The Tesserae Project: Intertextual Analysis of Latin Poetry

Neil Coffee[1], J.-P. Koenig[2], Shakthi Poornima[2], Christopher Forstall[1], and Roelant Ossewaarde[2]

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

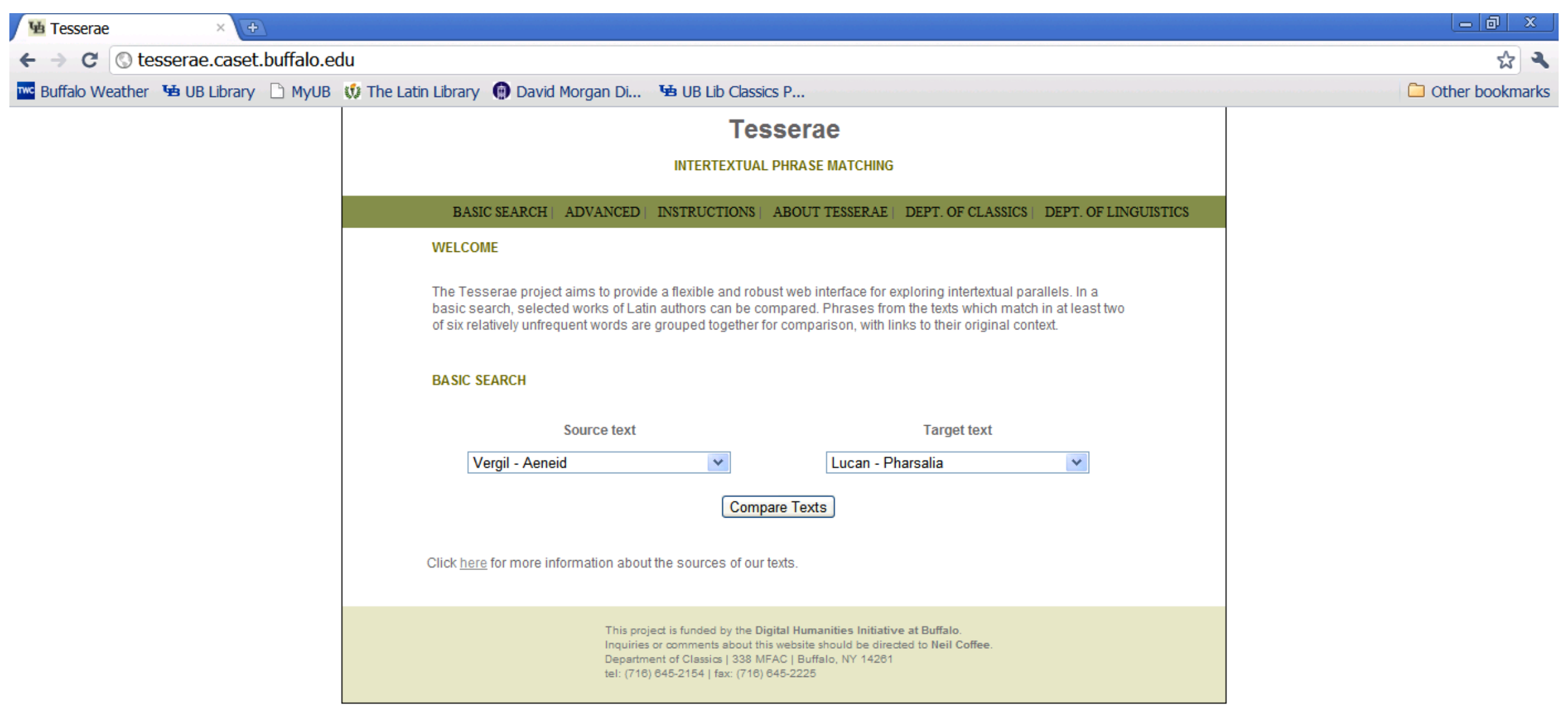
Project Goals

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.[1,2] Existing digital humanities work on text reuse has tended to concentrate on authorship attribution[3] and the identification of explicit quotations.[4] 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.[5] The Tesserae group has built upon this research to offer a web tool (http://tesserae.caset.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. 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



Tesserae

tesserae.caset.buffalo.edu

Tesserae

INTERTEXTUAL PHRASE MATCHING

BASIC SEARCH | ADVANCED | INSTRUCTIONS | ABOUT TESSERAE | DEPT. OF CLASSICS | DEPT. OF LINGUISTICS

WELCOME

The Tesserae project aims to provide a flexible and robust web interface for exploring intertextual parallels. In a basic search, selected works of Latin authors can be compared. Phrases from the texts which match in at least two of six relatively unfrequent words are grouped together for comparison, with links to their original context.

BASIC SEARCH

Source text: Vergil - Aeneid

Target text: Lucan - Pharsalia

Click here for more information about the sources of our texts.

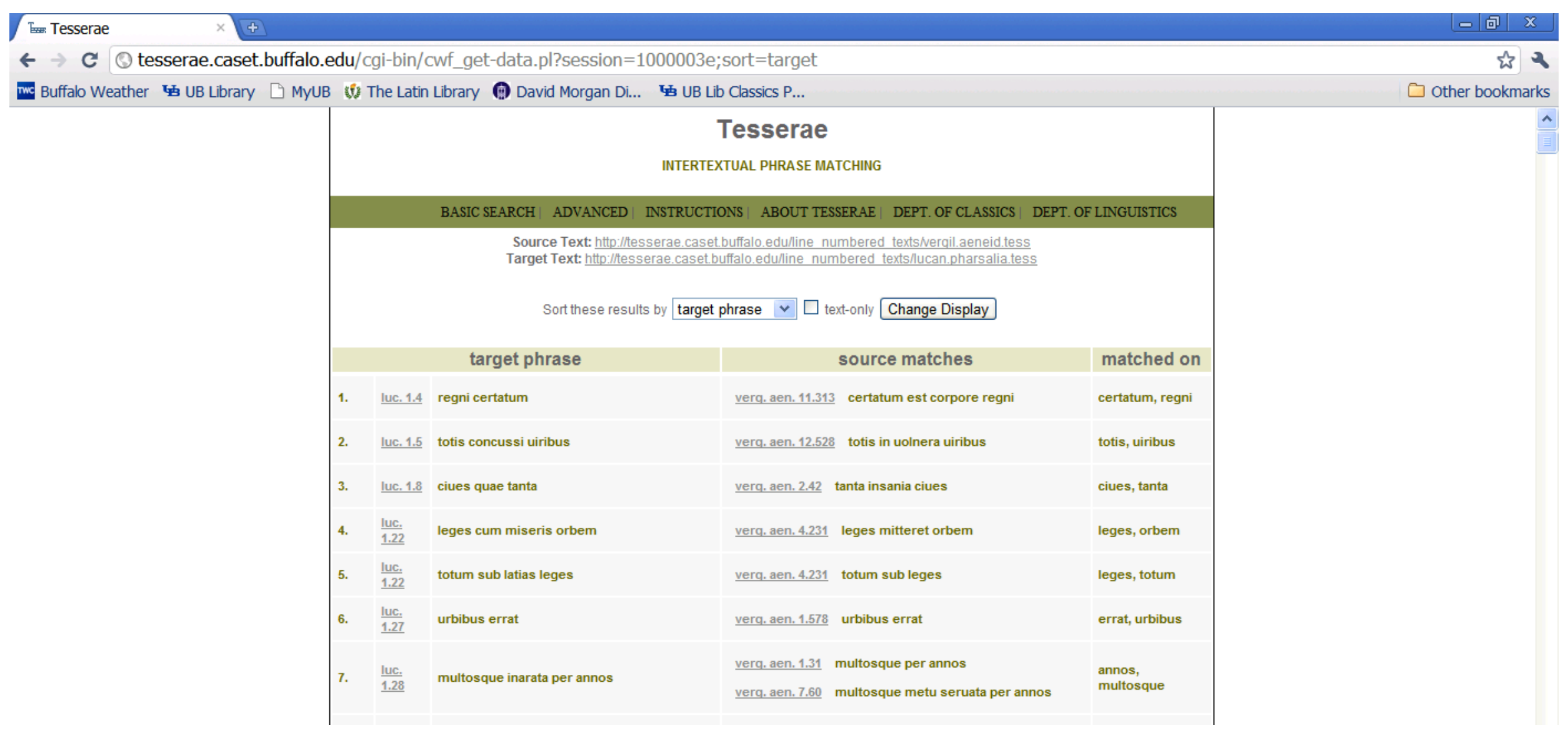
This project is funded by the Digital Humanities Initiative at Buffalo.
Inquiries or comments about this website should be directed to Neil Coffee.
Department of Classics | 338 MFAC | Buffalo, NY 14261
tel: (716) 645-2154 | fax: (716) 645-2225

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	Stem (Headword) Identity
Text A is examined six words at a time to see if any from a pre-existing list of exact 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.	Text A is examined whole sentences at a time to see if any from 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.

Results are delivered to the user as a list of two-word phrases for inspection.

Tesserae Web Interface: Results Page



Tesserae

tesserae.caset.buffalo.edu/cgi-bin/cwf_get-data.pl?session=1000003e;sort=target

Tesserae

INTERTEXTUAL PHRASE MATCHING

BASIC SEARCH | ADVANCED | INSTRUCTIONS | ABOUT TESSERAE | DEPT. OF CLASSICS | DEPT. OF LINGUISTICS

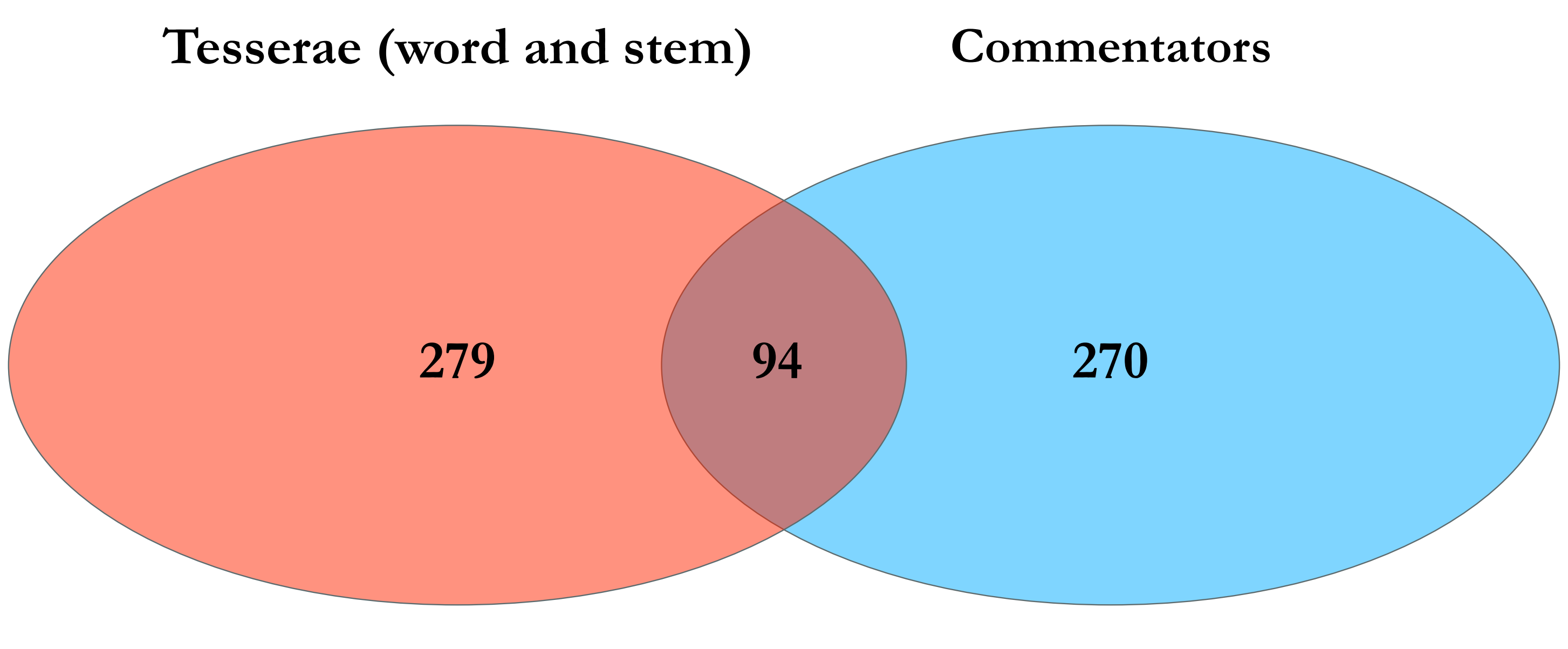
Source Text: http://tesserae.caset.buffalo.edu/line_numbered_texts/vergil.aeneid.tess
Target Text: http://tesserae.caset.buffalo.edu/line_numbered_texts/lucan.pharsalia.tess

Sort these results by: target phrase ☐ text-only

	target phrase		source matches		matched on
1.	luc. 1.4	regni certatum	verg. aen. 11.313	certatum est corpore regni	certatum, regni
2.	luc. 1.5	totis concussi uiribus	verg. aen. 12.528	totis in uolnera uiribus	totis, uiribus
3.	luc. 1.8	ciues quae tanta	verg. aen. 2.42	tanta insania ciues	ciues, tanta
4.	luc. 1.22	leges cum miseris orbem	verg. aen. 4.231	leges mitteret orbem	leges, orbem
5.	luc. 1.22	totum sub latias leges	verg. aen. 4.231	totum sub leges	leges, totum
6.	luc. 1.27	urbibus errat	verg. aen. 1.578	urbibus errat	errat, urbibus
7.	luc. 1.28	multosque inarata per annos	verg. aen. 1.31 verg. aen. 7.60	multosque per annos multosque metu seruata per annos	annos, multosque

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 ciuili* 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.



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

Conclusions from Comparison with Existing Methods

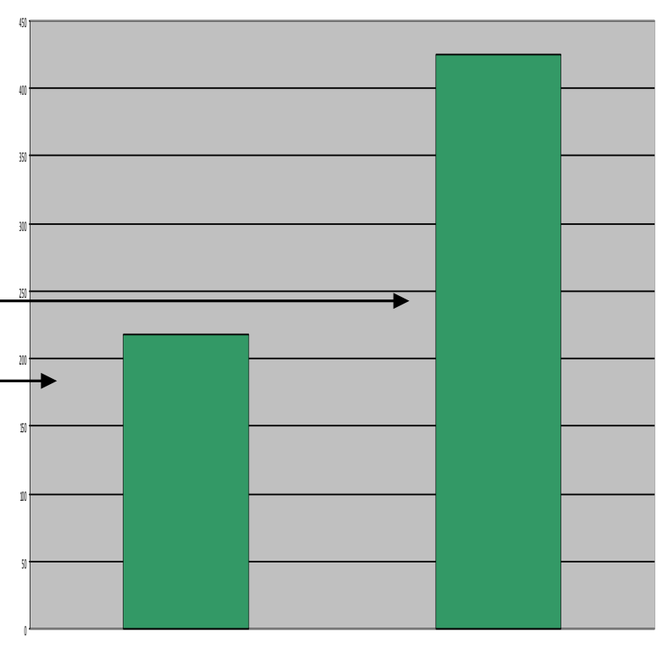
Testing shows that combined searching with both Tesserae algorithms can identify parallels found by commentators, as well as parallels not yet found even in two well-studied texts. Further work is needed to account for other feature sets, such as meter and semantics, that will allow for capture of more 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?

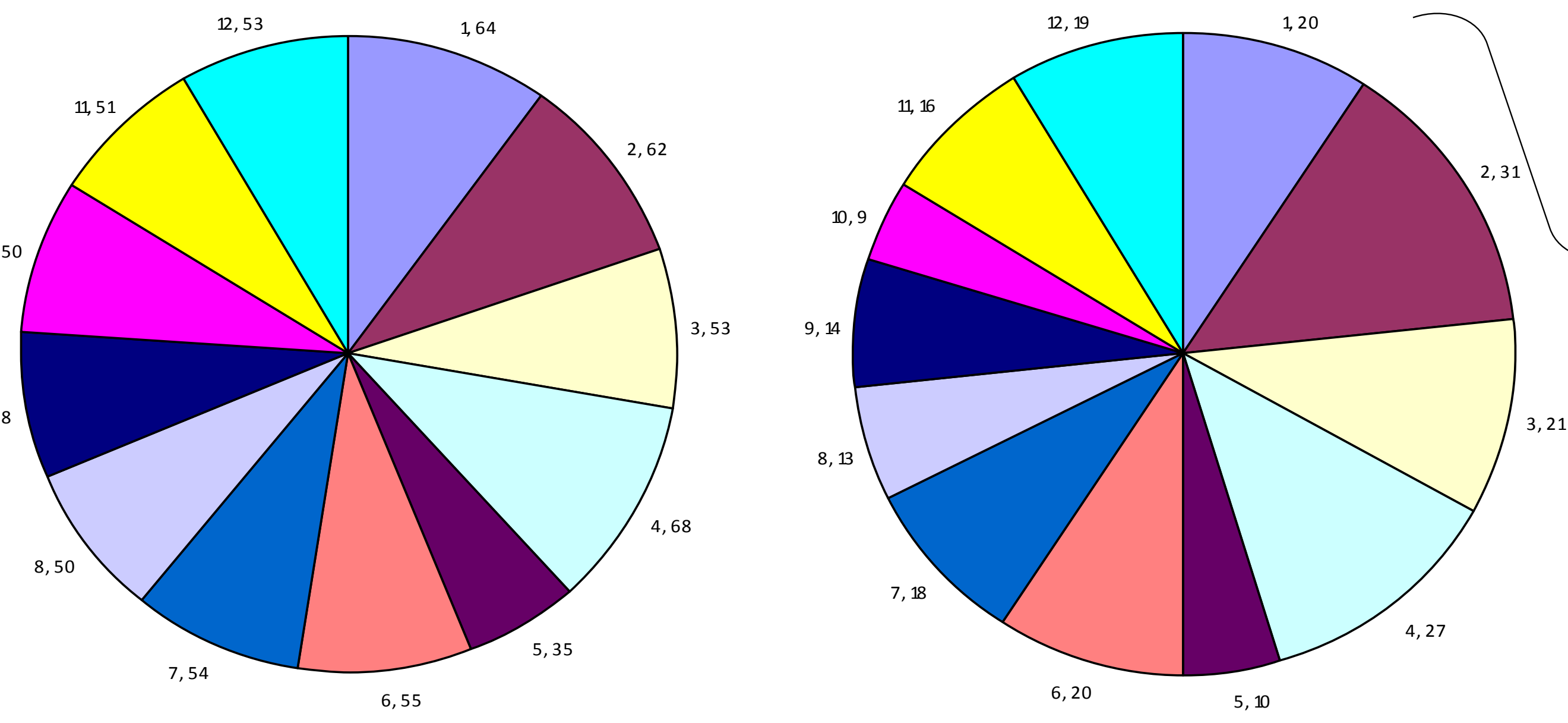
Detected parallels can be classed into those that are interpretable (i.e. generate meaning through analogous contexts) and those that are linguistic borrowings (i.e. where non-analogous or generic contexts show they are just reuse of language). In the case of 643 parallels found in the Lucan 1 – *Aeneid* comparison, **425 were judged as linguistic borrowings** and **218 as interpretable**. Lucan was clearly both steeped 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.

Relatively **even pattern** of all significant references to books of *Aeneid*



Greater use of first books of *Aeneid* for interpretable references

All Significant References

Interpretable References Only

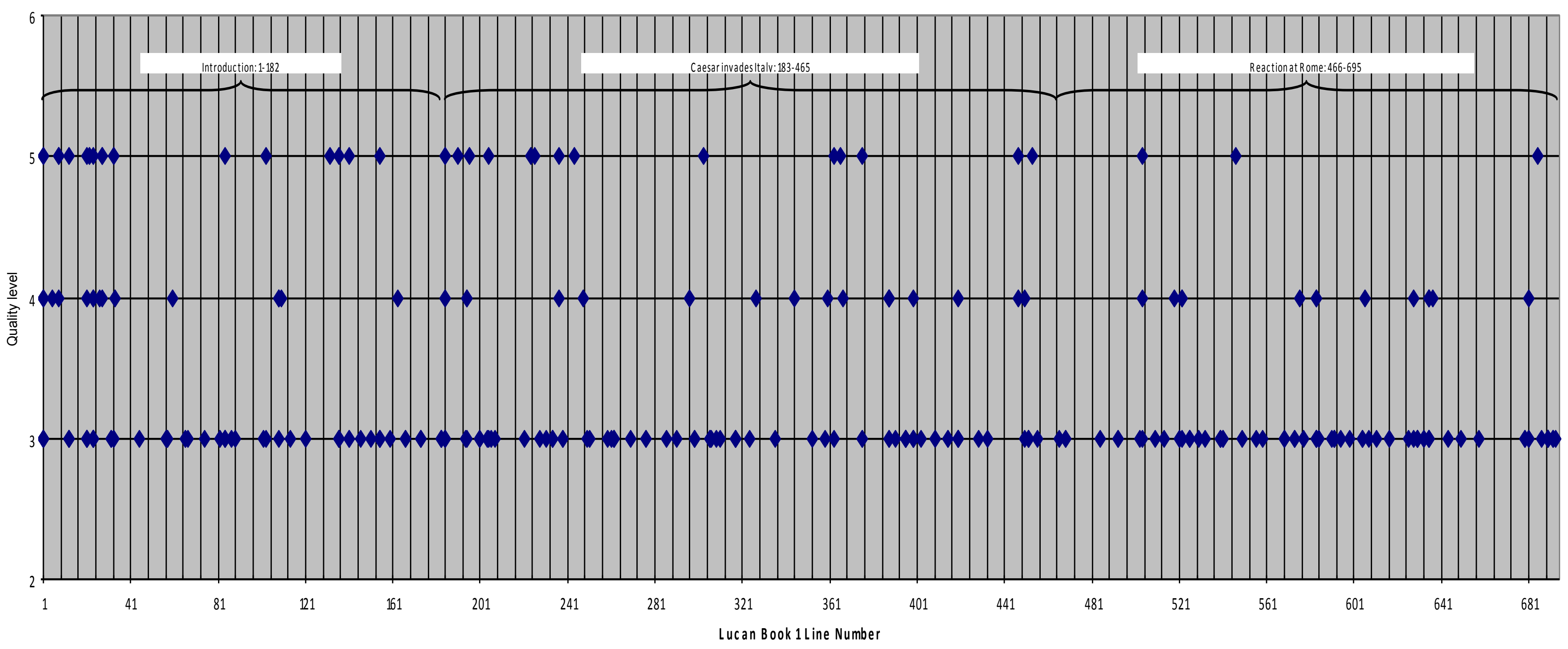
Lucan *Civil War* 1 References to *Aeneid* by book of *Aeneid*

***In What Situations* Does an Author Refer to Another?**

A final question we can answer is in what situations in book 1 Lucan chooses to make significant references to Vergil. The following chart shows Lucan making such references much more frequently in the first two thirds of book 1. A more detailed look shows Lucan avoiding Vergilian references that might ennoble the emperor Nero and the traitor Curio, while using a density of references to reverse Vergil's positive vision of empire in his proem and the scene of Caesar crossing the Rubicon.

Distribution of References to *Aeneid* in Lucan *BC* 1 by Line Number and Quality

5 or 4 = higher and lower quality significant parallel; 3 = linguistic parallel



References

1. S. Hinds. 1998. *Allusion and Intertext: The Dynamics of Appropriation in Roman Poetry*. New York, Cambridge University Press.
2. L. Edmunds. 2001. *Intertextuality and the Reading of Roman Poetry*. Baltimore, Johns Hopkins University Press.
3. M. Büchler, A. Geßner, et al. 2010. "Unsupervised detection and visualization of textual reuse on Ancient Greek texts." *Proceedings of the Chicago Colloquium on Digital Humanities and Computer Science*, 1(2).
4. R.Trillini, S. Quassdorf. 2010. A 'key to all quotations'? a corpus-based parameter model of intertextuality. *Literary and Linguistic Computing*, 25(3):269–286.
5. D. Bamman, G. Crane. 2008. "The Logic and Discovery of Textual Allusion". LaTeCH (Language Technology for Cultural Heritage Data), Marrakech Morocco.