

Modeling the Scholars: Detecting Intertextuality through Enhanced Word-Level N-Gram
Matching

Christopher Forstall, Neil Coffee, Thomas Buck, Katherine Roache, Sarah Jacobson

NOTE: THIS IS A PRE-PRINT DRAFT VERSION. The published version contains several editorial changes. Interested readers are advised to consult the forthcoming version of this paper in LLC ©: 2014. Published by Oxford University Press. All rights reserved.

Intertextuality is an important part of linguistic and literary expression, and has consequently been the object of sustained scholarly attention from antiquity onward. The definition of intertextuality has been much debated, but it is commonly understood as the reuse of text where the reuse itself creates new meaning or has expressive effects, distinct from the unmarked reuse of language.¹ In recent years, digital humanists have taken various approaches to detecting forms of intertextuality.² This article reports on an advance in

¹ In the area of Latin literature, which we focus on here, key works on intertextuality include Conte 1986, Martindale 1993, Wills 1996, Hinds 1998, Pucci 1998, Edmunds 2001, Barchiesi 2001, Farrell 2005, and Hutchinson 2013. More general studies include Ben-Porat 1976, Genette 1997, Irwin 2001, Ricks 2002, and Allen 2011. The term “intertextuality” was coined by Kristeva 1986. An annotated bibliography on intertextuality surveying these and other works is provided by Coffee 2012.

² Bamman and Crane 2008, Büchler, Geßner et al. 2010, Trillini and Quassdorf 2010, Büchler, Crane et al. 2011, Berti 2013.

automatic detection of a subset of intertextuality, namely, instances of text reuse determined by scholars of classical Latin to bear literary significance. This work was carried out by the Tesserae Project research group, whose approach is distinctive for combining: 1) efforts to use digital methods to emulate scholarly intertextual reading, 2) corresponding procedures for testing results against scholarship, and 3) an evolving free website for intertextual detection and analysis, <http://tesserae.caset.buffalo.edu/>.³

Tesserae Version 1 matched exact word strings within moveable word windows. Version 2 added the capacity for lemma matching by line or sentence. Deployment of these versions on the Tesserae website provided scholars with a means of automatically finding phrase parallels that were candidates for instances of intertextuality. A previous test comparison of two Latin epic poems demonstrated that the word-level n-gram matching employed by both versions could detect the majority of intertexts identified by scholars.⁴ The search lacked precision, however, so intertexts lay undifferentiated in long lists of candidate parallels, the vast majority of which were not meaningful. Version 3 now provides a filtering function that ranks parallels by significance, making it substantially easier to find those of greater potential interest. The Version 3 search algorithm is now the default method for searching the newly expanded corpus of Latin, ancient Greek, and English available on the Tesserae site. This article describes the performance of Version 3 search.

³ The complete code is available at <https://github.com/tesserae/tesserae>.

⁴ Coffee, Koenig et al. 2012a, Coffee, Koenig et al. 2012b.

METHODOLOGY

Tesserae search proceeds in two stages. In the first stage, the search identifies all instances where a given unit in one selected text shares at least two words with a unit in another selected text. The units can be either lines of poetry or “phrases,” where a phrase is equivalent to a sentence or text demarcated by a semicolon or colon. Words can be matched by exact word form (for Latin, *cano*, “I sing” = *cano*) or dictionary headword (*cano*, “I sing” = *cecini* “I sang”). Users can choose to exclude common words using a stop list, the size and source of which (one text, both texts, or the corpus) can be adjusted. This first stage of the Version 3 search is conceptually identical to that of previous versions, but incorporates some modifications to the code that produce a greatly increased number of phrase matches.

To achieve better precision than provided by the stop list alone, Version 3 introduces a second stage scoring system that ranks results by two additional criteria: the relative rarity of the words in the phrases shared by the two texts (“word frequency”), and the proximity of the shared words in each text (“phrase density”). We privileged word frequency because we observed that, with notable exceptions, phrases identified by scholars as intertexts consist of words that are relatively rare in their contexts. We privileged phrase density because we observed that scholars generally found intertexts to consist of compact rather than diffuse collocations. The equation given in Figure 1 represents our attempt to express the relationship of these criteria as a measure of intertextual significance. The inputs to this equation are the frequency of each matching word in its respective text, and the distance between the two most infrequent words in each of the two phrases. The output is a prediction of interpretive

significance generally falling between 2 and 10. The effect of the equation is that, for a given parallel, the rarer the shared words are, and the closer together in their respective texts, the higher its score will be.

TESTING

Search Stage 1: Phrase Matching

To assess the Version 3 search, we conducted a test that compared our results to a benchmark set of scholarly parallels between two Latin epic poems considered to have a high level of intertextual relation, Vergil's *Aeneid* (9,896 lines of hexameter verse) and book 1 of Lucan's *Civil War* (695 lines of hexameter verse). We performed the search using the Tesserae Corpus-wide search interface (<http://tesserae.caset.buffalo.edu/multi-text.php>, Fig. 2). The interface allowed us to generate a list of parallel passages with common phrases, and also to see where else in the corpus those phrases appeared, as an aid to the hand-ranking process described below. We selected relatively unrestricted settings for our search to capture the greatest number of meaningful results. We compared texts by phrases rather than lines, since phrases were generally longer and so could find a broader range of intertexts. We searched by lemma rather than exact word, at the cost of some false matches,⁵ to allow for the detection of intertexts with identical roots but different forms, a necessary measure for a highly inflected language like Latin. We chose a stop list that excluded only the 10 most common lemmata in

⁵ Lemmatization is at present unsupervised. In cases where an inflected form is ambiguous (e.g. Latin *bello* could mean "war" or "handsome"), it is allowed to match on any of the possible lemmata.

Civil War 1 and the *Aeneid* taken together. The stop list words were: *et, qui, quis, in, hic, sum, tu, per, neque, and fero*.⁶ The resulting search generated a list of 23,617 phrase parallels between the *Aeneid* and *Civil War* book 1, each with an automatically assigned score.

Comparison of these parallels with the benchmark set showed that the search captured 62% of the intertexts recorded by scholars.⁷

⁶ Users can replicate the search discussed here by using the following parameters on the Corpus-Wide Search page. Source: Vergil *Aeneid*; target: Lucan *Bellum Civile* book 1; unit: phrase; feature: lemma; number of stop words: 10; stop list basis: target + source; maximum distance: 50 words; distance metric: frequency; drop scores below: 0; filter matches with other texts: no filter; texts to search: all. The original distance metric counted both words and non-word tokens such as spaces and punctuation marks. Since word and non-word tokens generally alternate, one should cut this number in half to estimate the number of intervening words in the “sparsest” parallels. The current, revised metric counts only words, and produces comparable results when set to a maximum of 23.

⁷ Our list of scholarly parallels was compiled from the Lucan commentaries of Heitland and Haskins 1887, Thompson and Bruère 1968, Viansino 1995, and Roche 2009. These were supplemented by a list of parallels not recorded by scholars that had been generated in previous testing and graded according to the scoring system described below. Note that the 62% recall reported here excluded matches on the list of stop words, as well as phrases in which matching words were very far apart (see below). Without these restrictions, recall would be higher, around 72%, though at the expense of substantially decreased precision.

We further attempted to determine if the search had revealed new meaningful intertexts. This required assessing the quality of the parallels returned in the search that had not been noted by scholars. For the assessment, we used a hand-ranking scale we had previously developed for this purpose, given in Table 1.⁸ The scale has five ranks, from least to greatest significance for the literary interpreter. For testing purposes, we concentrated principally on whether parallels passed one of two thresholds. To clear the first threshold, a phrase parallel needed to have marked language and therefore be of potential interest for its artistry. This standard excluded both erroneous matches (type 1) and instances of unmarked, ordinary language (type 2). The determination as to whether a given phrase parallel had marked language was made in part through consideration of how often it appeared elsewhere in the corpus, as indicated by results from the Corpus-Wide search function. All other things being equal, a phrase parallel between the two texts that was rare in the corpus was considered of greater interest than a parallel common in the corpus.⁹ Parallels passing this

⁸ For a full explanation of the scale, see Coffee, Koenig et al. 2012, 392-398.

⁹ This criterion is meant to exclude very common collocations. For example, forms of the expression “lift oneself up” (*se tollere*) occur at *Civil War* 1.142 and *Aeneid* 2.699, but also in 82 other texts in our corpus, confirming that it is a common expression and uninteresting in and of itself. At the same time, classicists have recognized instances where an intertext in fact becomes more meaningful by having been repeated, generally with variation, in multiple locations. A distinction is commonly made between a parallel consisting of two (or few) textual loci, called an *allusion* or *intertext*, and a set of multiple occurrences with close similarities, called a *topos*. Homer initiates the “many mouths” topos by declaring that he could not name

threshold were awarded a minimum score of 3 and deemed, in our terms, “meaningful.” To clear the second threshold, a phrase parallel needed, in addition to marked language, sufficient contextual analogy between its two passages that a reader could interpret significance in their interaction.¹⁰ Parallels passing this threshold were awarded a minimum score of 4 and deemed, in our terms, “interpretable.”

all the Greek forces at Troy even if he had ten tongues, ten mouths, an unstoppable voice, and a heart of bronze (*Il.* 2.488-90). The Roman poets Lucretius, Vergil, Ovid, Persius, Silius Italicus, Statius, and Valerius Flaccus later pick up and rework the conceit into a commonplace (Hinds 1998, 34-47). Overall, it would seem that the sense of a continuum from fewest to greatest number of phrase repetitions underlies the qualitative labels allusion / intertext, topos, generic language, and ordinary language, even if there is more to these categories than phrase repetition. It may be possible to incorporate phrase frequency into a future scoring system, in which case this issue would need closer examination. For this test, phrase frequency was considered by human evaluators, which allowed for the possibility of discrimination between these types.

¹⁰ Our criteria for meaningful and interpretable parallels draw upon existing theoretical distinctions. Fowler 2000, 122 has written that the two fundamental criteria for an intertext are “markedness and sense.” Markedness is the quality that makes a parallel “stand out” and makes it “special.” We take Fowler’s criterion of markedness to refer principally, if not exclusively, to the sort of distinctive shared language features required to make a parallel “meaningful” in our terms. Fowler further explains that for a parallel to have “sense,” the interpreter must “make it mean.” Fowler’s criterion of “sense” corresponds to our requirement

Evaluating all the parallels in the test set was prohibitive, so we chose instead to rank a random sample consisting of 5% of the results at each automatic score level, amounting to 1,194 parallels, distributed as shown in Table 2.¹¹ The resulting quality distribution of the sample set was as follows, from most to least meaningful: Type 5: 7 (1% of results sampled); Type 4: 39 (3%), Type 3: 145 (12%), Type 2: 879 (74%); Type 1: 124 (10%). Figure 3 shows these proportions projected onto the full set of 23,617 results returned. Based on this projection, between Lucan's first book and the *Aeneid* we should expect to find 2,770 instances of phrase parallels that constitute more or less distinctive generic language (type 3) and 899 interpretable intertexts (739 type 4 and 160 type 5). Although this may appear to be an unduly large number of intertexts to be found in 695 hexameter lines, two considerations make it seem less so. First, we counted every set of parallel loci between the two texts separately. So when a given locus in the *Civil War* had parallels with multiple passages in the *Aeneid*, these each counted as separate parallels. The 899 interpretable intertexts are thus constituted by fewer than 899 separate loci

that an "interpretable" parallel have a contextual similarity in the parallel passages that generates significance.

¹¹ Of the parallels thus selected, 1,078 had already been hand-ranked in previous testing. The remaining 116 were ranked for the first time in this study. The previously ranked and newly ranked results were then combined to make a sample set where each parallel had both an automatic score and a hand rank. All results were collated into a spreadsheet that is posted on the Tesseræ blog (<http://tesseræ.caset.buffalo.edu/blog/benchmark-data/> under "Tesseræ 2012 Benchmark").

in the *Civil War*. Second, a high level of interaction is not surprising for verse (hexameter) and genre (epic) traditions generally regarded as densely intertextual.

Figure 4 illustrates the projected recall of meaningful parallels (types 3-5) from our test in relation to those recorded by commentators, showing that Version 3 is projected to increase the number of recognized meaningful intertexts substantially. Figures 5 and 6 illustrate the recall of interpretable parallels (types 4-5) produced by the Versions 1 and 2 combined (Fig. 5) and the projected recall produced by Version 3 (Fig. 6), both again in relation to those recorded by commentators. Comparison of Figures 5 and 6 illustrates the significant improvement in recall of Version 3 over even the combination of the two previous Tesserae versions. Overall, the projections from our sample suggest that Version 3 improves considerably upon previous versions in discovering meaningful and interpretable intertexts, including many that have not previously been recorded.¹²

An example of these results is a parallel found in our Tesserae Version 3 test sample, but neither noted by commentators nor discovered with previous Tesserae versions, which was assigned an automatic score of 7 and a hand-rank of 5. In *Civil War* 1, Lucan narrates the abandonment of Rome at the advent of Caesar, comparing the panicked reaction of Romans to the fear of Hannibal generations earlier:

non secus ingenti bellorum Roma tumultu
 concutitur, quam si Poenus transcenderit Alpes
 Hannibal.

¹² The total number of commentator parallels is lower in the Version 3 test because review of the earlier commentator parallels for the current test found some that were judged duplicates.

(*Civil War* 1.303-5)

Rome was rocked by the massive upheaval of war,
no less than if the Carthaginian should cross the Alps.

This passage bears some similarity to an episode in the underworld narrative of *Aeneid* book 6.

In the *Aeneid* episode, set in Rome's mythical prehistory, Aeneas's father Anchises looks forward over the centuries to the birth of the great general Marcellus who saved Rome from the Carthaginians in the First Punic War and fended off Gallic incursions:

hic rem Romanam, magno turbante tumultu,
sistet, eques sternet Poenos Gallumque rebellem,
tertiaque arma patri suspendet capta Quirino.

(*Aeneid* 6.857-9)

This [Marcellus] will keep Roman affairs standing
When it is threatened by great upheaval,
He will lay low the Carthaginian horsemen, the rebellious Gaul,
He will offer a captured general's arms to Father Quirinus,
For only the third time ever.

There are other sources, beyond this Vergilian passage, that Lucan may be drawing upon and alluding to, including some with lines that also end with the word *tumultu*.¹³ But several

¹³ In his comment on the Lucan passage, Roche 2009, 248 ad 1.303-4 does not mention this possible Vergilian parallel, but observes that "the allusion to Hannibal is compounded by the intertextual allusion to Lucretius' description of the effects of the Punic war at 3.834f. *omnia cum belli trepido concussa tumultu / horrida contremuere sub altis aetheris altis.*" Horace

features make for a distinctive recollection of the description of Marcellus by Anchises: the pairing of Rome and upheaval (*tumultu*) in the same line, the enjambment of the verb for the first line at the beginning of the second, and the placement of a form of the word “Carthaginian” (*Poenus / -os*) in the same metrical position before a caesura, in a line with identical metrical rhythm.¹⁴

The similarity of language features in the two passages meets our requirements for a meaningful intertext. There is also sufficient analogy in context to make the parallel interpretable. Both passages deal overall with the possibility of the destruction of Rome through foreign invasion and the corresponding Roman response (or lack thereof). The analogy invites the reader’s interpretation. We can thus observe that the echoing of *Aeneid* 6 in this *Civil War* passage figures Romans as not only fleeing from Caesar as they might have done from Hannibal, but also fleeing as Marcellus did *not* do when faced with an earlier Carthaginian

Carmina 4.4.45-52 has a similar combination of thought and language: *Romana pubes crevit et impio / vastata Poenorum tumultu / fana deos habuere rectos, / dixitque tandem perfidus Hannibal*. . . . The ancestor of all expressions of upheaval in Africa with *tumultu* at line-end would seem to be Ennius’s *Africa terribili tremit horrida terra tumultu* (*Annales* 309 Skutsch), a line that stuck in Cicero’s memory (*De oratore* 3.42).

¹⁴ Among the variable first four feet, both lines have an initial dactyl and then spondees. *Poenus / -os* takes up the end of the third foot and beginning of the fourth foot.

threat in the First Punic War. The resonance compounds Lucan's criticism of Romans for deserting their city.¹⁵

Search Stage 2: Scoring

Having demonstrated that Tesseract Version 3 can capture intertexts with some success, we then wished to evaluate how these intertexts could be identified among all the phrase parallels returned, the majority of which were not meaningful. This part of the testing involved evaluating how the scoring system developed for Version 3 could improve precision.

Our procedure for calculating precision was to divide the number of meaningful (type 3-5) or interpretable (type 4-5) results in our test set by the total number of results of all types (1-5). To provide a baseline, we began by calculating precision for our sample set before engaging the automatic scoring system, with results illustrated in Table 3. The published commentaries that were our model naturally had a very high rate of precision: 86% of the parallels they record are meaningful, and the remaining 14% are instances of ordinary (metrically compatible) language (type 2). For interpretable parallels (types 4-5), Version 1 gave the highest precision among Tesseract versions, since it matched by exact words, whereas the lemma matching of Version 2 and Version 3 without the scoring system, though capturing a broader range of parallels, had lower precision.

¹⁵ We have chosen to focus on the *Civil War 1 – Aeneid* comparison precisely because it is well-studied, and so allows comparison of automatic methods with existing scholarship. As is true in this case, therefore, any new parallels between the two poems revealed by Tesseract contribute to, and must be interpreted within, a larger set of recognized connections.

We then tested how effective the automatic scoring system was at identifying the most meaningful parallels. Table 4 shows how automatic scores in our sample set correspond to hand-rankings. If we average the automatic scores at each hand-rank level, we find the correlation illustrated in Figure 7. As this figure shows, overall the scoring system succeeds in distinguishing the more meaningful intertexts given higher hand ranks by assigning them higher scores. In other words, the automatic scoring system replicated the trends in assessment of intertexts performed by human readers.

To get a more concrete sense of the performance of Version 3 search, we further assessed our results in terms of recall and precision. Figures 8 and 9 illustrate how recall and precision of meaningful (types 3-5, Figure 8) and interpretable (types 4-5, Figure 9) parallels vary when we discard results below certain score levels. In both cases, discarding results with increasingly higher score levels steadily increases the proportion of interpretable or meaningful intertexts in the remaining set, leading toward consistently higher precision. Raising the score threshold also reduces recall, however, by progressively eliminating meaningful and interpretable intertexts. At this stage of development, then, the scoring system may best be employed to allow the user to filter results according to his or her needs. For example, by discarding all parallels below an automatic score level of 6 in our test set, the user can eliminate nearly three-quarters (727/1003) of the non-meaningful types 1 and 2 and yet retain some three-quarters of type 3 parallels (107/145), 90% (35/39) of type 4 parallels, and all type 5 parallels. On the other hand, those who wished to get only a high quality sample could choose to consider results only at a higher score level.

Another way to choose a score cutoff level would be to consider the combined measure of recall and precision known as an F-measure. For our F-measure assessment, we used the following equation:¹⁶

$$F_1 = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 2$$

Figure 10 illustrates the F-measure scores produced when we progressively discard results below increasingly higher automatic score levels. Though the results fall considerably below the perfect F-measure of 1 at any score cutoff level, this measurement does suggest that those interested in a relatively economical investigation into meaningful parallels would be best served by investigating those at a score level of 6 or above, while those interested in a range more likely to be interpretable could investigate those at a score level of 7 or above.

CONCLUSIONS

The Version 3 algorithm behind the current default Tesseract search is designed to identify meaningful intertexts through word-level n-gram lemma matching, word frequency, and phrase density. Our tests demonstrate that Version 3 search has considerable success in identifying intertexts in a sample comparison from two Latin epic poems. It gives higher scores to phrase parallels of greater interest, pointing users to those more likely to constitute an intertext. With relatively unrestricted settings, it can identify a majority of the intertexts recorded by scholars. These results, along with our further informal experimentation, suggest Version 3 can be similarly employed for other comparisons of Latin texts in our corpus, as well

¹⁶ Rijsbergen 1974.

as for comparisons of ancient Greek and English texts, making Tesseract search a substantial aid to intertextual study. Our results also suggest that the three criteria of lemma identity, word frequency, and phrase density are important formal components of what constitutes an intertext. When scholars identify two or more passages as intertextual, they may be using the presence or absence of these three features as implicit, if not explicit criteria.

FIGURES

Figure 1. Equation for Tesseract Version 3 scoring system

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

where

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

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


d_t is the distance in the target;

d_s is the distance in the source.

Frequency is the number of times a word occurs in its respective text divided by the total number of words in that text. The frequency of the same word may thus be different in different texts.

Distance is measured between the two lowest-frequency matching words in a phrase. We assume that, where an allusion involves more than two shared words, the lowest-frequency words are likely the most important.

Figure 2. Screenshot of Tesseract Corpus-Wide Search Interface Used in Testing



TESSERAE

[SEARCH](#)
[HELP](#)
[BLOG](#)

[BASIC SEARCH](#) [ADVANCED FEATURES](#) [ENGLISH](#) [GREEK](#) [OTHER TOOLS](#) [SOURCES](#)

Corpus-wide Search

This experimental search cross-checks your results against all other texts in the corpus. This will allow you to see whether a particular parallel is unique to your two selected works, or whether there is a broader precedent for the repeated expression.

SOURCE:

TARGET:

UNIT:

FEATURE:

NUMBER OF STOP WORDS:

STOPLIST BASIS:

MAXIMUM DISTANCE:

DISTANCE METRIC:

DROP SCORES BELOW:

FILTER MATCHES WITH OTHER TEXTS:


TEXTS TO SEARCH: All Prose Verse

Caesar - De Bello Gallico

Caesar - De Bellum Civile

Caesar Augustus - Res Gestae Divi Augusti

Catullus - Carmina



University at Buffalo
The State University of New York

NATIONAL ENDOWMENT FOR THE
Humanities

Tesseractae is a collaborative project of the University at Buffalo's Department of Classics and Department of Linguistics, and the VAST Lab of the University of Colorado at Colorado Springs.

This project is funded by the Office of Digital Humanities of the National Endowment for the Humanities and 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






Figure 3: Projected distribution by type of all 23,617 *Aeneid* – *Civil War* candidate parallels, prior to application of scoring algorithm

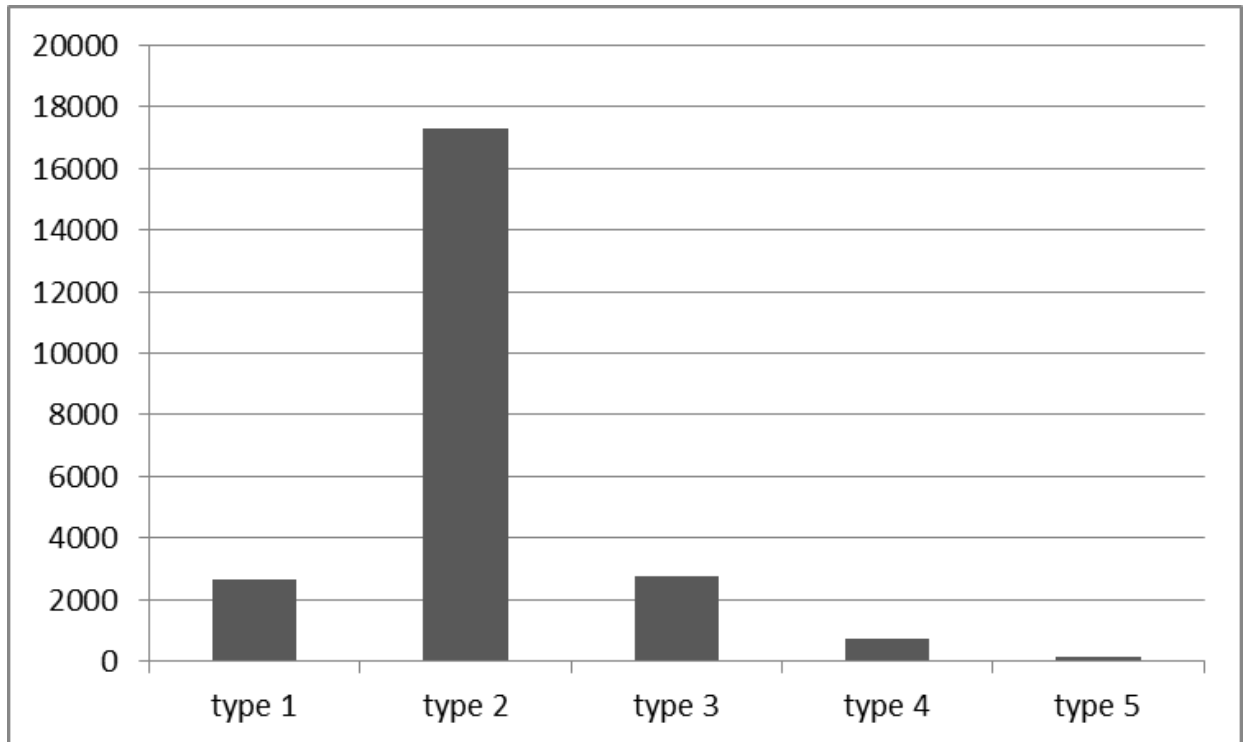


Figure 4: Numbers of meaningful (types 3-5) parallels between Lucan *Civil War* 1 and Vergil *Aeneid* found by Tesserae Version 3 (projected) and by commentators. Projected figures are produced by projecting the quality scores for a test sample over the whole larger test set.

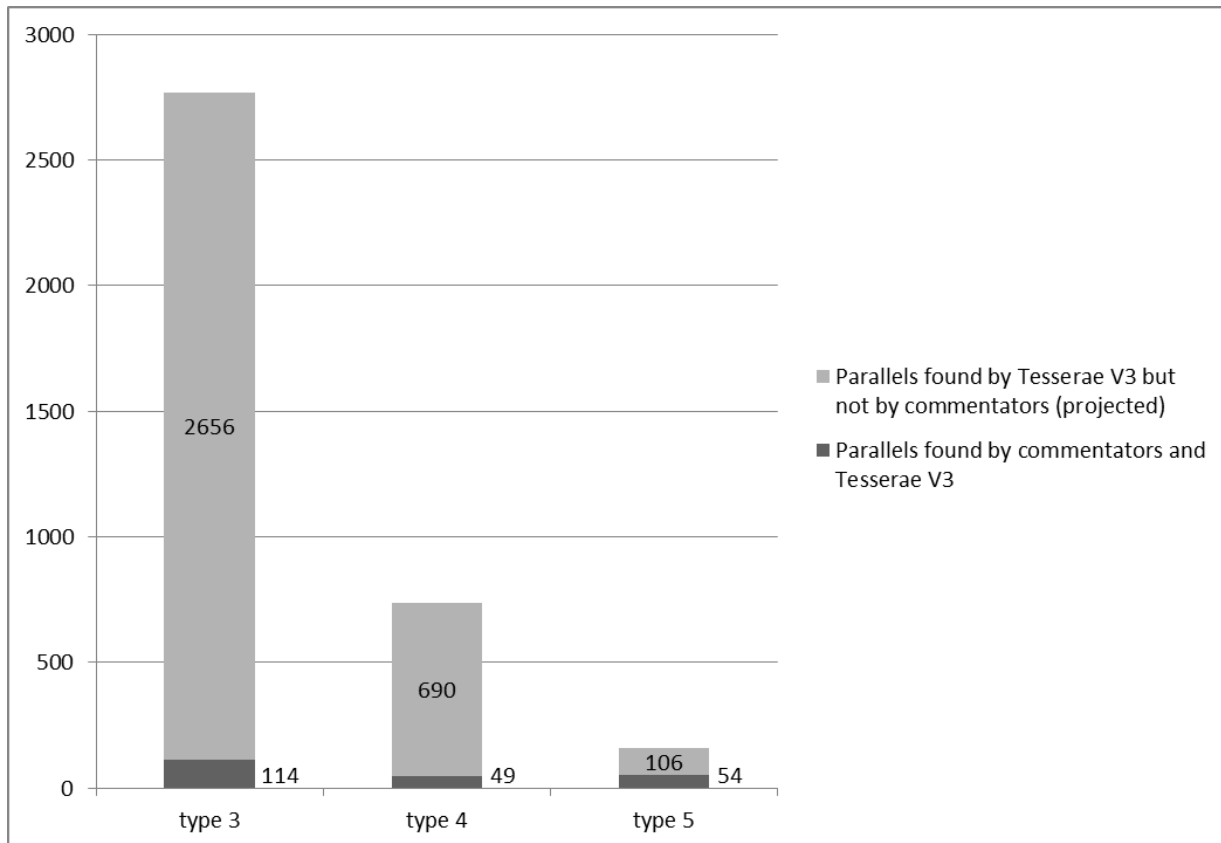


Figure 5: Unique interpretable (type 4-5) parallels between Lucan *Civil War* 1 and Vergil *Aeneid* found by Tesserae Versions 1 and 2, commentators, and both, as reported in Coffee, Koenig et al. 2012, 398.

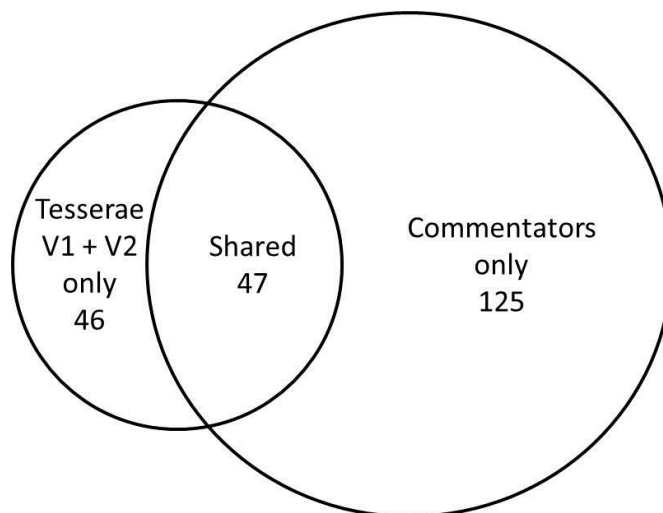


Figure 6: Unique interpretable (type 4-5) parallels between Lucan *Civil War* 1 and Vergil *Aeneid* found by Tesserae Version 3 (projected), commentators, and both.

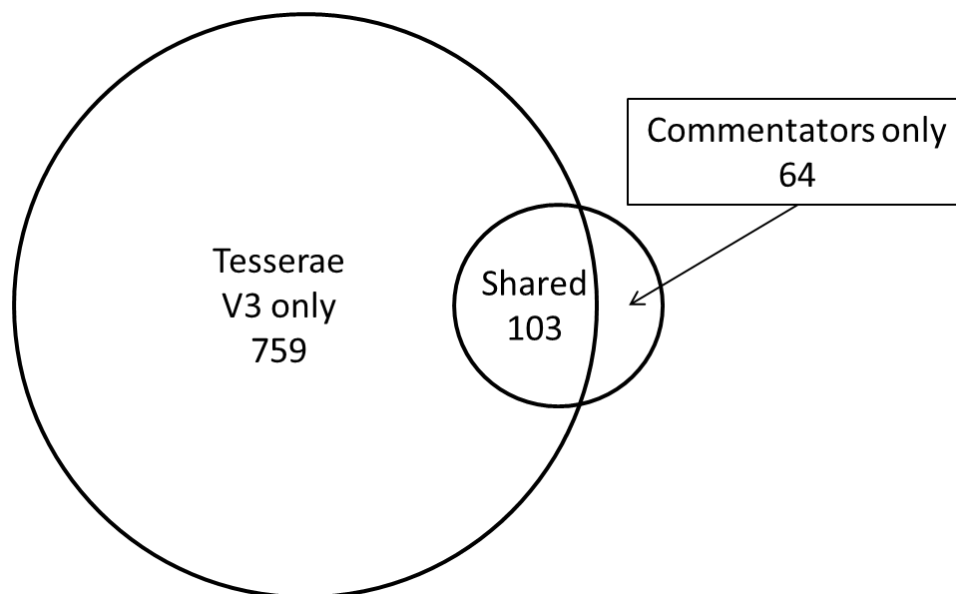


Figure 7: Correlation of Tesserae Automatic Scoring System with Hand Ranking of Intertextual Significance

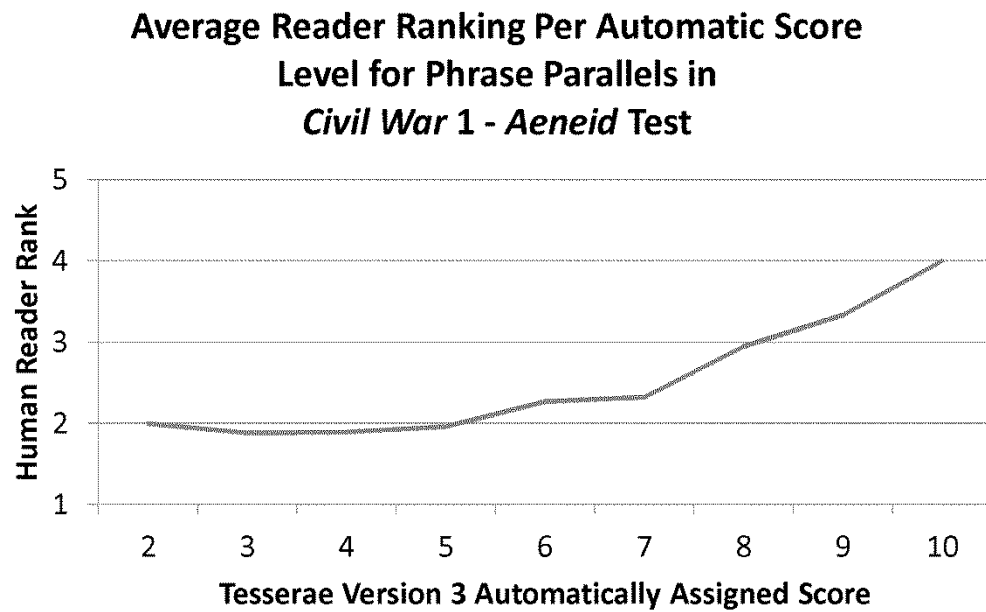


Figure 8: The effects of score cutoff on recall and precision rates for meaningful (type 3–5) parallels.

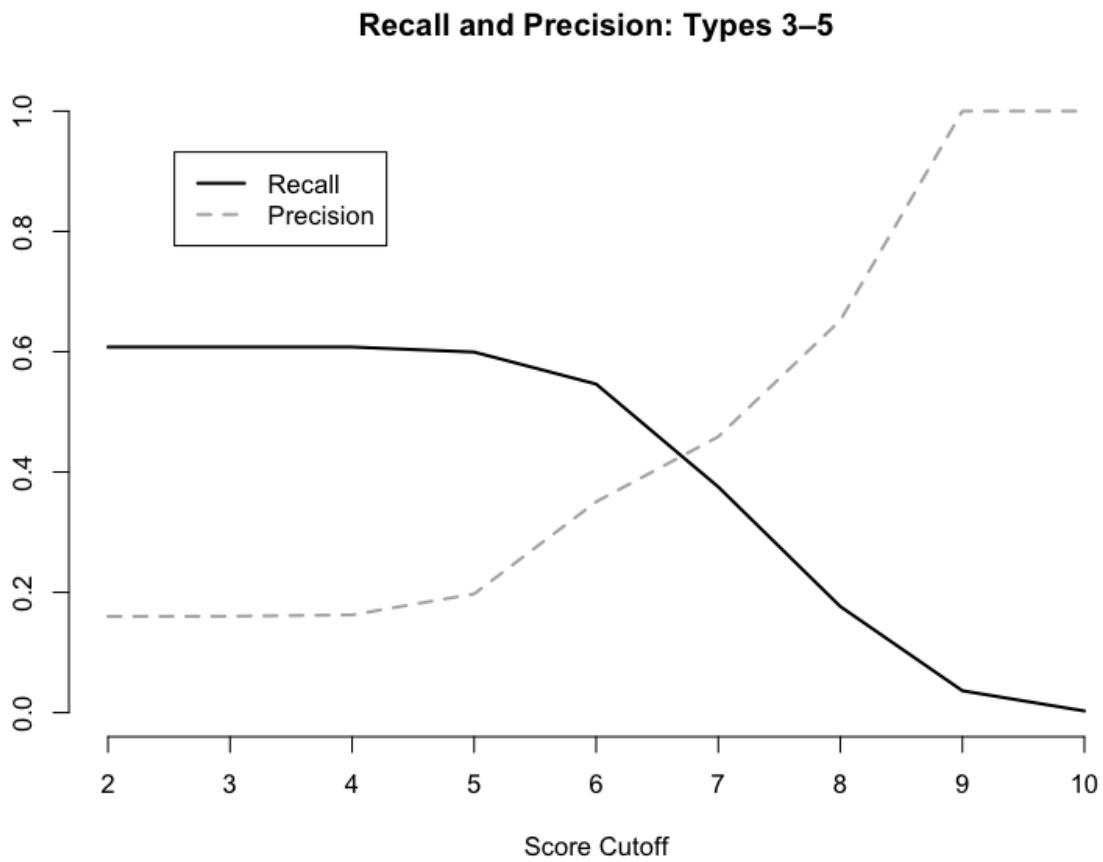
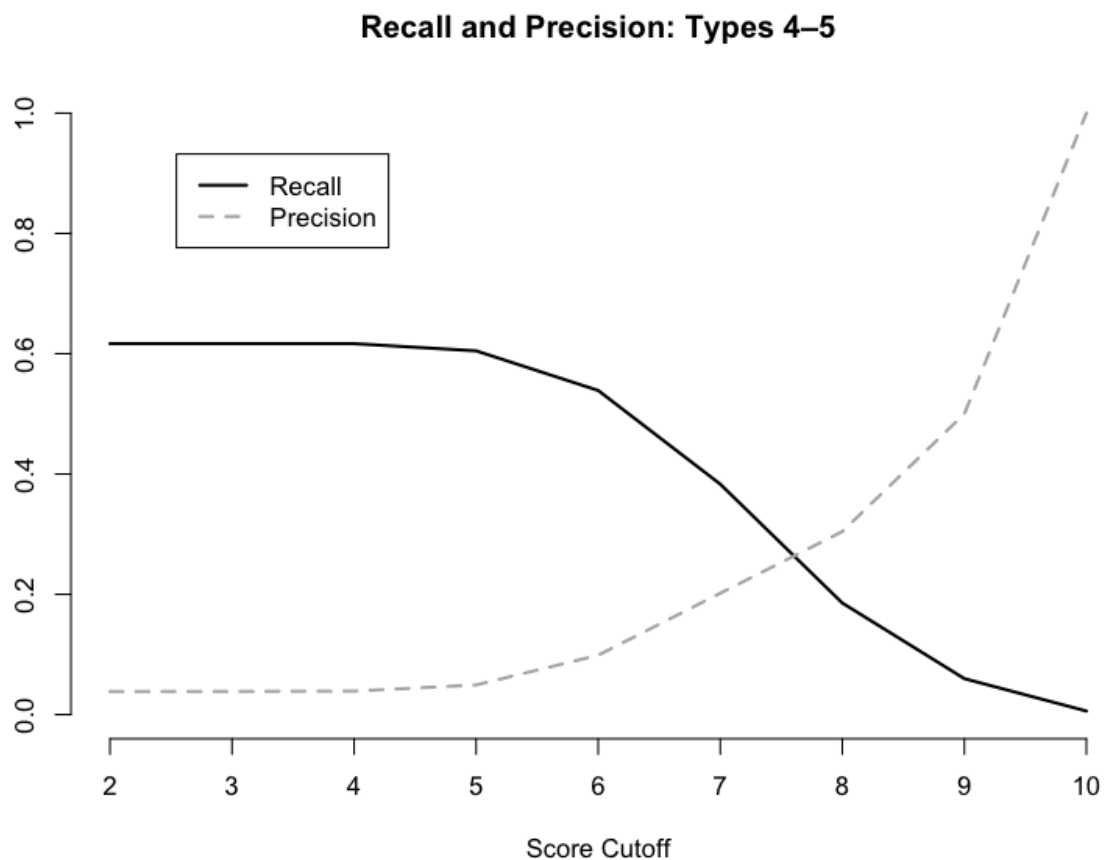
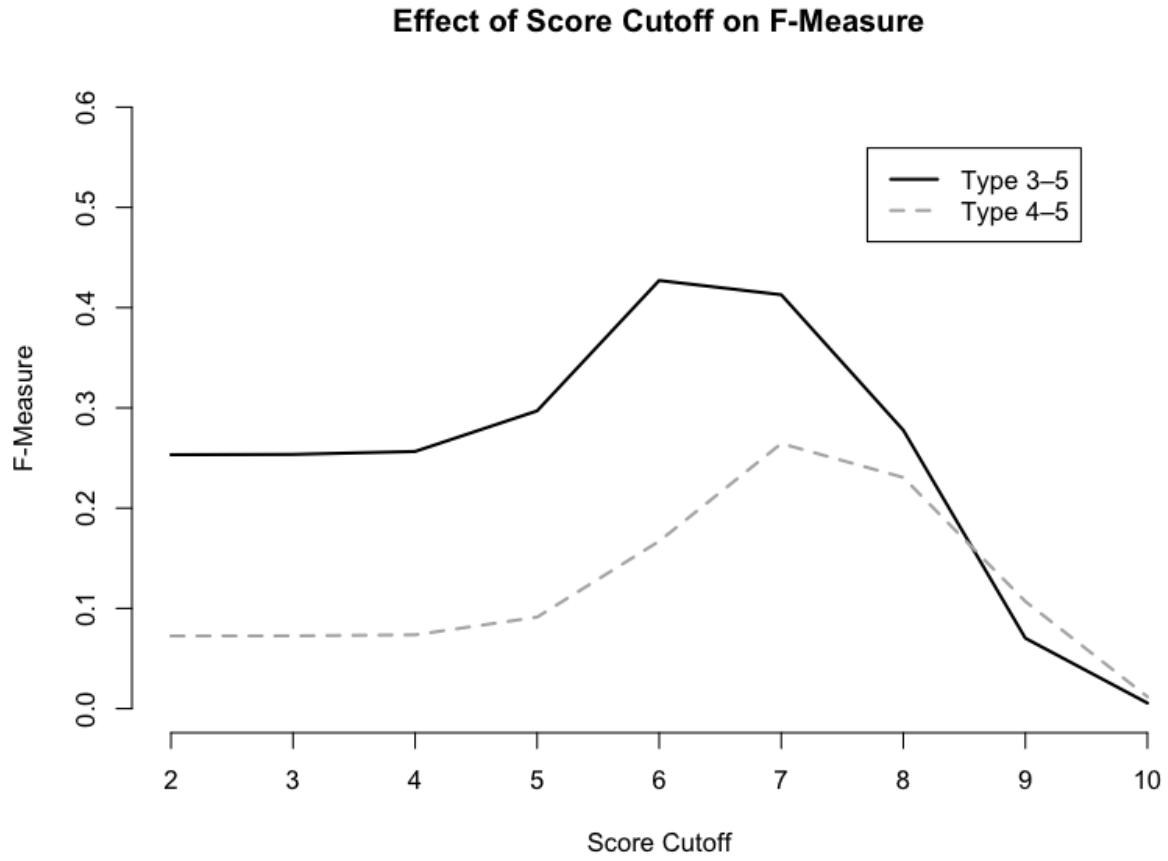


Figure 9: The effect of score cutoff on recall and precision rates for interpretable (type 4-5) parallels.¹⁷



¹⁷ Note that the stoplist and distance restrictions apply to all points on this and the following two graphs. If these constraints were removed, recall would be slightly higher and precision slightly lower, with little or no change to F-measure.

Figure 10: Effects of score cutoff on F-measure for type 4-5 parallels and for type 3-5 parallels.



TABLES

Table 1. Tesserae scale for ranking significance of intertextual parallels, from Coffee, Koenig et al. 2012, 392-398.

Type	Characteristics	Significance Category	
5	High formal similarity in analogous context.	Meaningful	Interpretable
4	Moderate formal similarity in analogous context; or High formal similarity in moderately analogous context.	Meaningful	Interpretable
3	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.	Meaningful	Not interpretable
2	Very common words in very common phrase; or Words too distant to form a phrase.	Not-meaningful	Not interpretable
1	Error in discovery algorithm, words should not have matched.	Not-meaningful	Not interpretable

Table 2. Total number of Version 3 results and number hand-ranked

Automatic Tesserae Score	Total in Test Set	Number Sampled (approx. 5%)
10	1	1
9	32	3
8	342	19
7	1721	86
6	6314	316
5	10004	507
4	4942	243
3	259	17
2	2	2

Table 3. Rates of precision for various sources in *Civil War 1 – Aeneid* test search. Given V3 precision rates are prior to application of the secondary scoring system.

Quality (Rank)	Commentators	V1 (exact form match)	V2 (lemma match)	V3 (lemma match)
Meaningful (3-5)	86%	53%	11%	17%
Interpretable (4-5)	41%	27%	2%	5%

Table 4. Comparison of automatic scores and hand-ranks for Tesseræ Version 3 sample set of parallels between *Civil War 1 – Aeneid*.

Automatic Score		Hand Rank Type				
	Total	5	4	3	2	1
10 (highest)	1		1			
9	3		1	2		
8	19	2	3	6	8	
7	86	5	10	20	44	7
6	316		20	79	184	33
5	507		4	31	412	60
4	243			7	214	22
3	17				15	2
2 (lowest)	2				2	
Total	1194	7	39	145	879	124

Works Cited

- Allen, G. 2011. *Intertextuality*. London, Routledge.
- Bamman, D. and G. Crane. 2008. "The Logic and Discovery of Textual Allusion." *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008) (Marrakesh)*.
- Barchiesi, A., Ed. 2001. *Speaking Volumes: Narrative and Intertext in Ovid and Other Latin Poets*. London, Duckworth.
- Ben-Porat, Z. 1976. "The Poetics of Literary Allusion." *A Journal for Descriptive Poetics and Theory of Literature* 1: 105-126.
- Berti, M. 2013. "Collecting Quotations by Topic: Degrees of Preservation and Transtextual Relations among Genres." *Ancient Society* 43: 269-288.
- Büchler, M., G. Crane, M. Mueller, P. Burns and G. Heyer. 2011. "One Step Closer To Paraphrase Detection On Historical Texts: About The Quality of Text Re-use Techniques and the Ability to Learn Paradigmatic Relations." In *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*. Eds. G. K. Thiruvathukal and S. E. Jones.
- Büchler, M., A. Geßner, T. Eckart and G. Heyer. 2010. "Unsupervised Detection and Visualization of Textual Reuse on Ancient Greek Texts." *Proceedings of the Chicago Colloquium on Digital Humanities and Computer Science* 1.
- Coffee, N. 2012. "Intertextuality in Latin Poetry." In *Oxford Bibliographies in Classics*. Ed. D. Clayman. New York, Oxford University Press.

Coffee, N., J.-P. Koenig, S. Poornima, C. W. Forstall, R. Ossewaarde and S. L. Jacobson. 2012a.

“The Tesserae Project: Intertextual Analysis of Latin Poetry.” *Literary and Linguistic Computing* doi:10.1093/lc/fqs033.

Coffee, N., J.-P. Koenig, S. Poornima, R. Ossewaarde, C. Forstall and S. Jacobson. 2012b.

“Intertextuality in the Digital Age.” *TAPA* 142: 381-419.

Conte, G. B. 1986. *The Rhetoric of Imitation: Genre and Poetic Memory in Virgil and Other Latin Poets*. Ithaca, Cornell University Press.

Edmunds, L. 2001. *Intertextuality and the Reading of Roman Poetry*. Baltimore, Johns Hopkins University Press.

Farrell, J. 2005. “Intention and Intertext.” *Phoenix* 59: 98-111.

Genette, G. 1997. *Palimpsests: Literature in the Second Degree*. Lincoln, University of Nebraska Press.

Hinds, S. 1998. *Allusion and Intertext: The Dynamics of Appropriation in Roman Poetry*. New York, Cambridge University Press.

Hutchinson, G. O. 2013. *Greek to Latin: Frameworks and Contexts for Intertextuality*. Oxford, Oxford.

Irwin, W. 2001. “What Is an Allusion?” *The Journal of Aesthetics and Art Criticism* 59: 287-297.

Kristeva, J. 1986. ““Word, Dialogue and Novel”.” In *The Kristeva Reader*. Ed. T. Moi. New York, Columbia University Press: 34-61.

Martindale, C. 1993. *Redeeming the Text: Latin Poetry and the Hermeneutics of Reception*. Cambridge, Cambridge University Press.

Pucci, J. 1998. *The Full-Knowing Reader: Allusion and the Power of the Reader in the Western Literary Tradition*. New Haven, Yale University Press.

Ricks, C. B. 2002. *Allusion to the Poets*. Oxford, Oxford University Press.

Trillini, R. H. and S. Quassdorf. 2010. "A 'Key to All Quotations'?: A Corpus-based Parameter Model of Intertextuality." *Literary and Linguistic Computing* 25: 269-286.

Wills, J. 1996. *Repetition in Latin Poetry: Figures of Allusion*. Oxford, Oxford University Press.