of the ModERN project.

¹⁷ For a discussion of our text-matching parameters and experimentation, see Fedchenko, Nicolosi and Roe (2024).

¹⁸ Our thanks to the ARTFL project for its collection of dictionaries (<https://artfl-project.uchicago.edu/content/dictionnaires-dautresfois>), *Electronic Enlightenment* for its correspondences (<https://www.e-enlightenment.com>), the Newberry Library for its collection of French pamphlets (<https://www.newberry.org/collection/research-guide/french-pamphlets>) and the BnF DataLab for the 18th-century press (<https://www.bnf.fr/fr/bnf-datalab>).



To take one example, we came across an unexpected case of plagiarism involving the Greek poet Sappho, a figure whose reception in the 18th century is somewhat problematic, and whose texts were included in the sub-corpus of classical translations mentioned above. While the scandal of Sappho's homosexual relationships was largely mitigated in the 18th century, leading to a more pathos-driven portrayal of her character in numerous rewritings that depict her as an unlucky lover, old and exiled, her erotic and passionate dimension remained ever-present. So much so that Mercier, in his utopian work *L'An 2440*, included her among the ancient texts that were unanimously burned as harmful. As Joan DeJean states, 'the eighteenth century may have continuously reaffirmed Sappho's heterosexuality because the fear of her sapphism had not been eradicated' (1989, p.118). Her fragments were thus, on the one hand, excluded from the canon of classical authors, and on the other, frequently translated into French and Latin, often in anthologies of Greek poets (Sappho 1681; Sappho 1712; Sappho 1758; Sappho 1781). These translations, which lack deep philological rigour, significantly altered the original text to promote this new characterisation, making the new versions scarcely recognisable compared to the originals (DeJean 1989, pp.116–67). Sappho is therefore published, but read with suspicion, or simply misread; like the poet herself, her texts are subject to misinterpretations and rewritings. Sappho's case is therefore exemplary in understanding how, without the use of automatic textual comparison systems, it would be impossible to find all the traces of her literary dissemination.

Given this context, we discovered that in 1788 her poems were the subject of an almost comical 'appropriation' in the poetry section of the provincial *Journal du Hainaut et du Cambrésis*: a certain M. Parent de Saint-Amand published not just one, but at least three poems by Sappho ('À ma bouteille', 'Ma mort', 'À Éléonore', this last rechristened 'La Discretion à Mlle de C. L. '), explicitly presenting himself as the author, with small changes to the texts that in no way justify this claim (see Parent de Saint-Amand 1788; Parent de Saint-Amand 1789a; Parent de Saint-Amand 1789b). In Figure 2 we see the text of Sappho compared to the *Journal* in the Text-PAIR web interface, and in Figure 3, the original text published in the *Journal*. Due to OCR errors present in the two compared texts, Text-PAIR only identified the part highlighted in red and marked the minor differences in green. However, the researcher can easily observe that the entire poem has been fully copied, with only the modification of the title and the change of addressee of the poem (Éléonore becomes Constance). This is one of the main advantages of Text-PAIR, which allows for the identification of textual reuse, even partial, when the texts are incorrect or fragmented. Clearly, in this case, this is plagiarism no matter how one looks at it, involving only the modification of the title and the change of the addressee of the poem.

In its exceptionality, this case raises several key points:

First, we must not forget the nature of the texts we are analysing: the 18th century had a very special relationship with Antiquity, both of proximity and acclimatisation (theories of artistic perfection, syncretism and the principle of the *belle infidèle*, for example). This allowed for great liberty in the reuse of ancient texts, which could be cut, transformed and distorted for any aesthetic purpose (Zuber 1968; Grell 1995, pp.307–24). A project such as ours can thus not only detect extreme cases of plagiarism such as the



1

Source

Sappho • *Poésies de Sappho, suivies de différentes poésies dans le même genre.* • 1781

, le'ons dans nos plaisirs la luiicre & le biruit; N: difois point au jour les secrets de la nuu'ir Aux regards inquiets dérobons nos caresss. L'aniour heureux se tranhit ainnement ! J'! cians pour toi les yeux d'nie nmere attnctive; Ik crains ce vieil Aigus, au ccuir de diamantr, Dont la vertu brusque & rétive Ne s'adoucit qu'a prix d'argent. Durant le jour tu n'es plus mon amante. Si te im'oitte tes yeux; garde-toi de rougir; Pscends à ton aimour le plus leger soupir; Aeffeei un air distrait; que ta voix féduisante Eite de frapper mon oreille & mon coeur; Xe mets dans tes regards ni trouble; ni langueur. Helas ! de mes conseils je me repens d'avance. Ma chere Elconore, au non de nos amoirs, NVimite pas trop bien cet air d'indif'rence, Il dirois, c':st un jeu i mai jc craindrois toujours. A LA MÊME. A LA ME ME. IAU sein d'un afylc champêtre Oû Damis trouvoi: le repos, Le plus arifibl; des

Figure 2 The text of Sappho as it appears in TextPair.

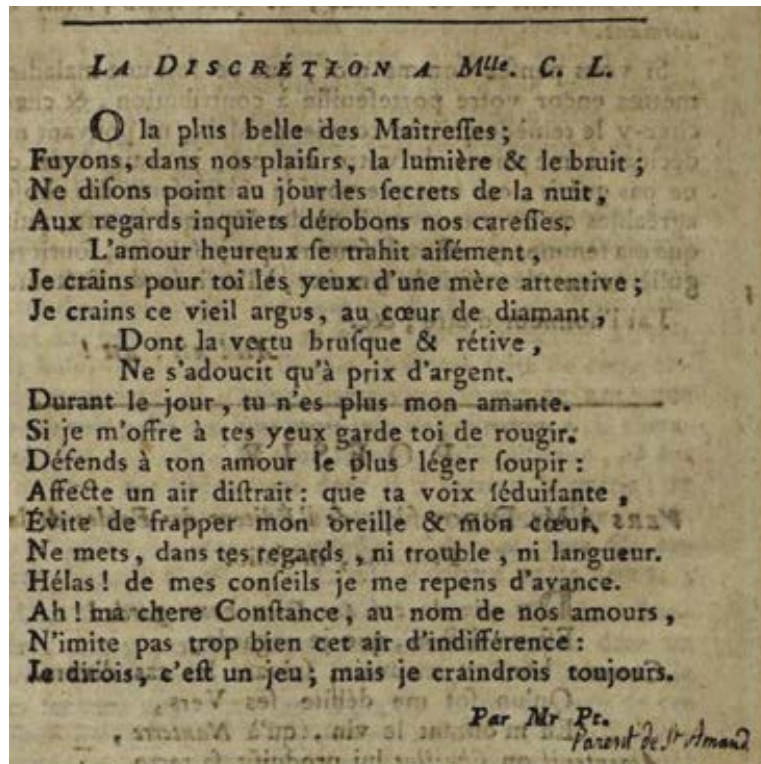


Figure 3 M. Parent's plagiarism as it appears in the *Journal du Hainaut et du Cambrésis*.



one above, but also identify co-occurrences and rewritings that, given the extreme nonchalance with which sources were often reused in the 18th century, would often otherwise have remained hidden.

Second, and most importantly, the detection of intertextual links is highly dependent on the ‘culture’ of the analyser: just as the readers of the *Journal* may not have recognised Sappho’s texts, similarly, we researchers today would have a difficult time identifying such references if not for its automatic detection across a large and heterogenous corpus that includes both classical translations and issues of obscure provincial newspapers. But, the importance of these small, almost serendipitous discoveries is significant, as they open up unpredictable and stimulating fields of research: who was this M. Parent? Why did he choose Sappho? How transparent was this plagiarism for the reader of the time? etc. Or it can equally serve as a starting point for new more general research questions: is it possible to find networks of dissemination of classical texts in the provinces, where the processes of cultural diffusion were different from those in the capital? Despite moral controversies, how extensive was the dissemination of Sappho’s texts in the 18th century? etc. It is precisely these sorts of questions, and their potential answers, that we hope will emerge once the project’s data is analysed and released to the public.

5. Quantitative analysis – uncovering the influence of individual authors

Aside from identifying and uncovering direct (or indirect) examples of text reuse, our project seeks more generally to understand the notion of authorial or textual ‘influence’ in the 18th century. Confronted with many lesser-known figures whose reception is ambiguous, this type of analysis is often difficult. It is undeniable, for instance, that Cicero was a key reference for 18th-century culture, both in terms of oratory and moral reflexions, but understanding the impact of less prominent figures, or those whose biographies might significantly affect contemporary judgment, is altogether more complex. Again, it comes down to a question of scale: Cicero is everywhere, and the possible variations in references to his works tend to lose their significance; Catullus much less so, and each individual reference takes on much greater importance, making the discovery of multiple examples particularly valuable.

Take Julius Caesar as one such example: his reception in the 18th century is highly ambiguous, both as a historical figure and as a writer (Mercier and Bièvre-Perrin 2024). These two aspects are often connected: for instance, in Rollin’s *Traité des études*, an important pedagogical text of the time, Caesar is simultaneously praised for his style as a historian (Grell 1995, pp.100–106) and condemned for his arrogance and for his political coup that undermined the institutions of the Roman Republic (Bedon 1985). Praised for his military achievements and for civilising Gaul (Grell 1995, pp.1113–19), Caesar is also highly criticised, on one hand politically, given his status as a ‘tyrant of usurpation’ and, on the other, as an historian, whose works are often considered devoid of concrete details (Poignault 1985). We need only think about Voltaire’s equivocal treatment of Caesar in *Rome sauvée*, where the character is both an alternative to Cicero’s passivity and one of the first possible accomplices in Catiline’s thirst for power (Nicolosi 2024), or the different nuances his image takes in revolutionary speeches (Parent 2022). How should one assess the period’s interest in such an



ambiguous figure? In this case, traditional exegesis can be enriched by the data that a project like ours can provide, confirming and nuancing existing interpretations.

The simplest method for assessing the influence of authors is to quantify and evaluate their presence in the texts of the time, either as subjects of theoretical works or as protagonists in literary texts, or to the extent that their works and words are cited or reused for their exemplarity or appropriateness. However, while the history of Julius Caesar is evidently the subject of countless comments and analyses (in our corpus, the expression ‘Jules César’ occurs 620 times), the case is different when examining his ‘active’ presence in the period’s imagination as a ‘speaking’ subject or ‘agent’, and thus, a more direct definition of influence in terms of symbolic impact or direct reuse of his statements.

We can start by evaluating the number and type of plays that are dedicated to Caesar, presented in Table 2. Based on Brenner’s catalogue of all 18th-century plays (1947), we notice that Caesar is relatively marginalised compared to other Roman historical figures (Laplace 1985): out of eight plays about him, two are translations of Shakespeare, where Caesar is little seen; two others give him larger roles but were not performed on Parisian or institutional stages; and the rest are primarily about Caesar’s death, where the protagonists are actually Brutus and Cassius. His appearances in works on other subjects are rare: besides the aforementioned *Rome sauvée*, we could also mention *Caton d’Utique* (1715) by François-Michel-Chrétien Deschamps. Clearly, we have here a historical character who is much talked about for the greatness of his deeds, but whom an 18th-century public seemingly does not want to ‘see’ or ‘hear’.

Author	Title	Genre	Acts	Year (and place if known) of first performance or publication
M.-A. Barbier	<i>La Mort de César</i>	tragedy	5	1710 (Paris)
Banières	<i>La Mort de Jules César</i>	tragedy	5	1728 (Toulouse)
Voltaire	<i>La Mort de César</i>	tragedy	3	1735 (Paris)
Abbé Saulx	<i>La Mort de César</i>	tragedy	unknown	1737 (Reims, Collège des Bons-Enfants)
P.-A. de La Place	<i>Jules César</i> (translation of Shakespeare)	tragedy	5	1746 (published in <i>Le Théâtre anglais</i>)
J.-B.-C. Delisle de Sales	<i>César ou les deux vestales</i>	play	1	1774 (at the residence of the prince d’Hénin)
P.-P.-F. Le Tourneur	<i>Jules César</i> (translation of Shakespeare)	tragedy	5	1776 (published in <i>Shakespeare traduit de l’anglais</i>)
Anonymous	<i>L’Héroïsme sénonais ou le siège de Sens sous Jules César</i>	drama	3	1781

Table 2 Table of all 18th-century plays dedicated to Julius Caesar.



But what about the presence of Caesar's words in other texts? The extracted reuse data from our project would seem to confirm the hypothesis we have just formulated. Caesar's two main works are the *Commentarii de Bello Gallico*, and the *Commentarii de Bello Civili*, which we consider first by finding quotations directly in Latin (the texts in the original language are easily found online). It is immediately evident that most references to these two historiographical treatises are extracted from the *De Bello Gallico* (57 co-occurrences), rather than the *De Bello Civile* (five co-occurrences), which as the history of an insurrection was clearly less popular in the context of French absolutism. Quantitatively, the quotations are not particularly numerous, and often appear in works by non-French authors and military texts, or those that take an interest in pre-Roman Gaul (Table 3). All told, Caesar seems to be used mainly as documentary support for historical or proto-ethnological enquiries, and rarely taken up or commented on for his rhetoric and sentences.

Author and nationality	Title and date of publication	Number of quotes
F.-R. Pommereul (French)	<i>Recherches sur l'origine de l'esclavage religieux politique du peuple, en France</i> (1783)	12
C. Guischartd (French)	<i>Mémoires critiques et historiques sur plusieurs points d'antiquités militaires</i> (1774)	8
G. Stuart (British)	<i>Dissertation historique sur l'ancienne constitution des Germains, Saxons et habitants de la Grande-Bretagne</i> (1794)	6
J.-R. Sinner (Swiss)	<i>Voyage historique et littéraire dans la Suisse occidentale</i> (1781)	5
R. Wallace (British)	<i>Essai sur la différence du nombre des hommes dans les temps anciens et modernes</i> (1754)	4
J.-B. de Mirabaud (French)	<i>Le Monde, son origine, et son antiquité</i> (1751)	2
H. Gautier (French)	<i>Traité des ponts</i> (1728)	2
T. Shaw (British)	<i>Voyages dans plusieurs provinces de la Barbarie et du Levant: contenant des observations géographiques, physiques, philologiques...</i> (1743)	2

Table 3 Table of works that most frequently cite Julius Caesar in Latin

Using our sub-corpus of translations, we find a similar set of practices of reuse concerning Caesar's works (Caesar 1678; Caesar 1763; Caesar 1785; Caesar 1786): most of the references appear, again, in military texts, confirming how Caesar was appreciated in the 18th century as a brilliant general rather than as a politician; his historical treatises are again used mainly to extract information about ancient Gaul (Table 4).

Interestingly, our alignments also unearthed a maxim attributed to Madame Des Houillères that recurs frequently in the various translations of Caesar: 'Nul n'est content de sa fortune, ni mécontent de son esprit'. The significant presence of this maxim, which has become proverbial, in many peritexts of Caesar's translations suggests a negative perception of this character, of the leader who, out of his personal



Author	Title and year of publication	Number of quotes
J.-B. Dubos	<i>Histoire critique de l'établissement de la monarchie française dans les Gaules</i> (1734)	11
J. Pagès	<i>Manuscrits de Pagès, marchand d'Amiens, écrits à la fin du 17^e et au commencement du 18^e siècle, sur Amiens et la Picardie</i> (1820)	10
A.-F. Boureau-Deslandes	<i>Essai sur la marine des anciens</i> (1768)	10
C. Guischart	<i>Mémoires critiques et historiques sur plusieurs points d'antiquités militaires</i> (1774)	8
M. de Saxe	<i>Les Rêveries dédiées à Messieurs les officiers généraux par Mr. de Bonneville</i> (1757)	3
D. Lescallier	<i>Vocabulaire des termes de marine anglois et françois</i> (1777)	3
Anonymous	<i>Un bon François de l'ordre des patriciens, aux bons François de l'ordre des plébéiens</i> (1789)	2

Table 4 Table of works that most frequently cite Julius Caesar translated in French

affirmation, overthrew the legitimate, albeit republican, state. Finally, our data confirm and corroborate what we had empirically perceived when looking at the theatrical output of the century: Caesar's 'voice' remained largely unheard in the 18th century, which preferred to discuss his exploits (mainly through Plutarch and historians of the Imperial period) than to use directly the expressions of a controversial figure in the context of monarchical France.

Here as above, the ability to compare a large number of texts allows us to confirm a hypothesis that would be otherwise difficult to prove in absolute terms – remaining, as these so often do, at the level of a 'hunch' (in this case, correct). Certainly, due to the nature of our data, it is possible that for technical reasons (OCR texts with many errors), some of Caesar's quotations may remain unidentified. But given the very low number of quotations found in both French and Latin in comparison to other Roman authors, the nature of the texts reusing the Roman general's treatises, and the massive presence of Madame Des Houillères couplet, there seems to be little doubt as to Caesar's scarce presence within the 18th-century literary field. As with mixed-mode methods in the social sciences, quantitative analysis, applied to narrow or inherently complex cases, becomes a solid ally of qualitative hypotheses and research.

6. The heuristic potential of networks

Finally, what about network analysis? How can it serve literary studies? At our current stage of research, the amount of data and its complexity make it difficult to obtain reliable results. The various obstacles that have appeared, and that we intend to overcome, concern, for example, the difficulties in classifying co-occurrences: when are they significant? When do they represent a true reuse and not simply a repetition of common or formulaic language with no intertextual value? While these questions remain very much



open, we are nonetheless encouraged by some of our preliminary results, which confirm known premises of 18th-century literature while hinting at the broader potential of the project as a whole.

Let us take, for example, one of the typical dichotomies of the theatrical and literary world of the 18th century, the subject of countless specialist debates: who is the most important point of reference for Enlightenment dramatists, Corneille or Racine? And how did these two giants of French classicism come to influence 18th-century playwriting? The parallel, already posed at the end of the 17th century (Mortgat-Longuet 2003) and continued in the 18th century (Goldzink 2003), accompanies the entire history of French literature. Many concessions would need to be made, but in general, anyone who has dealt with the history of 18th-century theatre would likely answer Racine, insofar as his use of the pathetic and the spectacular (e.g. *Athalie*) inform the main aesthetic developments of the century (Perchellet 2004a; Perchellet 2004b; Viala and Tunstall 2015, p.274). Can our current data confirm this first, intuitive hypothesis? And if so, how? To answer these questions, we need to first take into account our use of graph metrics for understanding network ‘influence’.

Our dataset of textual reuses allows us to generate graphs in which each text or author (i.e. the totality of texts attributed to the same author) represents a node, and in which the links between two points indicate an intertextual exchange that has taken place between two texts (or between the entire production of both connected authors). Each generated graph will thus have characteristics that can be analysed mathematically, and which give us information on the function and weight that each node (and thus each text or author) assumes within the system of intertextual exchanges.

The first metric that can be analysed is the degree of a node, i.e. the number of exchanges in which it is a protagonist, either as a quoting subject or as a quoted object. Using this degree measure, it is possible to generate a simple relative ranking by number of intertextual interactions. In our case, and with the current data at our disposal, Corneille appears in 14th position, while Racine comes in 11th. Thus, in the absolute, Racine appears in more exchanges than Corneille. A first confirmation of our hypothesis, but from which no conclusion can be drawn: beyond this purely quantitative measure, it is the quality and importance of these links in the general context of the intertextual network that can give us more and better indications.

Another measure that can be taken into consideration is PageRank, a measure for directed graphs which depends on the number and quality of links to a node (Liu et al. 2017; Labatut and Bost 2019). The underlying hypothesis behind this measure is that the most important nodes are likely to receive more links from other important nodes. An author has a high PageRank if a large number of authors reuse their text and these authors are themselves often reused and, therefore, considered important in the system. In other words, if an author who is ‘widely read’ quotes me, my ‘importance’ and the possibility of other people reading me increase. There is more chance of readers ‘stumbling across’ my text if Voltaire reuses me, than if 20 minor and little-read authors do. Now, if we take the ranking by PageRank of the authors in our project, Corneille is in sixth place, while Racine is in seventh: we could therefore deduce that Corneille is less cited



in the absolute, but cited by more 'important' authors, and that therefore his impact on the literary world of the 18th century is slightly stronger than that of Racine. But it is possible to refine this result even further.

Another fundamental measure in network analysis is betweenness centrality, which calculates the importance of a node with respect to its position in the graph (Labatut and Bost 2019; Grandjean and Jacomy 2019). More specifically, it calculates the number of times a node is on the shortest paths between two other nodes; the more central a node is and therefore the more it can act as a bridge between other nodes in the network, the higher its betweenness centrality. The more peripheral and isolated a node is from the rest of the nodes in the system, the lower its betweenness will be and the lower its impact on potential paths. Now, this measure emphasises the identification of paths, as in the case of information transmission and distribution flows (e.g. of people, energy, goods): if to get from point A to point C, the fastest route goes through point B, then B appears as a central node in the distribution of a resource, and will have a high betweenness – it becomes a hub. An airport of a large city allows for the connection of many airports of smaller cities, not otherwise connected to each other: its betweenness and importance in the transportation network are very high, enabling the passage of travellers between disconnected places. In an intertextual network, however, the relationships between texts/points do not describe a flow of information or represent a path between various points. If text B cites text A, and is itself cited by text C, the representation of this interaction ($A \rightarrow B \rightarrow C$) seems to suggest that B is a bridge between A and C; but in reality, it makes little sense to say that A and C are connected thanks to B, for numerous reasons (what B cites from A is not necessarily what C cites from B; and if the citation is the same, it is impossible to establish whether C cited A through B, or whether C directly cited A). More generally, even though the direction of the arrows may suggest a path from A to C, in reality, what is represented is only the relationship between A and B, and that between B and C.

In our network, betweenness does not indicate the importance of a text, but rather its ability to be involved in intertextual exchanges with different groups of texts or literary communities. A high betweenness implies that the node is connected to many nodes of the network which would be disconnected otherwise: a text that cites or is cited in politics, economics, theatre, religion, etc. will come into contact with very different parts of our intertextual graph, even in isolation. Conversely, a low betweenness implies that the node is poorly connected, or connected to a homogeneous group of texts that tend to quote each other, without connections with other areas of the network. Typical cases of low betweenness are found in religious texts or legal documents, which are very present in texts of the same nature, but rarely found in texts dealing with other subjects.

Our network's betweenness centrality ranking finds Racine in sixth position and Corneille 196th. The latter's position in the network is thus much more marginal than that of the former. The explanations for this large gap, even in the face of similar degree and PageRank measures, may be multiple: Racine seems to act as a 'bridge' between different groups in our network, both as a citing author (think of his links with the religious world of Port-Royal) and as a quoted author (his verses become part of popular culture, and quoted in contexts that do not concern theatrical dramaturgy). On the contrary, Corneille remains a highly



cited author (his degree measure is very high), but by homogeneous groups, or by a few individual authors in an intense manner. Voltaire's *Commentaires sur Corneille* (1764), as well as other poetic texts (La Harpe's *Lycée*, 1739, for example), quote his works extensively, and we thus understand why Corneille's PageRank is so high, but these texts all belong to the world of *Belles-lettres* alone, and links with other groups and areas of the network remain weak.

While it therefore remains difficult (and probably pointless) to answer definitively the question we have posed regarding the importance of Corneille or Racine in 18th-century culture, the possibility of enriching our knowledge through the integration of graph metrics extracted from network analysis has allowed us to imagine new and different research hypotheses. The two authors seem to participate, in a qualitatively different way, in the network of transmission of intertextual material: Racine seems to be a more transversal author, and his verses are extracted from their context and reused in different spheres; Corneille remains a very important literary reference point, but with a few exceptions, his words (poetic, but also theoretical, e.g. his *Trois discours sur le poème dramatique*) resonate mainly in literary circles. Even as we await more conclusive and extensive data, this simple insight opens up new avenues of research: which categories of texts quote Racine the most? How do his verses – extracts from tragic texts, and decontextualised – manage to take on different valences and become meaningful? How does Corneille become a point of reference – by adherence or contrast – to the dramatic poetics of the 18th century?

7. Conclusion: next steps

In this brief account, we have described the basic workings of our project, and presented some possible lines of research that it can help to identify and enrich. The analyses of the plagiarism of Sappho and the dissemination of Caesar's texts represent an initial demonstration of the advantages of applying digital methods in the discovery of intertextual connections, which would otherwise be either unrecoverable or too numerous for traditional close-reading approaches. As such, each piece of data becomes the basis for returning to the sources, leading to new interpretations that either reaffirm or challenge common critical assumptions. On the other hand, the Corneille/Racine comparison shows how the use of SNA paradigms and metrics in literature and its internal intertextual links is both possible and potentially capable of refining common exegesis, offering new evidence for established theories or uncovering new patterns that only large-scale distant-reading analyses can reveal (Underwood 2019).

Clearly, we are still at an early stage of our project, and much work remains to be done to make our results meaningful. Our goal is to create and define profiles for each text/author node in our network, created on a mathematical basis in relation to the various metrics presented (and others belonging to the field of graph studies – e.g. closeness centrality, clustering).¹⁹ For example, since ours is an oriented graph (the links

¹⁹ See Labatut and Bost (2019); Grandjean and Jacomy (2019).



connecting the works have a direction, there is a source and a target of each citation), it is possible to define an author-text node as an Authority (high number of texts citing it) or as an Observer (high number of texts cited); the same goes for the category of Mediator, applicable to a node whose betweenness centrality and PageRank measures are both high. Other possible categories will emerge through the combination of the various metrics we intend to calculate. We also envision analysing the context of each alignment, i.e. defining by topic modelling or sentiment analysis the ‘intention’ behind each quotation/reuse. For example, while Corneille is quoted at length and commented on in Voltaire’s *Commentaires*, the criticisms that the *philosophe* levies against the 17th-century playwright certainly contain a different set of value judgements than those found in La Harpe’s *Éloge de Racine* (1772).

These future perspectives notwithstanding, our general hope is that our large-scale treatment of intertextual links in the 18th century will allow us to verify, in a new and novel way, some of the most widespread literary hypotheses, and to offer the entire scholarly community the tools to conduct such research themselves. Once organised in a database, all our data will be available online and interrogated in, we hope, an intuitive manner, allowing any researcher to verify their own hypotheses on the circulation and diffusion of texts and authors in the 18th-century French literary field.

References

- 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>.
- Barron A., Huang J., Spang R. and DeDeo S. 2018. ‘Individuals, institutions, and innovation in the debates of the French Revolution’. In: *Proceedings of the National Academy of Sciences* 115:18, 4607–12. <https://doi.org/10.1073/pnas.1717729115>.
- Barthes R. 1984. *Le Bruissement de la langue*. Paris: Seuil.
- Bedon R. 1985. ‘César dans le *Traité des études* de Charles Rollin’. In: Chevallier R. (ed.) *Présence de César*. Paris: Les Belles Lettres, 275–85.
- Brenner C. 1947. *A Bibliographical List of Plays in the French Language 1700–1789*. Berkeley CA: Edwards brothers.
- Brockliss L. W. B. 2002. *Calvet’s web: Enlightenment and the Republic of Letters in Eighteenth-century France*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780199247486.001.0001>.
- 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. 2018. *The French Book Trade in Enlightenment Europe*. London: Bloomsbury Academic.
- Burrows S. 2020. ‘The FBTEE revolution: mapping the *Ancien Régime* book trade and the future of historical bibliometric research’. 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, 167–94.



- 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.
- Buscaldi D., Felhi G., Ghouli D., Le Roux J., Lejeune G. and Zhang X.** 2020. 'Calcul de similarité entre phrases: quelles mesures et quels descripteurs?'. In: Cardon R., Grabar N., Grouin C. and Hamon T. (eds) *Actes de la 6^e conférence conjointe Journées d'études sur la parole (JEP, 33^e édition), Traitement automatique des langues naturelles (TALN, 27^e édition), Rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues (RÉCITAL, 22^e édition). Atelier Défi Fouille de Textes*. Nancy: ATALA and AFCP, 14–25. <https://aclanthology.org/2020.jeptalnrecital-deft.2>.
- Caesar J.** 1678. *Les Commentaires de César, de la traduction de N. Perrot, sieur d'Ablancourt. Édition nouvelle revue et corrigée*. Perrot d'Ablancourt N. (trans.). Amsterdam: A. Wolfgang.
- Caesar J.** 1763. *Les Commentaires de César [...]. Nouvelle édition augmentée de notes historiques et géographiques, et d'une carte nouvelle de la Gaule et du plan d'Alise, par M. Danville*. Perrot d'Ablancourt N. and Le Mascrier J.-B. (trans.). Amsterdam: Arkstee & Merkus.
- Caesar J.** 1785. *Commentaires de César, avec des notes historiques, critiques et militaires*. Turpin de Crispé L. (trans.). Montargis: C. Lequatre and Paris: C.-G. Leclerc.
- Caesar J.** 1786. *La Guerre de Jules César dans les Gaules*. De Précis (trans.). Paris: Imprimerie royale.
- Coffee N., Koenig J.-P., Poornima S., Forstall C., Ossewaarde R. and Jacobson S.** 2013. 'The Tesseræ Project: intertextual analysis of Latin poetry'. In: *Literary and Linguistic Computing* 28:2, 221–28. <https://doi.org/10.1093/llc/fqs033>.
- Comsa M. T., Conroy M., Edelstein D., Edmondson C. S. and Willan C.** 2016. 'The French Enlightenment network'. In: *The Journal of Modern History* 88:3, 495–534. <https://doi.org/10.1086/687927>.
- Darnton R.** 1982. *The Literary Underground of the Old Regime*. Cambridge MA: Harvard University Press.
- Darnton R.** 2021. *Pirating and Publishing: The Book Trade in the Age of Enlightenment*. Oxford: Oxford University Press.
- DeJean J.** 1989. *Fictions of Sappho, 1546–1937*. Chicago IL: University of Chicago Press.
- 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. <https://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>.
- 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>.
- Genette G.** 1992. *Palimpsestes: la littérature au second degré*. Paris: Seuil.
- Goldzink J.** 2003. 'Le torrent et la rivière'. In: Declercq G. and Rosellini M. (eds) *Jean Racine, 1699–1999*. Paris: Presses Universitaires de France, 719–28.
- Grandjean M. and Jacomy M.** 2019. 'Translating networks: assessing correspondence between network visualisation and analytics'. *Digital Humanities* 2019. Utrecht: HALSHS. <https://shs.hal.science/halshs-02179024>.
- Grell C.** 1995. *Le Dix-huitième siècle et l'antiquité en France: 1680–1789*. SVEC 330–31. Oxford: Voltaire Foundation.



- Hamzehei A., Jiang S., Koutra D., Wong R. and Chen F.** 2017. 'Topic-based social influence measurement for social networks'. In: *Australasian Journal of Information Systems* 21. <https://doi.org/10.3127/ajis.v21i0.1552>.
- Kristeva J.** 1969. *Sēmeiōtikē. Recherches pour une sémanalyse*. Paris: Seuil.
- Labatut V. and Bost X.** 2019. 'Extraction and analysis of fictional character networks: a survey'. In: *ACM Computing Surveys* 52:5, 1–40. <https://doi.org/10.1145/3344548>.
- Laplace R.** 1985. 'Le personnage de César à la Comédie-Française'. In: Chevallier R. (ed.) *Présence de César*. Paris: Les Belles Lettres, 293–304.
- Li Y. and Mullen L.** 2020. *textreuse: Detect Text Reuse and Document Similarity*. <https://docs.ropensci.org/textreuse>.
- Liu Q., Xiang B., Jing Yuan N., Chen E., Xiong H., Zheng Y. and Yang Y.** 2017. 'An influence propagation view of PageRank'. In: *ACM Transactions on Knowledge Discovery from Data* 11:3, 1–30. <https://doi.org/10.1145/3046941>.
- McCarty W.** 2018. 'Modeling the actual, simulating the possible'. In: Flander J. and Joannidis F. (eds) *The Shape of Data in Digital Humanities*. London: Routledge, 264–84.
- Mercier C. and Bièvre-Perrin F.** (eds) 2024. *Jules César, construction d'une image de l'Antiquité à nos jours*. Besançon: Presses universitaires de Franche-Comté.
- Moretti F.** 2008. *Graphes, cartes et arbres. Modèles abstraits pour une autre histoire de la littérature*. Paris: Les Prairies ordinaires.
- Moretti F.** (ed.) 2017. *Canon/Archive. Studies in Quantitative Formalism from the Stanford Literary Lab*. New York: n+1 Foundation.
- Mortgat-Longuet E.** 2003. 'Aux origines du parallèle Corneille-Racine: une question de temps'. In: Declercq G. and Rosellini M. (eds) *Jean Racine, 1699–1999*. Paris: Presses Universitaires de France, 703–17.
- Most G. W.** 2008. 'Réflexions de Sappho'. Rabau S. and de Gandt M. (trans.). In: *Fabula-LhT* 5. <https://doi.org/10.58282/lht.832>.
- Nicolosi D. M.** 2024. 'La valeur symbolique de l'espace scénique dans les tragédies romaines et grecques de Voltaire'. In: *Revue Voltaire* 22, 121–36.
- Norman L. F.** 2011. *The Shock of the Ancient. Literature and History in Early Modern France*. Chicago IL: University of Chicago Press.
- 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. <https://doi.org/10.16995/dscn.258>.
- Paige N.** 2021. *Technologies of the Novel: Quantitative Data and the Evolution of Literary Systems*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108890861>.
- Parent H.** 2022. *Modernes Cicéron. La romanité des orateurs révolutionnaires et de l'Empire (1789–1807)*. Paris: Classiques Garnier.
- Parent de Saint-Amand [given name unknown]** 1788. 'À ma bouteille'. In: *Journal du Hainaut et du Cambrésis, par M. le Cher de Limoges, membre de plusieurs académies* 45, 378.
- Parent de Saint-Amand [given name unknown]** 1789a. 'Ma mort'. In: *Journal du Hainaut et du Cambrésis, par M. le Cher de Limoges, membre de plusieurs académies* 4, 35–36.
- Parent de Saint-Amand [given name unknown]** 1789b. 'La Discrétion à Mlle de C. L.'. In: *Journal du Hainaut et du Cambrésis, par M. le Cher de Limoges, membre de plusieurs académies* 7, 60.



- Perchellet J.-P.** 2004a. *L'Héritage classique: la tragédie classique entre 1680 et 1814*. Paris: Honoré Champion.
- Perchellet J.-P.** 2004b. 'Corneille et ses publics au XVIII^e siècle.' In: *Dix-septième siècle* 225, 549–57. <https://doi.org/10.3917/dss.044.0549>.
- Poignault R.** 1985. 'Napoléon I^{er} et Napoléon III lecteurs de Jules César.' In: Chevallier R. (ed.) *Présence de César*. Paris: Les Belles Lettres, 329–45.
- Romanello M. and Hengchen S.** 2021. 'Detecting text reuse with Passim.' In: *Programming Historian*. <https://doi.org/10.46430/phen0092>.
- 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>.
- Sappho** 1681. *Les Poésies d'Anacréon et de Sappho, traduites de grec en François, avec des remarques*. Dacier A. (trans.). Paris: D. Thierry & C. Barbin.
- Sappho** 1712. *Les Odes d'Anacréon et de Sappho en vers François, par le poète sans fard*. Gacon F. (trans.). Rotterdam: Fritsch & Böhm.
- Sappho** 1758. *Anacréon, Sappho, Moschus, Bion, Tyrthée, etc., traduits en vers François*. Poinset de Sivry L. (trans.). Nancy: P. Antoine.
- Sappho** 1781. *Poésies de Sappho, suivies de différentes poésies dans le même genre*. Billardon de Sauvigny E.-L. (trans.). London: [n.pub.].
- Tharsen J. and Gladstone C.** 2020. 'Using Philologic for digital textual and intertextual analyses of the *Twenty-Four Chinese Histories* 二十四史.' In: *Journal of Chinese History 中國歷史學刊* 4:2, 558–63. <https://doi.org/10.1017/jch.2020.27>.
- Underwood T.** 2019. *Distant Horizons, Digital Evidence and Literary Change*. Chicago IL: University of Chicago Press.
- Van Praet J. B. B.** 1783. *Catalogue des livres de la bibliothèque de feu M. le duc de La Vallière*. 4 vol., Paris: Guillaume De Bure. <https://catalogue.bnf.fr/ark:/12148/cb365379105>.
- 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>.
- Viala A. and Tunstall K.** 2015. *L'Âge classique et les Lumières*. In: Viala A. 2014–2017. *Une histoire brève de la littérature française*. Paris: Presses universitaires de France.
- Zuber R.** 1968. *Les Belles infidèles et la formation du goût classique*. Paris: A. Colin.