An Agenda for the Study of Intertextuality

Neil Coffee

An Agenda for the Study of Intertextuality[*]

NEIL COFFEE

State University of New York at Buffalo

SUMMARY: The study of intertextuality has been a central pursuit of scholars of Greek and especially Latin literature. It promises to reveal the meaning of texts for original audiences, trace authorial influence, and illuminate an aspect of literary artistry. Yet inconsistent standards and the scattering of insights across publications have hindered progress. This article proposes restoring momentum toward the goals of intertextual study through an agenda of representing intertexts in a standard digital form susceptible to complex and systematic analysis.

INTRODUCTION

FROM ANCIENT ALEXANDRIA TO TODAY, LITERARY SCHOLARS HAVE SPENT considerable energy studying intertextuality, illuminating how texts relate to one another to create literary effects. They do this to understand authors'

artistic practices; to see how authors influenced one another; and to understand what literary artifacts meant to original readers and listeners. When, for example, we recognize an echo of the famous opening of the *Aeneid*, *arma virumque cano* ("arms and the man I sing") at the beginning of the *Amores*, *arma gravi numero* ("arms in a heavy meter . . ."), we uncover a new level of significance in Ovid's first published collection of poetry.[1]

In the past, scholars relied on their memories and libraries to discover textual connections. Nowadays, they also use a variety of digital resources, from simple searches to tools that allow for rapid, automatic identification of potential intertexts. Yet these same digital tools can exacerbate existing problems with intertextual study. Scholarly notations of intertexts are already scattered across commentaries, articles, and monographs. Now digital tools can produce many more candidate intertexts that need to be scrutinized and potentially published.

This article proposes that, rather than overwhelming the study of intertextuality, digital resources can instead provide a means of more fully attaining its goals. It begins with a survey of the state of digitally-enhanced intertextual study. It then suggests a way forward, building on the tools surveyed, that centers on the adoption and creation of digital standards for referring to intertexts. Classical literature, with its highly developed digital resources, is the focus of the article, but it presents ideas that can be applied to other literatures.[2]

THE DIGITAL DEFINITION OF INTERTEXTUALITY

In order for a digital tool to find an intertext, it must have a description of what an intertext is. A digital approach necessarily proceeds from detectable language features, namely those that make one piece of text sufficiently marked to recall another.[3] An increasing number of features are becoming detectable. Exact quotations and repeated words can be found with text string searches

---

[1] *Aeneid* 1.1 and *Amores* 1.1. For more on Ovid's programmatic interaction with the opening of the *Aeneid* see McKeown 1987. For a recent survey of work on intertextuality, focusing on Latin literature, see Coffee 2012.

[2] This article builds upon the perspective on the digital study of intertextuality offered by Coffee et al. 2012.

[3] See Hinds 1998 and Farrell 2005 for two efforts at defining intertextuality in classics. Computationally compatible definitions derive in spirit from the Russian formalist tradition, in particular Jakobson's assertion that literary and poetic language can be understood as being built up from linguistic elements (Jakobson 1960; Erlich 1980 on the Russian formalists). For "markedness" and "sense" as the necessary features of an intertext, see Fowler 1997.

carried out on the Perseus website for Greek or Latin, or on the Packard Humanities Institute (PHI) website for Latin.[4] A variety of intertextual search tools, described further below, provide for searching for similarities of context, lemma, meaning, meter, sound, word frequency, and word order.[5] Syntactic similarities have been studied in a focused way.[6] Efforts to create general syntactical parsers for Greek and Latin have resulted in data sets of fully syntactically parsed Greek.[7] Word co-occurrence frequency, namely how often words occur together, is another feature being researched.[8] Work has been done toward defining the verbal markers of intertextuality known as *verba dicendi*, or words that introduce quotations or paraphrases.[9]

In addition to searching for these standard language features, a computational approach enables the discovery of latent features not intuitively recognized by human readers. For example, Levenshtein edit distance refers to the number of changes necessary to turn one word into another. The edit distance between *ensis* ("sword") and *mensis* ("month") is one: to get from *ensis* to *mensis*, we must make only one change, adding the letter "m" at the

---

[4] http://www.perseus.tufts.edu/hopper/search; http://latin.packhum.org/search.

[5] Tesserae provides an experimental search for semantic relatedness for Greek to Latin (http://tesserae.caset.buffalo.edu/cross.php), as well as an experimental search for context similarity using a topic modeling approach (http://tesserae.caset.buffalo.edu/cgi-bin/lsa.pl). No intertextual search yet accounts for section boundaries. Sound similarity is potentially highly tractable, but subject to various definitions. Approaches to sound similarity have been developed by both *Tesserae* and *Musisque Deoque*. *Tesserae* uses three-letter sequences ("character trigrams"), an option available from its main search pages. Other features could be listed. On the more tractable end of the spectrum, we could add wordplay, as in the case of acrostics imitated across works, which are difficult for humans to discern but relatively simple for machines. See, e.g., Castelletti 2014 on the acrostics of Valerius Flaccus's *Argonautica* in their relation to Aratus. At the more subjective end, we could add the repetition of complex notions such as plot, tone, or character types. The last has been described in digital terms by Bamman, O'Connor and Smith 2013, Bamman 2014, http://tesserae.caset.buffalo.edu/blog/intertextuality-beyond-words-by-david-bamman/. This sort of approach could be used to meet the need in the study of Latin poetry, described already by Griffin 1986, for richer accounts of literary characters.

[6] Dexter et al. 2017.

[7] See the data set of Giuseppe Celano of Leipzig University Computer Science, available at https://github.com/gcelano/AncientGreekUD1. Descriptions of work on automated analysis of Greek and Latin syntax include Bamman and Crane 2011 and Gorman and Gorman 2016.

[8] This is ongoing work at the Tesserae Project.

[9] Berti et al. 2009, Almas and Berti 2013. Heinze 1993: 197–98 remarks that Roman poets had a particular habit of using such expressions to acknowledge their sources.

beginning of *ensis*. Edit distance does not correspond to a language feature, but detects multiple features to various degrees. It can capture most lemma similarities, since most words from the same lemma share the letters of a stem (e.g., *amico* and *amicis*, two forms of "friend"), but misses some lemma similarity, such as *tuli* as a form of *fero* ("bear" or "carry"). It accounts for string similarities—how many letters the words have in common—and so sound similarities as well, as in the case of *ensis* and *mensis*, though it does not distinguish among these features.[10]

In many cases, automatic detection of language features still needs refinement. Computational identification of a word's lemma (dictionary headword) can produce errors when two different lemmata have the same inflected forms. In a test of Tesserae intertext discovery, the form *equis*, which Latin readers would ordinarily recognize as "[to/for/with] horses" from the stem *equus*, was also associated with the dictionary entry for the rare verb *equio*, and so taken to mean "you are in heat." Variant manuscript readings must be addressed to ensure correct word matching and metrical scansion.[11]

These sources of imprecision can in principle be addressed, but we can nevertheless already accurately detect a large percentage of language feature similarities. How, though, do we determine which similarities constitute a meaningful intertext? One solution is to take anything that scholars have called an intertext and derive its formal features. For example, the *Aeneid* and *Amores* phrases cited above, which form a well-recognized intertext, have similarities of lemma (*arma*), sound (*virumque~gravi*, *cano~numero*), metrical rhythm (dactyl-dactyl-long vowel), and proximity to the beginning of the work.[12]

---

[10] For a discussion of edit distance in intertextual study through examples, see Chaudhuri and Dexter 2017. A notable computer science article on Levenshtein edit distance is Wagner and Fischer 1974.

[11] Among the variety of language analysis tools offered by The Classical Language Toolkit (http://cltk.org) are lemmatizers that can address the problem of ambiguous lemmata. *Musisque Deoque* has its own text editions that include manuscript variants and emendations.

[12] In a study of Lucan *Civil War* book 1 and Vergil's *Aeneid*, Coffee et al. 2012: 415 found that sensitivity to the following features would capture all the intertexts recorded by commentators: two-word (exact word or lemma) identity (58%), one identical word + semantic context (16%), one identical word + synonym (9%), semantic context only (8%), two synonyms (7%), one identical word + syntax (1%), and one identical word + sound (1%). These figures differ slightly from those found in the article, accounting only for parallels found in commentaries, and so excluding interpretable parallels found by Tesserae search, reducing the number of instances in the first group, lemma or exact-word identity, from 146 to 100. This change is meant to give a clearer view of the composition of parallels as identified by commentators.

After type of similarity, frequency is the most important determiner. In general, the rarer the features and the rarer their combination, the more striking the similarity is likely to be, and so the more likely to be deemed a meaningful intertext. In this case, Ovid combines relatively common features (a dactyl-dactyl-long vowel rhythm; the sounds *vi* and *o*) with relatively rare ones (appearance of *arma*, proximity to work opening) to make for an overall rare combination that distinctly echoes the *Aeneid*. Detection systems use frequency measures to reduce the false positives that occur when collecting all instances of feature similarity.

Of course, if we searched other texts for just the rare combination of features common to the openings of the *Aeneid* and *Amores*, we would miss the majority of significant intertextuality. What is needed is a large benchmark set of recognized intertexts with a full set of features on which to train a search. This requires first locating and collecting the intertexts. It is not always obvious which scholarly works contain them and where they are in the works. Even when a scholar marks a connection between two texts, it is often unclear whether it is to be understood as a meaningful intertext. The abbreviation "cf.," for example, could mean that the passage under consideration echoes another one or just that it contains a similar grammatical usage. Finally, the collated intertexts must be formatted so that their features are uniformly analyzable. Despite these challenges, work has already begun on collecting such benchmark sets.[13] By tuning search to recognize such features, we can refine computational sensitivity to language features to more closely emulate scholarly sensitivity to intertexts. Greater computational sensitivity will enable further definition, and possibly expansion, of the concept of intertextuality, and so better searches.[14]

Such an effort will not be fully successful until it can at least detect the subtler forms of intertextuality explicitly recognized by ancient readers. For example, the ancient critic Macrobius found Vergil, in his description of a plague in *Georgics*, echoing an analogous description in Lucretius's *De Rerum Natura*.

> praeterea iam <u>nec</u> mutari pabula refert
> quaesitaeque nocent artes; cessere magistri.

---

[13] See, e.g., http://tesserae.caset.buffalo.edu/blog/collected-benchmark-sets/.
[14] On the latter point, see Coffee et al. 2012: 414–19.

Besides, it makes <u>no</u> difference now to change their feed,
healing arts do harm when applied; their masters withdraw in defeat.
Verg., *G.* 3.548–9

<u>nec</u> requies erat ulla mali: defessa iacebant
corpora, mussabat tacito medicina timore.

<u>Nor</u> did the evil know any respite: their bodies lay exhausted,
physicians reduced to muttering in silent fear.[15] Lucr. 6.1178–9

On the surface, at least, this is a difficult parallel to recognize. Vergil employs only one word used by Lucretius, the common conjunction *nec,* and even this word is in a different metrical position in the line. Both are of course in the same dactylic hexameter, but to single them out as significantly intertextual would require recognition of a shared combination of the specific themes of medicine, plague, and potentially failure, a task not impossible with current methods but demanding greater precision than approaches like semantic analysis have shown to this point.[16]

EXISTING TOOLS

Having understood the basic challenge of identifying intertexts computationally, we can now review how the tools currently available meet the challenge. Basic textual search tools in Greek and Latin remain valuable for allowing users to carry out string (character-sequence) searches including wildcards and Boolean operators. We might, for example, want to find every instance of the Latin stem *habit-* to see if any authors echo others when writing about the idea of "dwelling," whether with the verb *habito*, the noun *habitatio*, or other forms. The *PHI* (Packard Humanities Institute) website offers this sort of search for Latin literature, as does *Perseus* for both Greek and Latin, and the *TLG* (*Thesaurus Linguae Graecae*) *for* Greek.[17] *Diogenes* is a downloadable

---

[15] *Saturnalia* 6.2.13.

[16] For an example of semantic analysis applied for the study of intertextuality, see Scheirer et al. 2016. This example also suggests why the term "intertextuality" may be more useful for conceptualizing the phenomenon under study than the alternative term "text reuse" that is preferred by some digital humanists. "Text reuse" sounds more concrete and seemingly avoids the tricky question of determining just what an "allusion," "reference," or "intertext" is. Taken literally, however, it excludes meaningful forms of intertextuality such as the relationship that Macrobius identifies, which is not an instance of text reuse, at least in the strict sense.

[17] http://latin.packhum.org/search.

program that allows for further searching within versions of the *PHI* and *TLG* databases.[18] Using a large-scale search, the *Proteus Project* has produced a database of possible quotations of classical texts in later literature. This was done through an automatic comparison of classical Latin texts and English translations of Greek texts from the *Perseus Digital Library* with the books in the Internet Archive to produce a list of candidate intertexts.[19]

Newer tools offer searching for more varied features. *Fīlum* allows users to input a particular string for searches in Latin prose and poetry. It returns passages not only with the same exact text, but with text within a defined edit distance of the query text, permitting the discovery of quotations but also passages with identical lemmas and similar sound features.[20]

The *Musisque Deoque* (*MQDQ*) website provides the ability to search for intertexts in Latin poetry.[21] Users can choose an individual line or passage from a poem and retrieve a set of passages with parallel language from either one other work or from the whole corpus. *MQDQ* search takes into account similarity of word forms, their distance in their respective texts, word order, and metrical shape and position. The project offers an extensive, hand-curated corpus of Latin poetry through late antiquity including textual variants, allowing for more accurate and complete search.

The *Tesserae* website allows for search across a large corpus of Greek and Latin prose and poetry, mainly from the *Perseus Project*. Users search one whole text against another to return the most similar passages in the two texts. They can search by exact form, lemma, and sound, with the results ranked by the frequency of the target feature and the proximity of matched words. Users can also employ semantic matching to find passages where words have similar meanings. For example, in a comparison of the poems of Catullus with Vergil's *Georgics*, *tacet nox* (Catullus 7.7) and *silet nox* (*Georgics* 1.247) were returned as matches, both meaning "the night is quiet," although they use two etymologically unrelated words for "is quiet," *tacet* and *silet*. Semantic matching is available for paired Greek texts, paired Latin texts, and paired Greek and Latin texts. This last search type makes it possible for the first time

[18] https://community.dur.ac.uk/p.j.heslin/Software/Diogenes/.

[19] http://books.cs.umass.edu/mellon/index.html.

[20] https://qcrit.org/filum. For further information on Fīlum see Chaudhuri et al. 2015 and Chaudhuri and Dexter 2017.

[21] http://www.mqdq.it/public/cooccorrenze/cooccorrenze. The site has a variety of other features. Mastandrea and Spinazzè 2011 describe the work of *MQDQ* to that point. A companion website, *Pede Certo* (http://www.pedecerto.eu/), allows for full metrical searching of Latin hexameters.

to conduct automatic intertextual search across languages. Users can find, for example, passages of Vergil's *Aeneid* that contain words with similar meanings to those in passages of Homer's *Iliad*.[22]

Finally, *TRACER* is a downloadable software package for detecting intertextuality that accounts for a wide range of language features, including synonyms and syntax. The comprehensive approach makes *TRACER* both powerful and computationally intensive. Users prepare their own texts then employ *TRACER* as a command line tool, choosing from a range of customizable retrieval options and running their detection. The results can then be inspected individually and viewed with *TRACER* visualizations.[23]

These sites take various approaches to the openness of their data and code. *PHI* allows users to search its texts and use its site under "fair use" principles but not download the whole text collection. *Musisque Deoque* allows users to access texts and use search results but does not offer its texts for download or provide its code online. *Diogenes* is free to download and its code can be modified, but it requires the *PHI* and *TLG* databases. *Fīlum* uses public domain texts from the *Tesserae* collection and does not currently provide its code, but may do so in the near future. *TRACER* requires users to ingest their own texts and makes its code publicly available on its site.[24] The *Proteus* Project uses public domain texts and makes its code downloadable, reusable, and alterable. *Tesserae* uses public domain texts transformed into plain texts. Its texts and project code are available on Github and can be freely reused and modified.

A LIFECYCLE OF INTERTEXTUAL STUDY

As we turn to how we can further develop these tools and possibly create others, it will help to review the practical work of scholarship that the tools are designed to aid.

Traditional intertextual reading consists of four activities often blended together:

---

[22] http://tesserae.caset.buffalo.edu/. On general Tesserae search methodology see Forstall et al. 2015. Catullus – Vergil *Georgics* search results available at: http://tesserae.caset.buffalo.edu/cgi-bin/read_bin.pl?session=100017f3. For a blog post on semantic matching see: http://tesserae.caset.buffalo.edu/blog/version-3–1-capturing-more-parallels-and-sorting-them-better/.

[23] TRACER is available at: http://www.etrap.eu/research/tracer/. Description in Büchler et al. 2017 forthcoming. See also the presentation of G. Franzini and M. Büchler: http://www.etrap.eu/wp-content/uploads/2017/08/2017–08–10-montreal-dh2017-orosius.pdf. On visual text analysis, see Jänicke et al. 2016.

[24] The TRACER code is available by registering at: http://vcs.etrap.eu/users/sign_in.

- Discovery
- Interpretation
- Storage
- Representation

The easiest way to understand these stages is by sketching typical scenarios under which intertextual study is carried out. We begin from two common scenarios, as summarized in the first two columns of Table 1.

In Scenario 1, a scholar reading a text recalls another passage with some language similarity (discovery). Both texts are investigated to develop an understanding of how they are related (interpretation). In Scenario 2, the scholar reads a text but rather than recalling another passage, conducts a targeted digital search to find other passages with similar language (discovery) and then develops an understanding of how these passages relate to the original text (interpretation).

From Scenario 1 to Scenario 2, a gap emerges between discovery and interpretation. In Scenario 1, the recollection is triggered by similar words, sounds, etc., in the mind of the scholar, who may also consciously or unconsciously sense some interpretive meaning. Discovery can be followed closely by, or even bound up with, interpretation. In Scenario 2, discovery and interpretation are separated. The scholar first carries out discovery with a targeted search, then begins the process of interpretation.

Under Scenarios 1 and 2, the media for storage and representation are identical. The scholar publishes the intertexts found in an article, book, or commentary where they are stored and represented on the page. The format of the publication is determinative. The parallels are mentioned in the analytical narrative of a book chapter, for example, or gathered under lemmata keyed to individual words in a commentary. Due to limited publication space, and, in the case of articles and books, the need to follow an argumentative thread, some connections the scholar has discovered may not be stored or represented in publicly accessible way but remain unpublished among personal notes.

Scenario 3 is based on the current capacities for unsupervised discovery. In this scenario, the researcher begins by choosing whole texts for automatic comparison, generating a list of potential intertexts. The researcher starts not from apprehension of a passage ("that's an interesting line") but rather from a research question that prompts one to compare texts ("which passages in Apollonius' *Argonautica* share phrasing with Homer's *Iliad*?").

The Scenario 3 procedure introduces a stage of interpretation that precedes discovery. In Scenarios 1 and 2, the scholar may have an interpretive framework in mind when reading, but the scholar engaging in Scenario 3 is compelled to

TABLE 1. INTERTEXTUAL READING SCENARIOS WITH CURRENT AND PROJECTED DIGITAL TOOLS

Scenario 1 Unassisted Reading	Scenario 2 Reading and Targeted Search	Scenario 3 Global Automatic Search	Scenario 4 Search, Repository, Visualization
DISCOVERY AND INTERPRETATION			
Recall of passage while reading. Reflection on parallels and interpretation	DISCOVERY Targeted digital search for other instances of passage found while reading	INTERPRETATION Selection of research query (e.g., which texts to compare)	INTERPRETATION Selection of research query (e.g., which texts to compare)
	INTERPRETATION Reflection on parallels and interpretation	INTERPRETATION Search algorithm for matching selected, representing (minimal) interpretive schema	INTERPRETATION Search algorithm for matching selected, interpretive schema representing (minimal) interpretive schema
		DISCOVERY Unsupervised detection of parallels by algorithm	DISCOVERY Unsupervised detection of parallels by algorithm
		INTERPRETATION Reflection on parallels and interpretation	INTERPRETATION Reflection on parallels and interpretation

	Column 1	Column 2	Column 3	Column 4
STORAGE	Article, commentary, monograph. Some loss of information due to publication formats.	Article, commentary, monograph. Some loss of information due to publication formats.	Search results can be stored online for public access. Curated results stored in article, commentary, monograph. Some loss of information due to publication formats.	Entirety of research results can be stored online for public access in different formats, with no loss of information.
REPRESENTATION	Article, commentary, monograph.	Article, commentary, monograph.	Article, commentary, monograph.	Conclusions can be represented in a variety of digital formats, including graphs. Can also be presented in traditional formats of article, commentary, monograph.

start from at least a minimal interpretive scheme simply by choosing discovery tools, with their inherent assumptions (e.g., start from matching lemmata), and by choosing how to use the tools (e.g., search settings). In Scenario 3, discovery occurs instantaneously when the search is executed. This is followed by another stage of interpretation when the researcher examines the returned parallels and attempts to explain them. At any of these points the researcher will supplement the process with prior knowledge. Storage remains virtually identical with Scenarios 1 and 2, involving publication in articles, books, and commentaries, though queries and results can be archived and made available online for further search and examination.[25]

We can use a final Scenario 4 to sketch out what a next generation inter-textual research environment and process could look like in order to provide one possible agenda for expanding research horizons much farther than digital tools have already done. In order to establish how this Scenario 4 could be plausible, we must first take account of another piece of digital infrastructure that could become important in particular for storage and representation.

To gather and compare intertexts, we need to refer accurately to their tex-tual locations and categorize them in digital terms. One current standard for referring to locations in classical texts is the Canonical Text Services (CTS) protocol.[26] It provides every classical text with its own unique identifier, a Uniform Resource Name (URN), which allows for designating the language of the work, its author, its title, and the particular section, word or words referred to.[27] The Classical Works Knowledge Base (CWKB) has its own protocol of canonical identifiers for authors, works, and parts of works to make it possible to find a queried location in several existing free and paid databases.[28] One research team has recently employed the CTS protocol to provide a standard for describing intertexts. The standard includes CTS URNs for the referring text, the text referred to, and the specific piece of text where the reuse takes place.[29] Challenges remain in simply defining a location in a text, as when we wish to indicate exactly where a thematic resemblance begins and ends.

---

[25] So, for example, the data for Coffee and Forstall 2016 are available at (http://tesserae. caset.buffalo.edu/blog/data-for-claudian-lucan-study/).

[26] http://www.homermultitext.org/hmt-doc/cite/texts/cts.html. The lead authors of the protocol are C. Blackwell and N. Smith. A guide to CTS can be found at Smith and Blackwell 2012.

[27] A set of CTS URNs for classical texts can be found in the Perseus Catalog, http://catalog.perseus.org/.

[28] http://cwkb.org/home.

[29] Berti, Blackwell et al. 2016: 126–127. More precisely: 1. A unique identifier for the instance of intertextuality, 2. A number indicating which instance of the referred to text

More difficult still is the definition of the relationship between two pieces of text. One starting point is provided by the *Sharing Ancient Wisdoms* (*SAWS*) project, which has developed a set of standard descriptors for textual relationships. These include tags for situations when one piece of text is a verbatim repetition of another, one is a shorter or longer version of another, and for loose or close renderings.[30]

Applying such digital standards to describe the locations and types of intertexts would open the door to more easily saving, sharing, and viewing intertexts. Emerging digital editing environments for classical texts, such as the *Digital Latin Library*, *Perseids*, and *Recogito* already allow users to annotate their texts with various forms of embedded information.[31] Intertext standards could permit editors of digital texts to embed intertextual links that not only take readers to the connected text and possibly indicate the type of intertextual relationship, but also enable search across entities tagged as intertexts, in order to answer larger-scale questions.[32] We might ask, for example, whether Ovid draws on the *Aeneid* more or less in the successive books of the *Metamorphoses*, and how the quality of his intertexts change over the course of the work. Provided there was a standard format for designating an intertext, answers to this question could come not only, or not even primarily, from manual annotations of texts and commentaries, but also from automatic search.

If intertexts were published openly in a standard format with unique URNs, it would then be possible to gather and search them in very large numbers. One model for such search is the *Peripleo* website of the *Pelagios* project.[33] *Peripleo* allows users to search across all the *Pleiades* place URNs online. If, for example, anyone has attached the canonical place "Athens" to a textual locus, archeological site report, image of a work of art, or anything else, one can retrieve all this information together and proceed to search it more precisely

---

this is in the referring text, 3. A unique identifier to the locus of the referring text, 4. A string of text used from the text referred to, 5. A unique identifier for the locus of text referred to, and 6. A unique identifier for this intertext among a set of intertexts from the text referred to. They employ Canonical Text Services IDs for the texts, and CITE IDs for the intertext.

[30] Hedges et al. 2017: 7: https://arxiv.org/pdf/1602.00234.pdf. Full set of tags: http://www.ancientwisdoms.ac.uk/media/ontology/sawsOntology.owl.

[31] http://digitallatin.org/, http://sites.tufts.edu/perseids/, http://recogito.pelagios.org.

[32] Existing work on large-scale intertextual trends among whole individual Latin works or parts of the classical Latin corpus includes Bernstein et al. 2015 and Burns 2017.

[33] http://pelagios.org/peripleo/map.

and effectively. One could imagine a similar search for all published intertexts that could answer questions about trends in intertextuality across all of Greek, Latin, or a combination of languages. The new corpus of marked intertexts would also provide a vastly expanded benchmark set for training machine learning algorithms to better detect intertextuality. If the intertext standard included information about the scholar or scholars recording the intertext, using standard ORCIDs to identify researchers, then an intertextual search service could also produce something like a registry for intertexts, which could also function as a form of micro-publication.[34]

Digital standards for intertexts can also aid in visualizing intertextuality. *TRACER* is the one intertextual search engine that already offers its own visualization tool, TRAViz.[35] Figure 1 is an example of one visualization available from *TRACER*. It illustrates language shared between English editions of the Gospels of Mark and Luke. The first image is a dotplot of the phrases that overlap, with darker circles showing greater similarity. The second image gives a list of the parallel phrases. The final image is produced when clicking on one of the dots in the dotplot, and gives an in-line representation of the similarities and differences across one pair of textual loci, which in this case happen to be highly similar.[36]

The development and application of common standards would allow visualizations like these to be created independently of any search engine or editing environment, so that properly formatted intertexts from any source could be visualized.

AN AGENDA FOR THE STUDY OF INTERTEXTUALITY

Having considered these resources, we are now prepared to trace one possible path forward from here. In Scenario 4, a scholar beginning research might start from an overall network visualization of intertextual relations across the Greek and Latin languages, based on a weighted suite of similarity measures for lemma, meter, sound, syntax and other features, consisting of existing scholarly annotations and others newly highlighted by the search algorithm. The scholar might then choose to highlight connections by genre and foreground the interaction between the Greek elegiac poets and the poets of Latin elegy. A change of views would allow for inspection of individual poems with

---

[34] https://orcid.org/.

[35] http://www.traviz.vizcovery.org. For more, see Jänicke et al. 2015.

[36] Image from: http://www.informatik.uni-leipzig.de:8080/BibleViz/.

Figures 1A-1C. TRAViz visualization of a comparison of the Gospels of Mark and Luke.
Credit: Copyright (C) 2015, Stefan Jänicke. From: http://www.vizcovery.org/ Jänicke, S, Gessner, A., Büchler M. and Scheuermann, G. 2014. "Visualizations for Text Re-use." In Proceedings of the 5th International Conference on Information Visualization Theory and Applications, IVAPP 2014: 59-70.
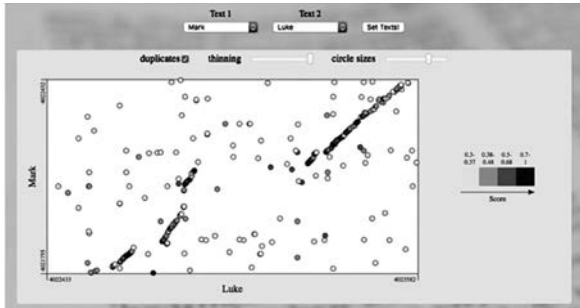

Figure 1A. A dotplot of overlapping phrases.
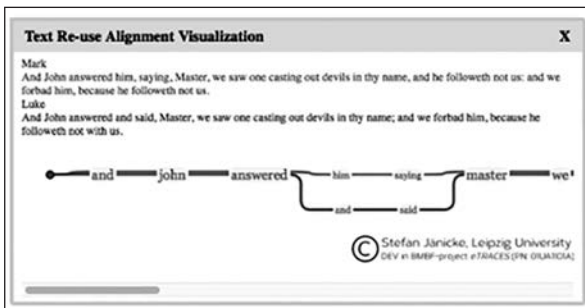

Figure 1B. A list of parallel phrases.

Figure 1C. An in-line representation of the similarities and differences across one pair of textual loci.

the strongest connections, proceeding down to the phrase level. Meaningful phrases could then be sifted out, to compare how language and sentiments from the Greek elegists resonated differently in Propertius, Tibullus, and Ovid. To understand these resonances better, it would then be possible to check for similar thematic material in other Augustan poetry and prose, as well as epigraphic material. A search forward into the later literary tradition would show where else in the later western literary corpus these themes took root. To complete the research, the scholar could select and review all the significant intertexts found, then record them publicly, with any added annotations, creating a set of micro-publications. The scholar could then proceed to write up an article on the findings to be published in an open access journal, where each intertext was tagged with a unique identifier that associated it with the other intertexts published by the same scholar. Other researchers could then access this work through the article, through a search for the scholar's annotations, searching or browsing the relevant texts, or visualizations created from the annotated intertexts.

Readers could then incorporate the scholar's notes into their reading experience. As they scrolled through the text of Ovid's *Amores*, sections of related text would appear and disappear beside the passage in focus. They could filter out later texts that adapt the *Amores* to concentrate on earlier texts that Ovid used. They could further filter out Latin sources for the *Amores* in favor of Greek ones. They would have at their disposal a short list of Greek sources for their *Amores* passage, all with translations. Some would have flags showing they were scholarly annotations, making it possible to view only the sources manually attested as most significant. In this way, a modern reader could simulate the experience of an ancient one, but customize the experience to answer a particular question: what are the Greek sources behind this passage of the *Amores*?[37]

If scholars and digital humanists can further develop the potential of existing technologies, they will be able to realize a vision of intertextual research like this one, as part of a new age of digital philology, and so advance much closer toward a deep understanding of how classical texts were composed and connected.

---

[37] This paragraph adapts a scenario envisioned by James Gawley. For a parallel vision of digitizing existing classical commentaries so that all their notes on passages could be compared at once, see Heslin 2016. M. Romanello and JSTOR have already produced a tool allowing users to click on lines of the *Aeneid* and bring up articles in JSTOR that refer to that line: http://www.aeneidinjstor.eu.

WORKS CITED

Almas, B. and Berti, M. 2013. "Perseids Collaborative Platform for Annotating Text Re-Uses of Fragmentary Authors." In *DH-Case 2013. Proceedings of the 1st International Workshop on Collaborative Annotations in a Shared Environments: Metadata, Vocabularies and Techniques in the Digital Humanities.* ACM Proceedings.

Bamman, D. 2014. "Intertextuality beyond Words." *Tesserae Project Blog* http://tesserae. caset.buffalo.edu/blog/intertextuality-beyond-words-by-david-bamman/.

Bamman, D. and Crane, G. 2011. "The Ancient Greek and Latin Dependency Treebanks." In Sporleder, C., Bosch, A. and Zervanou, K. eds. *Language Technology for Cultural Heritage Selected Papers from the LaTeCH Workshop Series.* Berlin, Springer-Verlag: 79–98, http://dx.doi.org/10.1007/978–3–642–20227–8.

Bamman, D., O'Connor, B. and Smith. N. 2013. "Learning Latent Personas of Film Characters." *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics:* (Volume 1: Long Papers), pages 352–61. Sofia, Bulgaria: Association for Computational Linguistics.

Bernstein, N. W., Gervais, K. and Lin, W. 2015. "Comparative Rates of Text Reuse in Classical Latin Hexameter Poetry." *Digital Humanities Quarterly* 9(3). http://www. digitalhumanities.org/dhq/vol/9/3/000237/000237.html.

Berti, M. 2013. "Collecting Quotations by Topic: Degrees of Preservation and Transtextual Relations among Genres." *Ancient Society* 43: 269–88.

Berti, M., Blackwell, C.W, Daniels, M., Strickland, S. and Vincent-Dobbins, K. 2016. "Documenting Homeric Text-Reuse in the *Deipnosophistae* of Athenaeus of Naucratis." *BICS* 59(2): 121–39.

Berti, M., Romanello, M., Babeau, A. and Crane, G. 2009. "Collecting Fragmentary Authors in a Digital Library." *Proceedings of the 2009 Joint International Conference on Digital Libraries (JCDL '09).* New York: ACM Digital Library. 259–62.

Büchler, M., Franzini, G., Franzini, E. and Bulert, K. Forthcoming 2017. "TRACER - a multilevel framework for historical Text Reuse detection." *Journal of Data Mining and Digital Humanities - Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages.*

Burns, P. 2017. "Measuring and Mapping Intergeneric Allusion in Latin Poetry using Tesserae" *Journal of Data Mining and Digital Humanities - Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages.* https://jdmdh.episciences. org/3821.

Castelletti, C. 2014. "Aratus and the Aratean Tradition in Valerius' *Argonautica*." In Augoustakis, A. ed. *Flavian Poetry and its Greek Past.* Leiden: Brill. 49–72.

Chaudhuri, P., Dexter, J. P. and Bonilla-Lopez, J. A. 2015. "Strings, Triangles, and Go-Betweens: Intertextual Approaches to Silius' Carthaginian Debates." *Dictynna* 12. https://dictynna.revues.org/1156

Chaudhuri, P. and Dexter, J. P. 2017. "Bioinformatics and Classical Literary Study." *Journal of Data Mining and Digital Humanities - Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages.* https://jdmdh.episciences.org/3807.

Coffee, N. 2012. "Intertextuality in Latin Poetry." In Clayman, D. ed. *Oxford Bibliographies in Classics.* New York, Oxford University Press.

Coffee, N. and Forstall, C. 2016. "Claudian's Engagement with Lucan in his Historical and Mythological Hexameters." In Berlincourt, V., Galli-Milic, L. and Nelis, D. eds. *Lucan and Claudian: Context and Intertext*. Heidelberg: Universitätsverlag Winter. 255–84.

Coffee, N., Koenig, J.-P., Poornima, S., Ossewarde, R., Forstall C. and Jacobson, S. 2012. "Intertextuality in the Digital Age." *TAPA* 142(2): 381–419.

Dexter, J. P., Katz, T., Tripuraneni, N. , Dasgupta, T., Kannan, A., Brofos, J. A., Bonilla Lopez, J. A., Schroeder, L. A., Casarez, A, Rabinovich, M., Haimson Lushkov, A. and Chaudhuri, P. 2017. "Quantitative Criticism of Literary Relationships." *Proceedings of the National Academy of Sciences* 114(16): E3195-E3204.

Erlich, V. 1980. *Russian Formalism: History, Doctrine*. The Hague: Mouton.

Farrell, J. 2005. "Intention and Intertext." *Phoenix* 59: 98–111.

Forstall, C., Coffee, N., Buck, T., Roache, K. and Jacobson, S. 2015. "Modeling the Scholars: Detecting Intertextuality through Enhanced Word-Level N-Gram Matching." *Literary and Linguistic Computing* 10.1093/llc/fqu01.

Fowler, D. 1997. "On the Shoulders of Giants: Intertextuality and Classical Studies." *MD* 39: 13–34.

Gorman, V.B. and Gorman, R.J. 2016. "Approaching Questions of Text Reuse in Ancient Greek Using Computational Syntactic Stylometry." *Open Linguistics* 2: 500–10.

Griffin, J. 1986. *Latin Poets and Roman life*. Chapel Hill: University of North Carolina Press.

Hedges, M., Jordanous, A., Lawrence, K. F. , Roueché, C. and Tupman, C. 2017. "Computer-Assisted Processing of Intertextuality in Ancient Languages." *Journal of Data Mining and Digital Humanities - Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages*. https://jdmdh.episciences.org/3905.

Heinze, R. 1993. *Virgil's Epic Technique*. Berkeley: University of California Press.

Heslin, P. 2016. "The Dream of a Universal Variorum: Digitizing the Commentary Tradition." In Kraus, C. S. and Stray, C. eds. *Classical Commentaries: Explorations in a Scholarly Genre*. Oxford: Oxford University Press. 494–511.

Hinds, S. 1998. *Allusion and Intertext: Dynamics of Appropriation in Roman Poetry*. Cambridge: Cambridge University Press.

Jakobson, R. 1960. "Closing Statement: Linguistics and Poetics." In Sebeok, T. ed. *Style in Language*. Cambridge, MA: MIT Press: 350–77.

Jänicke, S., Efer, T., Büchler, M. and Scheuermann, G. 2015. "Designing Close and Distant Reading Visualizations for Text Re-use." In Ranchordas, A., Madeiras Pereira, J., Araújo, H. J. and Tavares, J. eds. *Computer Vision, Imaging and Computer Graphics: Theory and Applications*. Berlin: Springer. 153–71.

Jänicke, S., Franzini, G., Cheema, M.F. and Scheuermann, G. 2017. "Visual Text Analysis in Digital Humanities." *Computer Graphics Forum* 36: 226–250. doi:10.1111/cgf.12873.

Mastandrea, P. and Spinazzè, L. eds. 2011. *Nuovi archivi e mezzi d'analisi per i testi poetici. I lavori del progetto Musisque Deoque Venezia 21–23 giugno 2010*. Amsterdam: Hakkert.

McKeown, J. C. 1987. *Ovid, Amores: Text, Prolegomena and Commentary in Four Volumes*. Liverpool: F. Cairns.

Scheirer, W., Forstall, C., and Coffee, N. 2016. "The Sense of a Connection: Automatic Tracing of Intertextuality by Meaning." *Digital Scholarship in the Humanities* 31: 204–17.

Smith, D.N. and Blackwell, C.W. 2012. "Four URLs, Limitless Apps: Separation of Concerns in the *Homer Multitext* Architecture." *A Virtual Birthday Gift Presented to Gregory Nagy on Turning Seventy by His Students, Colleagues, and Friends*. Washington DC: Center for Hellenic Studies. http://chs.harvard.edu/CHS/article/display/4846.

Wagner, R. A. and Fischer, M. J. 1974. "The String-to-String Correction Problem." *Journal of the ACM* 21(1): 168–73.