Introduction
The Tesserae Project attempts to detect literary allusions automatically by comparing two works of Latin poetry.

Allusions are often signalled by words borrowed from the alluded-to text, but not all cases of text re-use are allusions. Here we attempt to model and predict the significance readers attach to textual similarities, using easily quantifiable features of the text.

Benchmark Set
3,000 phrases were collected from Book 1 of Lucan’s Bellum Civile
• Each shares two or more words with a phrase from Virgil’s Aeneid
• Shared words may be differently inflected

400 additional allusions were drawn from professional commentaries on Lucan

Reader-Assigned Ranks
All 3,400 parallels were ranked by teams of graduate student and faculty readers for their literary significance: in each case, was Lucan making a pointed allusion to the Virgilian phrase?

<table>
<thead>
<tr>
<th>Rank</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Very meaningful allusion.</td>
</tr>
<tr>
<td>4</td>
<td>Moderately meaningful allusion.</td>
</tr>
<tr>
<td>3</td>
<td>Shared language points meaningfully to a genre or style, but not to a particular passage in the earlier text.</td>
</tr>
<tr>
<td>2</td>
<td>Shared words are not meaningful.</td>
</tr>
<tr>
<td>1</td>
<td>Error in discovery algorithm, words should not have matched.</td>
</tr>
</tbody>
</table>

The addition of the commentary results probably biases the benchmark towards better allusions, but even so, the benchmark set illustrates the general problem: most parallels are not allusions.

Scoring System
Can we predict reader-assigned significance? Our prototype scoring system uses the frequency of matching words and the distance between them:

\[
\text{score} = \ln\left(\frac{\sum f(t_i)}{\sum f(s_i)} + \frac{1}{d_s + d_t}\right)
\]

Where

- \(f(t_i)\) = Frequency of the \(i\)th matching term in the target phrase
- \(f(s_j)\) = Frequency of the \(j\)th matching term in the source phrase
- \(d_s\) = Distance between two lowest-frequency matching words in target
- \(d_t\) = Distance between two lowest-frequency matching words in source

In these figures, the dark lines show median scores for each rank, and for parallels noted/not noted by commentaries. Boxes shows the middle 50% of scores. Whiskers show the full score range for each group.

Precision-Recall
Parallels below a given score threshold are dropped, those above are kept. The following chart shows precision and recall for various thresholds as score cutoff is varied from 0 to 9.

Both measures declined where “good” results were restricted to Types 4–5. Commentary precision was undefined; the red line shows precision for Types 3–5 versus commentary recall.

Uniqueness of Parallels
Generally, allusions are more meaningful if the textual similarity is unique to the two texts in question. We found auto-assigned scores were roughly correlated with the uniqueness of parallels. As a parallel matched more texts, its score tended to be lower.