Evaluating the literary significance of text re-use in Latin poetry

James Gawley, Christopher Forstall, and Neil Coffee
The University at Buffalo, Buffalo NY.

Abstract for a poster presentation

The Tesserae Project is a collaborative, interdisciplinary effort to create software which can automatically detect literary allusions, with the aim of aiding researchers in the study of *intertext*, that is, the range of ways in which one literary text can refer to another. The project's mission is to automatically and quickly detect allusions across a large number of texts, and at the same time to develop a quantitative approach to modeling the literary process of intertextual reference. Our study corpus is made up primarily of Classical Latin poetry, but in theory the principles used should be broadly applicable. The current implementation of the software[1] successfully detects many interesting allusions based on realtively simple text re-use criteria, but includes a large number of false positives–instances of text re-use which have no literary significance.

The goal of the research presented here was twofold: to reduce the number of false or uninteresting allusions returned by Tesserae, and to sort the remaining results according to their significance to readers. We tested our results against two semi-independent benchmark sets: (1) a collection of professional commentaries; and (2) subjective ranking of literary significance by a group of Classics graduate students and faculty.

The ability to filter and score results will ultimately improve the performance and utility of the Tesserae search tool for the researcher into intertextuality. In addition, because the automated scoring system is being calibrated by and tested against the quality of allusions as assessed by professional commentators and expert annotators, it can also be claimed as a quantitative model of the subjective, literary art of intertextuality.

Benchmarks and Testing

All tests were performed on two Classical Latin texts: Book 1 of Lucan's *Bellum Civile*, an epic poem on the Roman Civil War written in the mid-first century CE, was checked for allusions to Vergil's *Aeneid*, an epic on the founding of Rome, which was written approximately 90 years earlier and had become an instant classic. The set of *commentator* allusions was drawn from four professional commentaries on Lucan[1, 2, 3, 4], and comprised approximately 300 distinct allusions.

The *annotator* benchmark was made up of approximately 3300 distinct cases of text re-use, gathered using two previous versions of Tesserae, and including all the commentator parallels. These were ranked by a graduate seminar on Latin intertextuality, working in small teams, on a scale of 1–5, where type 1 is least significant and type 5, most. While ranking was a subjective measure, it was agreed by all annotators that types below 3 were uninteresting from a literary perspective. It should be noted that the annotators were not aware of which allusions were in the

---

[1]Available at http://tesserae.caset.buffalo.edu.

Table 1: Recall data for each of the 5 annotator-assigned ranks (5=best) and for the commentator allusions, using a default Tesserae search without the scoring/filtering system.

type	tesserae returned	benchmark had	recall rate
1	377	491	0.77
2	1926	2280	0.84
3	316	396	0.80
4	72	96	0.75
5	46	70	0.66
commentators	205	336	0.61

commentator set, and that commentator allusions received a range of annotator rankings; this is what makes the two benchmarks at least semi-independent.

Tesserae identifies text re-use on the basis of two or more shared words in a pair of sentences, allowing differently inflected forms of the same dictionary word to consitute a match. A stop-list of the most frequent words is omitted. Table 1 gives recall rates for each of the 5 annotator types, as well as for the commentator set, for a default search.

Filtering / Scoring

A filter was created using two criteria: the presence of low-frequency words (where the cutoff was determined experimentally) and that of exact-form matches (as opposed to differently inflected forms of the same word).

- Textual parallels containing no low-frequency words were dropped; except,

- Any containing exact-form matches were retained.

Our scoring system was also based on only two criteria: word frequency and distance between matching words. Each case of text reuse is given a Tesserae score:

$$\ln \left(\frac{\sum \frac{1}{f(t_i)} + \sum \frac{1}{f(s_i)}}{d_t + d_s} \right) \tag{1}$$

where $f(t_i)$ represents the frequency of the ith matching term in the target (alluding) phrase, $f(s_i)$, the frequency of the ith matching term in the source (alluded-to) phrase, d_t, the greatest distance between two matching terms in the target, and d_s the greatest distance between two matching terms in the source. Word frequencies were specific to their respective texts. Score could be used to rank allusions retained by the filter, and also as a filter itself, with allusions scoring below a cutoff dropped from the results.

Table 2 gives recall values for all the five annotator types and for commentator allusions for one combination of scoring and filtering settings. It also gives, for each type, the mean score assigned by Tesserae. Figure 1 shows a precision-recall curve for annotator types 3–5 combined, over a range of filter/scoring settings.

Using only simple criteria such as word-frequency, distance between matching words, and presence of exact-form matches, we were able to dramatically increase the organization and precision
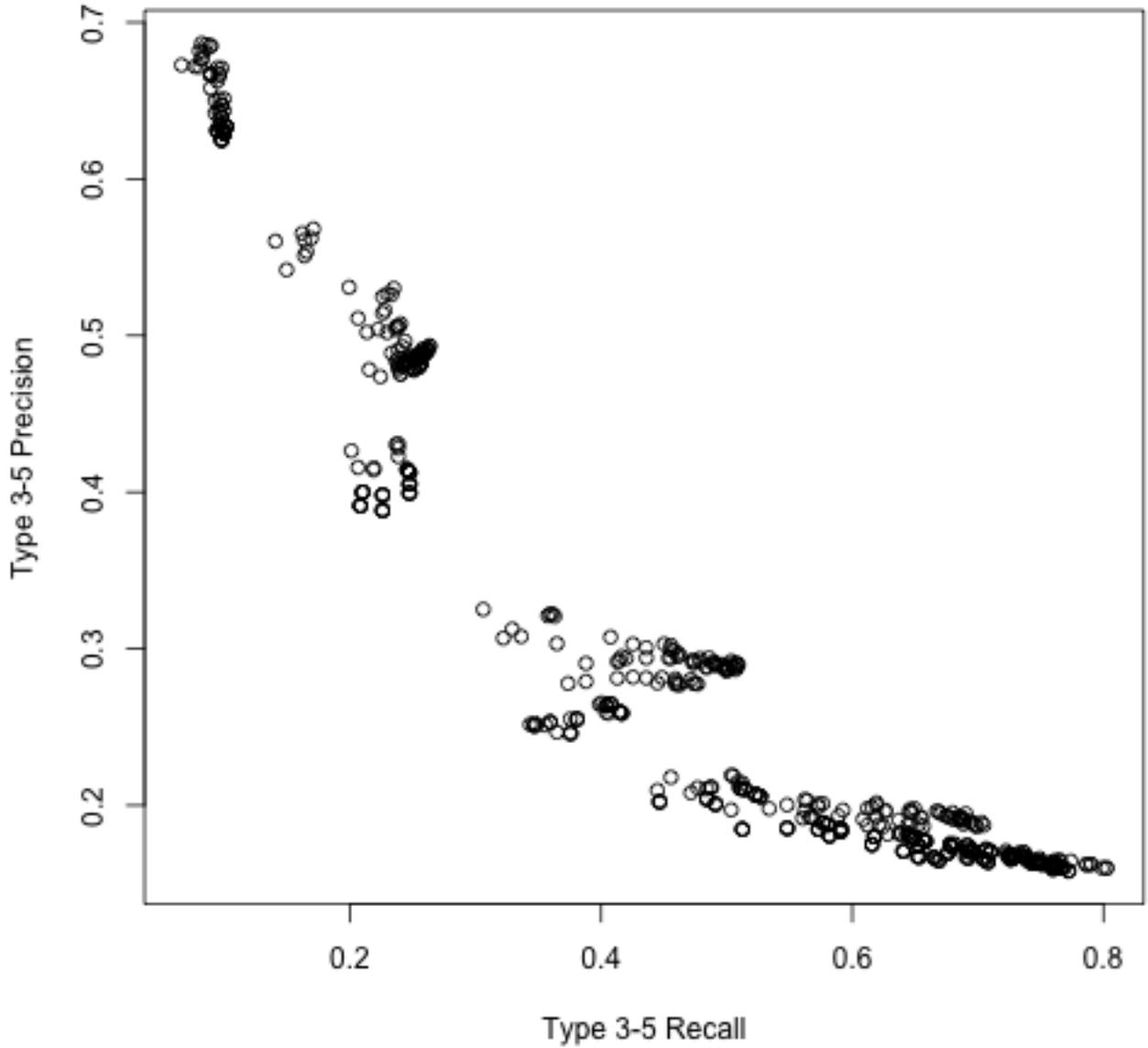
2

Figure 1: Precision vs. recall for annotator types 3–5 combined, across a range of filter/scoring system settings. Precision here is the number of allusions of type 3–5 as a proportion of all types among Tesserae results. Recall is the number of type 3–5 allusions returned as a proportion of those in the benchmark set.

Table 2: Recall data as in Table 1, with one implementation of the scoring/filtering system. The final column shows the mean score assigned by Tesserae to each type. It is interesting that commentator allusions are scored higher than any annotator type.

type	tesserae returned	benchmark had	recall rate	mean score
1	115	491	0.23	5.73
2	774	2280	0.34	5.77
3	215	396	0.54	6.30
4	59	96	0.61	6.59
5	36	70	0.51	6.46
comm.	179	336	0.53	6.62

of Tesserae results. This means more meaningful allusions near the top of the results list, a higher concentration of commentator results, with fewer results to look at overall. At the same time, this suggests that a large part of what commentators and other readers find interesting about text re-use is explained by these basic objective criteria. Although most readers would claim that they recognize allusion based on shared words alone, in fact they appear to be highly sensitive to the relative frequency of those words, unconsciously filtering out overly common collocations in favor of more interesting ones.

We continue to work on optimizing this system, as well as on discovering new textual features which will allow still more accurate scoring of allusions.

References

[1]   Heitland and Haskins (1887) *M. Annaei Lucani Pharsalia*. London: G. Bell.

[2]   Thompson and Brure (1968) Lucans Use of Vergilian Reminiscence. *Classical Philology* 63: 121.

[3]   Viansino (1995) *Marco Annaeo Lucano: La Guerra Civile (Farsaglia) libri I-V*. Milan: Arnoldo Mondadori.

[4]   Roche (2009) *Lucan: De bello civili. Book 1*. Oxford: Oxford University Press.