The Tesserae Project: Intertextual Analysis of Latin Poetry

Neil Coffee¹, J.-P. Koenig², Shakhti Poornima², Christopher Forstall¹, and Roelant Ossewaarde²

¹. Department of Classics, 2. Department of Linguistics, State University of New York at Buffalo

Project Goals
Free Online Intertextual search
The Tesserae Project is an effort to use computational means to better understand literary style and meaning. Its current focus is intertextuality (text reuse) in classical Latin poetry. The question of how authors make use of and refer to the works of their predecessors and contemporaries has been of considerable interest to ancient and contemporary scholars of Latin literature. Existing digital humanities work on text reuse has tended to concentrate on authorship attribution and the identification of explicit quotations. Recent research in Latin literature has laid the groundwork for further analysis of intertextuality as a matter of style and meaning by identifying five computationally tractable feature sets: meter, semantics, syntax, word form, and word order. The Tesserae group has built upon this research to offer a web tool (http://tesserae.csau.buffalo.edu) that allows for the identification of two-word parallels in two Latin poetic texts that are similar in one of these features, word form, and word order. Test results from the tool against scholarly commentaries have demonstrated the value of this approach and yielded new observations about the relationship of the two Roman poets considered in the test, Vergil and Lucan.

Search Process
Tesserae web interface: basic search page
The Tesserae web tool currently allows for the search for similar two-word phrases in two texts from a corpus of classical Latin poetry. Users can search using one of two algorithms. They begin by selecting two authors to compare. Words accounting for the top 30% most frequent in the Tesserae corpus are excluded from the search (to avoid, e.g. Latin et . . . et, and . . . and). The remaining search process depends upon the approach the user chooses.

Word form Identity
Text A is examined six words at a time to see if any form a pre-existing list of words from Text B are in the six word windows. If two or more similar words are found in the sentences, an entry is added. Results are delivered to the user as a list of two-word phrases for inspection.

Stem (Headword) Identity
Text A is examined whole sentences at a time to see if any form a pre-existing list of dictionary headwords from Text B are in the sentences. If two or more similar headwords are found in the sentences within specified distances, an entry is added.

Conclusions from Comparison with Existing Methods
Testing shows that combined searching with both Tesserae algorithms can identify parallels found by commentators, as well as parallels not yet found even in two well-studied texts. Further work is needed to account for other feature sets, such as meter and semantics, that will allow for capture of more parallels with greater precision. Future results will be tested against the compiled gold standard of commentator parallels and past Tesserae results.

Investigating Large-Scale Intertextuality
Existing Tesserae search and future improvements permit scholars for the first time to view comprehensively the intertextual relationship of large sections of works and whole works. The combination of Tesserae and commentator results provides an example of the new types of observations we can make from this broader perspective.

What Types of References Does an Author Make?
Detected parallels can be classed into those that are interpretable (i.e. generate meaning through analogous contexts) and those that are linguistic borrowings (i.e. where non-analogous or generic contexts show they are just reuse of language). In the case of 643 parallels found in the Lucan 1 – Aeneid comparison, 425 were judged as linguistic borrowings and 218 as interpretable. Lucan was clearly both immersed in Vergilian epic language and interested in making a significant number of meaningful parallels. Future comparison of these results with other books of Lucan’s epic can show how Lucan varied his text reuse over the course of his epic. Comparison with other authors will show differences in poetic technique.

What Parts of Another Work Does an Author Refer to?
Comprehensive assembly of parallels allows us to identify which parts of another work poets refer to most and least frequently, and so develop a picture of what passages they found most useful.

Results Compared with Traditional Methods
Separate tests were performed of word form and stem matches to assess the results of the algorithms. The tests examined potential references in the first book of the Roman poet Lucan’s unfinished 10-book epic On the Civil War (De bello civili or BC) to the Aeneid of his predecessor Vergil. These works were chosen as large corpora and with full commentaries for comparison with Tesserae results. Results were categorized into those that were just linguistic borrowings, and those that were judged to constitute significant references.

Tesserae (word and stem)   Commentators

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>279</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td></td>
</tr>
</tbody>
</table>

Total Unique Parallels found for Lucan Civil War Bk. 1 and Vergil Aeneid: 643

References