Comparative rates of text reuse in classical Latin hexameter poetry

report on paper by Neil Bernstein, Kyle Gervais and Wei Lin

The field

- The detection and interpretation of allusion currently represent the dominant mode of study of Latin poetry.

- The typical goal of intertextual study is to describe how links between texts affect the meaning of both the specific passages that contain them and the poems as a whole.

- Although intertextual associations may be signalled in many different ways (including similarity of action, character, or theme), verbal repetition, or text reuse, is the best studied and often the strongest type of signal.
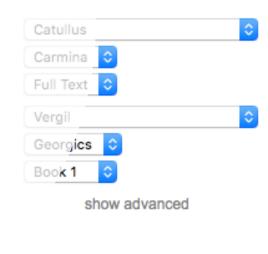
TESSERAE

Latin Search

The Tesserae project aims to provide a flexible and robust web interface for exploring intertextual parallels. Select two poems below to see a list of lines sharing two or more words (regardless of inflectional changes).

For an overview of all advanced features, see the Instructions page.

- search interface for a corpus of over 300 poetic and prose works

- discovers every instance where short passages (either lines of verse or grammatical periods) share two or more repeated lexemes

- Has a scoring system of potential interpretative significance

Source:

| Catullus |
| Carmina |
| Full Text |

Target:

| Vergil |
| Georgics |
| Book 1 |

show advanced

Compare Texts

Scoring system

$$\text{Score} = \ln\left(\frac{\sum\frac{1}{f(t)}+\sum\frac{1}{f(s)}}{d_t+d_s}\right)$$

$f(t)$ is the frequency of each matching term in the target text

$f(s)$ is the frequency of each matching term in the source text

d_t is the distance in the target text

d_s is the distance in the source text

Interpretative potential

- Statius, *Thebaid* 4. 260 *prosilit audaci **Martis** percussus **amore*** ("Parthenopaeus leapt up, struck by a bold desire for warfare").

- Vergil, *Aeneid* 7.550 *accendamque animos insani **Martis amore*** ("I'll inflame their minds with a desire for mad warfare").

- The words in the *Aeneid* are spoken by Allecto, a demon of the underworld

- Parthenopaeus' desire to fight in the Theban war in Statius is not only fatal, like the desire of Vergil's Euryalus to participate in Nisus' expedition; it is also infernal, like the war provoked by Vergil's Allecto.

Limitations

- Text reuse does not give the full picture of intertextuality: the allusions may be signalled by similarity of action, character, theme, and so on.

- Not all text reuse features the repetition of two or more lexemes.

- Repetitions with verbal variations that seem slight to a human reader are determinative for *Tesserae*.

- The Latin poets wrote for an audience of Roman elites that were literate in Greek [Hutchinson 2013], and so created numerous translingual calques on Greek phrases. The search for translingual allusions is not available yet.

- A human reader must be able to make a plausible interpretation of the allusion before it can be recognized as an allusion rather a chance repetition [Farrell 2005]

Objectives of the study

- The interpretation of specific allusions relies partly on the characterization of the overall intertextual relationship between texts, which is often hampered by a haphazard approach to gathering data.

- The paper presents a more consistent, quantitative picture of the interactions between poets in the Latin hexameter tradition.

- It uses *Tesserae* to generate **a statistical analysis of relative rates of text reuse in 24 Latin hexameter works** written from the 1st century BCE to the 6th century CE

- then compares the quantitative information about text reuse provided by *Tesserae* to the scholarly tradition of qualitative discussion of allusion by Latinists.

Questions considered

- Is it possible to quantify the verbal cohesiveness and distinctiveness of these genres?

- What other general factors affect text reuse across the entire hexameter tradition?

- Can the well-known influence of Vergil and Ovid on their epic successors be quantified?

- Which works in the classical hexameter tradition provide the most significant verbal resources for the hexameter epics of late antiquity?

Methods

- The sample included every possible source–target pair from a set of 24 Latin hexameter texts written from the 1st century BCE to the 6th century CE
- The number of "hits" in each source–target pair on Tesserae was recorded, with the following parameters of search:
  - phrases as the search unit
  - lemma as the matching feature
  - 20 stop words, determined by frequency in the *Tesserae* corpus scores calculated by stem
  - a maximum distance of 10, calculated by frequency
  - no score cutoff

Weighing of counts

- *C*7, *C*8, *C*9, *C*10, and *C*11 (counts of score 7, 8, 9, 10, and 11)
- strongly linear relationships between counts of every score

- conducted a series of linear regressions to characterize the relationship between *C*9 and the other four counts and obtain an initial composite count, *Cregr*
- applied principal component analysis (PCA) to the five counts, first correcting for their very different scales by dividing each count by its standard deviation, in order to obtain a second composite count, *Cpca*

Weighing of counts

$$C_{regr} = 0.057C_7 + 0.225C_8 + C_9 + 6.168C_{10} + 212.062C_{11}$$

$$C_{pca} = 0.057C_7 + 0.225C_8 + C_9 + 6.404C_{10} + 243.426C_{11}$$

$$C_{obs} = 0.057C_7 + 0.225C_8 + C_9 + 6.286C_{10} + 227.744C_{11}$$

Relative intensity

- impossible to compare directly: the length of the texts is different
- the expected count for any "average" pair of text (in logarithmic scale):

$$c_{exp} = -19.591 + 1.311w_s + 1.208w_t$$

- the measure of relative intensity of text reuse:

$$r = c_{obs} - c_{exp}$$

Centrality

- For each of our 24 chosen texts, the mean value of r for all pairs involving that text (23 pairs each time) was determined

- This was considered to be a measure of the "centrality" of each of our chosen texts within the 24–text set.

- A text strongly influenced by its predecessors and influential to its successors would have a higher mean r than a text more peripheral to the literary tradition of Latin hexameter poetry.

Results

- General trends (authorship, genre, time period – no influence)
- Vergil's Georgics and the Ilias Latina
- Post-Vergilian classical epic
- Didactic and satiric hexameter

General trends: Authorship

- The most important influence on text reuse intensity
- In all 13 pairs of texts by the same author, the reuse intensity was higher than average ($r > 0.000$)
- markedly so in 5 of the cases (standardized $r > 2.000$)

General trends: Genre

- Didactic, epic / panegyric, and satiric

- Within the small didactic and satiric genres, reuse intensity was higher than average for 5 of 6 pairs ($r > 0.000$; the exception, H*Sat* – J*Sat*, was slight: $r = -0.007$).

- Within the much larger (and more diverse) epic/panegyric genre, reuse intensity was higher than average in 66 of 78 pairs; the 12 remaining pairs had only slightly lower than average reuse intensity (standardized $r \geq -0.446$).

- In contrast, pairs comprising texts from different genres tended to display lower than average reuse intensity.

General trends: Genre

- The importance of genre was especially marked when pairing epic/panegyric with satiric texts

- These results indicate a strong separation between the genres, related to satire's pedestrian vocabulary and everyday concerns, which contrast with the more elevated style and subject matter of epic.

- The author signal appears to be stronger determinant than the genre signal for intensity of text reuse, as evidenced by higher r values for pairs of texts written by Horace than inter-author pairs within the satiric genre.

General trends: Time period

- No influence on text reuse intensity
- Not surprising, since the technical and aesthetic constraints of hexameter poetry discouraged changes in diction or syntax over time.

Vergil's Georgics and Ilias Latina

- The influence of Vergil's *Aeneid* on the tradition of Latin hexameter is well reflected in the results. (a high centrality score: 0.133 and higher than average reuse intensity: *r* > 0.000 when paired with 13 of 18 texts

- However, the results for Vergil's early work, the *Georgics*, are even more exceptional. (centrality score: 0.279, higher than average reuse intensity: *r* > 0.000) when paired with 16 of 20 subsequent target texts.

1. Aeneid is much longer that Georgics

2. Virgil reuses his own phrases from Georgics in Aeneid

(similar case with Ilias Latina – reused in Aeneid and Metamorphoses)

Vergil's Georgics and Ilias Latina

- The high scores for both the *Georgics* and *Ilias Latina* demonstrate that allusion in Latin literature is not always a case of a target text reusing a phrase from a single, specific source text.

- While scholars routinely privilege one source text at the expense of the others for the sake of interpretation, the automatic searches of *Tesserae* do not.

Post-Vergilian classical epic

- *Metamorphoses, Bellum Civile, Argonautica, Thebaid, Achilleid, Punica*
- the assumption has often been made that the earlier epics (*Metamorphoses* and *Bellum Civile*) responded to Vergil's influence in more creative and original ways, while the four later epics of the Flavian period tended to imitate Vergilian epic less creatively.
- the intensity of text reuse of Vergil's *Georgics* and *Aeneid* by the *Argonautica*, *Thebaid*, and *Achilleid* was higher than average ($0.160 \leq r \geq 0.299$)
- However, the intensity of reuse of any of Vergil's three works by the earlier *Metamorphoses* was much greater ($0.323 \leq r \geq 0.560$).

Didactic and satiric hexameter

- *De Rerum Natura* is considered to be a fundamental contribution to succeeding poetry [Hardie 2009]

- Yet the vocabulary of the *DRN* was not mined as extensively as the other foundational works of Republican and Augustan poetry: centrality = −0.151, $r < 0.000$ when paired with 21 of 23 succeeding target texts

- The only positive r values resulted from pairings with other didactic works: Vergil's *Georgics* ($r = 0.230$) and Manilius' *Astronomica* ($r = 0.023$)

- the low r values overall demonstrate the difference between the importance of Lucretius' poem as a conceptual resource and its importance as a verbal resource

Didactic and satiric hexameter

- There are traditional scholarly observations that *Astronomica* contents thematic adaptations of Lucretius, Vergil, and Ovid [Volk 2009].

- These thematic adaptations were accompanied by verbal reuse only for Vergil in the results: Vergil's *Georgics* yielded the highest reuse intensity (r = 0.342), followed by the *Eclogues* (r = 0.307).

- Unexpectedly, Horace's *Ars Poetica* had the next highest r value (0.213).

- There may be unexplored verbal connections between the poem on composing poetry and the poem of the stars, likely in the addresses of the didactic narrator.

Conclusion

- The majority of the results conform to the statements typically made by poetry scholars about the significance of various intertextual relationships in the Latin hexameter tradition.

- The high level of correlation between the quantified results and the scholarly tradition's qualitative assessments provides a strong vote of confidence for the methodology

- The new unexpected results may be considered as suggesting avenues for future research