

NOTE: THIS IS A PRE-PRINT DRAFT VERSION OF: “The Tesserae Project: intertextual analysis of Latin poetry” published in *Literary and Linguistic Computing*. The published version contains several editorial changes. Interested readers are advised to consult the published version of this paper:

Neil Coffee, Jean-Pierre Koenig, Shakthi Poornima, Christopher W. Forstall, Roelant Ossewaarde and Sarah L. Jacobson “The Tesserae Project: intertextual analysis of Latin poetry.” *Literary and Linguistic Computing* (2012)

doi: 10.1093/llc/fqs033

First published online: July 20, 2012

<http://llc.oxfordjournals.org/content/early/2012/07/20/llc.fqs033.abstract>

The Tesserae Project: Intertextual Analysis of Latin Poetry**Neil Coffee**

The University at Buffalo, The State University of New York

Jean-Pierre Koenig

The University at Buffalo, The State University of New York

Shakthi Poornima

The University at Buffalo, The State University of New York

Christopher W. Forstall

The University at Buffalo, The State University of New York

Roelant Ossewaarde

The University at Buffalo, The State University of New York

Sarah L. Jacobson

The University at Buffalo, The State University of New York

Correspondence:

Neil Coffee

Department of Classics

338 MFAC

University at Buffalo

Buffalo NY 14261-0026

USA

ncoffee@buffalo.edu

Abstract

Tesserae is a web-based tool for automatically detecting allusions in Latin poetry. Although still in the start-up phase, it already is capable of identifying significant numbers of known allusions, as well as similar numbers of allusions previously unnoticed by scholars. In this paper we use the tool to examine allusions to Vergil's *Aeneid* in the first book of Lucan's *Civil War*.

Approximately 3,000 linguistic parallels returned by the program were compared with a list of known allusions drawn from commentaries. Each was examined individually and graded for its literary significance, in order to benchmark the program's performance. All allusions from the program and commentaries were then pooled in order to examine broad patterns in Lucan's allusive techniques which were largely unapproachable without digital methods. While Lucan draws relatively constantly from Vergil's generic language in order to maintain the epic idiom, this baseline is punctuated by clusters of pointed allusions, in which Lucan frequently subverts or distorts Vergil's original meaning. These clusters not only attend the most significant characters and events, but also play a role in structuring scene transitions. We are working to incorporate the ability to match on word meaning, phrase context, as well as metrical and phonological features, into future versions of the program.

Introduction

The study of allusion has grown to become a core interest of classical—particularly Latin—literary studies over the past several decades. Beyond simply documenting instances of textual reuse, scholars such as Conte (1986), Hinds (1998), and Edmunds (2001) have enlarged the scope in which allusion is understood to create meaning, and presented several theoretical models for how allusion is both written and read.

A number of recent digital humanities projects have examined various aspects of text reuse. Bamman and Crane (2008) presented a model for identifying allusions based on multiple parameters, and detailed their methods for measuring two texts' similarity by words, word order, and syntax. Horton *et al.* (2010) created an algorithm for detecting text reuse in French and other languages, based solely on string similarity, which they have released under an open source licence. Büchler *et al.* (2010) examined larger-scale patterns of text reuse in the treatment of Plato by later Greek authors. Tesseract draws on these and other projects for models, yet distinguishes itself as an integrated effort to develop allusion detection software, undertake detailed case-studies, and bring the understanding of allusion to a non-specialist audience.

The current version of Tesseract identifies similar passages in two Latin poems based on word similarity: two passages constitute a parallel if they share two or more words. Word similarity is judged not only by the word's form in the text, but also by its dictionary headword. We used the Archimedes Morphology Service of the Max Planck Institute for the History of Science (<http://archimedes.mpiwg-berlin.mpg.de/arch/doc/xml-rpc.html>) to retrieve headword information for our texts. Texts themselves were drawn from the Latin Library (<http://thelatinlibrary.com>) and the Perseus Project (<http://www.perseus.tufts.edu>).

Users of Tesseract's web-based interface select two texts from simple drop-down lists

(Fig. 1). A list of parallel phrases is then returned (Fig. 2); this may be downloaded as an XML document or a list of comma separated values. The current version of the program is already online and freely accessible (<http://tesserae.caset.buffalo.edu>), and has received positive feedback from practising scholars of Latin allusion, including writers of textual commentaries who customarily note allusions.

In the remainder of this article we present some preliminary results from our application of the current version of the search tool to a case study of the Roman poet Lucan. Lucan was a poet of the time of Nero and left unfinished at his death an 8,000-line epic on the subject of Rome's civil war known as the *Bellum Civile*. In writing such an epic, it would have been impossible for Lucan to avoid comparison with the figure of Vergil, approximately 100 years his senior, whose monumental work, the *Aeneid*, had already become a classic. Lucan's relationship with his predecessor is far from simple: at times he relies on and reinforces Vergil's authority; at times he draws out ambiguity and paradox latent in Vergil's work; at times he deliberately opposes Vergil's artistic and ideological programs.

We formulated five questions to frame our analysis:

1. How often does Lucan refer to the *Aeneid*?
2. What kinds of reference does he make?
3. Where in the *Aeneid* does he turn most often, and for what kinds of references?
4. How are these references distributed within Lucan's text?
5. How do these results change our present understanding of the relationship between the *Bellum Civile* and the *Aeneid*?

Method

A list of parallels between the *Bellum Civile* (*BC*) and the *Aeneid* is where we begin to examine the contact between Lucan and Vergil. We concentrated our attention on *BC* book 1 (695 lines), considering parallels found anywhere in the entirety of the *Aeneid* (9,896 lines). We ran Tesseract on these texts, then compared the results with a list of parallels collated from four modern commentaries: Heitland and Haskins (1887), Thompson and Bruère (1968), Viansino (1995), and Roche (2009).

Each parallel identified either by the program or by the commentators was examined individually and given a type number between 1 and 5 according to its literary significance. While this was necessarily a subjective procedure, we formulated a general set of criteria for our classification (Table 1). The principal distinction was between meaningful (type 3–5) and not meaningful (1–2) parallels. This distinction follows the argument of Thomas (1986, 117) that references either are or are not “susceptible to interpretation or meaningful.” The set of meaningful parallels was further divided into those that simply reused distinctive language, and those that in doing so created new literary significance. Conte (1986, 31) proposed that an earlier work could provide either a *code model* or an *exemplary model* for a later one. In the first case, the model as a whole defines the idiom in which the later text speaks. In the second case, the referring author directs the reader’s attention to a particular moment in the earlier work. This distinction separates our type 3 from types 4–5. The final distinction, between types 4 and 5, less- and more significant allusions, was the most subjective. Other schemas are possible, but ours proved useful for broadly categorizing parallels to analyze the large-scale questions posed above, to which we now turn.

Results

Numbers of Parallels

The automated search returned a list of 3100 parallels across all types, while the combined efforts of the four commentaries produced 419 parallels of types 2–5. A comparison of results by type is given in Table 2. The number of type 3–5 parallels returned by the program was comparable to the work of the commentators, but the program reported vastly more of types 1–2 than did the commentaries. These results show that, with manual examination of the program’s output to filter out false positives, our automated search can already identify a significant portion of the parallels most interesting to literary scholars. Comparing the program to individual commentaries, we see that for interpretable allusions (types 4–5), it reports 103 to Viansino’s 48, but still fewer than Roche’s 151.

These numbers tell only half the story, however. While *Tesseractae* returned numbers of valuable parallels at similar rates to the commentators, the parallels themselves were often different from those found by the commentators. Only half of the interpretable allusions detected by *Tesseractae* were listed in the commentaries (Fig. 3). Thus although *Tesseractae* returned only 25% of the commentator’s allusions, it also increased the total number of allusions found by 25%.

Parallels by Type

The most obvious difference between our automated search and the commentaries was the number of less-meaningful parallels returned. Among the commentaries there is already a trend in this direction, with Roche (2009) surpassing his predecessors in the number of type 2–3 parallels reported. Unlike the other commentators, Roche examined only Book 1 of Lucan’s poem, effectively concentrating his efforts. He also used digital searches along with more traditional philological tools. These methods enabled Roche to look beyond the exemplary model

allusions most familiar to Latinists, and begin to represent the level of code model reference which underwrites Lucan's posture as an epic poet. *Tesserae* expands this perspective considerably.

In what proportions does Lucan use the various types of parallels? Combining results from *Tesserae* and the commentators, we start to get a comprehensive picture of the author's practice. The data presented in Table 2 suggest that in *BC* 1, Lucan relies on Vergil's generic epic language about twice as frequently as he alludes to specific passages in the *Aeneid*.

Parallels by Location in Source Text

Lucan does not draw evenly from all books of the *Aeneid*. Fig. 4 shows the distribution of all parallels in the *Aeneid*, by type. While Lucan draws relatively evenly on all books of the *Aeneid* for type 3 parallels, he clearly favors certain books for types 4–5. His most meaningful allusions are drawn above all from *Aeneid* 2, followed by books 4, 11, and 3.

It is natural that, in presenting the destruction of Rome as the major theme of the Roman civil war, Lucan should draw upon Vergil's portrayal of the fall of Troy in *Aeneid* book 2. *Aeneid* 11 describes hard fighting and internal conflict in the Latin assembly, and is also thematically apropos. The choice of *Aeneid* 4, the story of Dido, is less obvious. While Lucan uses material from this book for several purposes, a significant complex of allusions borrow notions of madness and ill rumor from the Dido story to suggest ill-starred similarities between Carthage and Rome. Thus *BC* 1.676, *attonitam rapitur matrona per urbem* ("The [prophetic] matron is swept through the awestruck city"), draws on *Aeneid* 4.666, *concussam bacchatur Fama per urbem* ("Rumor runs riot through the stunned city") to suggest that Romans of the civil war period were as mad and rumor-driven as Dido and her Carthaginians.

The wealth of allusions to the less-studied book 3 of the *Aeneid* gives further clues to Lucan's unique reading of Vergil. One significant strand of Lucan's use of this book involves reversing its optimistic prophecies of a new land for the Trojans in order to suggest the woeful future in store for the Romans. Thus, in a parallel identified only by Tesseræ, Vergil uses the image of Sicily's separation from Italy at the straits of Messina to foretell Aeneas' successful journey to found Rome (*Aeneid* 3.418), an image Lucan recalls and reverses when he depicts Sicily rejoined to Italy in an eruption of Mt. Aetna as a portent of the coming war (*BC* 1.547).

Parallels by Location in the Referring Text

We combined the automated results with those collated from commentaries to ask what large-scale patterns could be seen in Lucan's use of allusion within his own poem. Fig. 5 shows type 3–5 parallels by location in *BC* 1. Again, the baseline of code model references is relatively constant, punctuated by clusters of more significant allusions.

Lucan clusters significant references throughout the opening and closing sections, and in establishing the principal characters: at the outset of Lucan's text, where he sets out his theme and the artistic program for the work; in the opening descriptions of Caesar and Pompey, the principal belligerents; and in the prophecy of the matrona, which closes the book. In contrast, at the heart of Lucan's praise of Nero (lines 39-59), we find a pause in references to the *Aeneid*. Here, Lucan forgoes an obvious opportunity to ennoble Nero by association with the grandeur of the epic tradition, and instead creates a prosaic tone that flattens what should be the culmination of his praise.

Consideration of large-scale patterns also reveals how Lucan uses allusion to structure his narrative. He shows a tendency to cluster references at the beginning and ending of sections.

More specifically, he often closes a section with a Vergilian allusion, capped by his own pithy or moralizing statement. The next scene then opens with a fresh allusion to anchor and authorize it in the Vergilian tradition. Thus, in the transition from Rome's decline to Caesar's delay at the Rubicon (*Bellum Civile* 1.178–205), Lucan describes the prevalence of bribery (1.178) using language from Vergil's depiction of sinners in the underworld (*Aeneid* 6.622). He closes the scene with his own vision of avidity leading to war (1.82), before opening his section on Caesar's march with several new references to the *Aeneid*. Lucan draws on Vergil's authority to bring density of meaning to his transitions, yet he reserves the crucial end of the section to finish with his own master strokes.

Lucan's Bellum Civile 1 and the Aeneid

Do the results of automated allusion detection change our understanding of Lucan's relationship to the *Aeneid*? A full answer to this question will require analysis of the remaining books of Lucan's epic, but our results provide some initial responses. In existing scholarship, Lucan's references to the *Aeneid* have generally been taken as oppositional, subverting the imagery and language of Rome's founding to suggest that the construction of empire inevitably becomes a corrupt enterprise. Our study supports this picture, but also adds important detail. The constancy of type 3 parallels shows to what degree Lucan relied on Vergil even for the basic idiom of epic. At the same time, Lucan uses allusions to frame scenes, employs clusters of allusions to different themes within Vergil's poem, and shifts markedly from allusions to the *Aeneid* in favor of allusions to other works in his praise of Nero. These gestures all represent distinctive patterns in Lucan's large-scale use of meaningful allusions for artistic effect.

Future Work

The process of evaluating each of the 3000 results collected by the Tesseract program and commentators has created a benchmark set of parallels, including positive and negative examples, for training and testing future algorithms. It has also given us insight into which new feature sets would allow us to capture the greatest number of allusions currently missed by the program. Among these are the ability to match synonyms and the ability to match paragraph-level context, both of which seem to turn on semantics. Although sound-based allusions were not prevalent among the current test set, other examples have convinced us that the ability to match on character-level similarities and metrical shape would bring in additional high-grade allusions. As such feature sensitivity is incorporated, automatic detection of allusion, and of style and theme generally, will increasingly come to replicate the results of traditional scholarship and open up further new perspectives on literary meaning and artistry.

Tables

Table 1 The schema used to grade parallels reported by Tesseract and the commentaries

Meaningful		Not Meaningful		
Interpretable		Not Interpretable		
More significant	Less significant			
5	4	3	2	1
<ul style="list-style-type: none"> ⤴ High formal similarity in analogous context. 	<ul style="list-style-type: none"> ⤴ Moderate formal similarity in analogous context, or ⤴ High formal similarity in moderately analogous context. 	<ul style="list-style-type: none"> ⤴ High / moderate formal similarity with very common phrase or words, or ⤴ High / moderate formal similarity with no analogous context, or ⤴ Moderate formal similarity with moderate / highly analogous context. 	<ul style="list-style-type: none"> ⤴ Very common words in very common phrase, or ⤴ Words too distant to form a phrase. 	<ul style="list-style-type: none"> ⤴ Error in discovery algorithm, words should not have matched.

Table 2 All parallels reported by Tesseract and four commentaries, by type. The commentaries used were Roche (2009), Viansino (1995), Thompson and Bruère (1968) and Heitland and Haskings (1887). In adding columns, each unique parallel is only counted once; combined totals may be less than the sum of individual values.

Type	Tesseract	Commentaries					Total
		all	Roche	Viansino	T & B	H & H	
1	486	0	0	0	0	0	486
2	2241	55	50	8	1	1	2289
3	280	192	168	33	13	6	425
4	57	79	66	18	12	3	115
5	36	93	85	30	14	4	103
Total	3100	419	369	89	40	14	3418

Figures

Fig. 1 Tesserae user interface

Fig. 2 Tesserae results

Fig. 3 Type 4–5 parallels reported by Tesserae and four commentaries. Tesserae returned a significant number of matches overlooked by the commentaries.

Fig. 4 All type 3–5 parallels by book in the *Aeneid*

Fig. 5 All type 3–5 parallels by line in the *Bellum Civile*, book 1

Funding

This work was supported by funding from the Digital Humanities Initiative at Buffalo for its Textual Analysis Working Group; and from the University at Buffalo's Department of Classics.

References

- Bamman, D. and Crane, G.** (2008). The Logic and Discovery of Textual Allusion, *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, Marrakesh, Morocco.
- Büchler, M., Geßner, A., Eckart, T., and Heyer, G.** (2010). Unsupervised Detection and Visualisation of Textual Reuse on Ancient Greek Texts, *Proceedings of the Chicago Colloquium on Digital Humanities and Computer Science* 1(2).
- Conte, G. B.** (1986). *The Rhetoric of Imitation: Genre and Poetic Memory in Virgil and Other Latin Poets*. Cornell University Press, Ithaca, NY.
- Edmunds, L.** (2001). *Intertextuality and the Reading of Roman Poetry*. Johns Hopkins University Press, Baltimore, MD.
- Heitland, W. E. and Haskins, C. E.** (1887). *M. Annaei Lucani Pharsalia*. G. Bell, London.
- Hinds, S.** (1998). *Allusion and Intertext: The Dynamics of Appropriation in Roman Poetry*. Cambridge University Press, New York.
- Roche, P.** (2009). *Lucan: De Bello Civili: Book 1*. Oxford University Press, Oxford.
- Thomas, R. F.** (1986). Virgil's Georgics and the Art of Reference. *Harvard Studies in Classical Philology* 90: 171–198.
- Thompson, L. and Bruère, R. T.** (1968). Lucan's Use of Vergilian Reminiscence. *Classical Philology* 63: 1–21.