What makes an Allusion? A Digital Approach
Christopher W. Forstall, Neil Coffee, James Gawley, and Walter Scheirer
1 The University at Buffalo, State University of New York; 2 University of Colorado, Colorado Springs
http://tesserae.caset.buffalo.edu

What is an allusion?
This is a long-standing question in literary and linguistic studies. By training a computer system to detect the allusions that readers find between two texts, we arrive at formal, measurable characteristics of allusion—and discover new intertexts that traditional researchers have overlooked.

Tesserae detects the re-use of distinctive two-word phrases across texts, accounting for spelling differences and inflection.

What is an allusion?
This is a long-standing question in literary and linguistic studies. By training a computer system to detect the allusions that readers find between two texts, we arrive at formal, measurable characteristics of allusion—and discover new intertexts that traditional researchers have overlooked.

Tesserae detects the re-use of distinctive two-word phrases across texts, accounting for spelling differences and inflection.

What is an allusion?
This is a long-standing question in literary and linguistic studies. By training a computer system to detect the allusions that readers find between two texts, we arrive at formal, measurable characteristics of allusion—and discover new intertexts that traditional researchers have overlooked.

Tesserae detects the re-use of distinctive two-word phrases across texts, accounting for spelling differences and inflection.

What is an allusion?
This is a long-standing question in literary and linguistic studies. By training a computer system to detect the allusions that readers find between two texts, we arrive at formal, measurable characteristics of allusion—and discover new intertexts that traditional researchers have overlooked.

Tesserae detects the re-use of distinctive two-word phrases across texts, accounting for spelling differences and inflection.

What is an allusion?
This is a long-standing question in literary and linguistic studies. By training a computer system to detect the allusions that readers find between two texts, we arrive at formal, measurable characteristics of allusion—and discover new intertexts that traditional researchers have overlooked.

Tesserae detects the re-use of distinctive two-word phrases across texts, accounting for spelling differences and inflection.

English: William Wordsworth’s Prelude, Book 1 vs. William Cowper’s The Task

<table>
<thead>
<tr>
<th>Target Phrase</th>
<th>Source Phrase</th>
<th>Matched On</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>wordsworth prelude 1.233 Thoughtfully fitted to the Orphean lyre;</td>
<td>cowper task 3.587 By magic summons of the Orphean lyre.</td>
<td>lyre, orphean</td>
<td>9</td>
</tr>
<tr>
<td>wordsworth prelude 1.88 A backward glance upon the curling cloud</td>
<td>cowper task 4.472 There sit involved and lost in curling clouds</td>
<td>cloud, curl</td>
<td>9</td>
</tr>
</tbody>
</table>

Greek: Apollonius’s Argonautica vs. Homer’s Iliad

<table>
<thead>
<tr>
<th>Target Phrase</th>
<th>Source Phrase</th>
<th>Matched On</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.R. 1.836 ρηδίσιν, Κόμης γὰρ ἑτὶ γλυκύν ἴμμενον ὄρφον</td>
<td>hom. ll. 3.110 ὣς εἰσπόσσα ὀθαν γλυκύν ἴμμενον ἔμβαλε θυρώ</td>
<td>ἴμμενον, γλυκύν</td>
<td>9</td>
</tr>
<tr>
<td>A.R. 1.927 αὐτόπροσον κόρης Ἀθηναίινῳ αἰτή ρέθρα</td>
<td>hom. ll. 8.360 ὡκε ἀν ἄπειρουσα Στιγμον ὄδοσα αἰτή ρέθρα</td>
<td>ρέθρα, αἰτή</td>
<td>9</td>
</tr>
</tbody>
</table>

Latin: Vergili’s Georgics vs. Lucan’s Bellum Civile

<table>
<thead>
<tr>
<th>Target Phrase</th>
<th>Source Phrase</th>
<th>Matched On</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vergil. g. 1.495 exsae Inveniet scabra robigine pila</td>
<td>luc. 1.243 Et scabras nigrae morau rubiginis enses.</td>
<td>robigo, scaber</td>
<td>9</td>
</tr>
<tr>
<td>Vergil. g. 1.315 frumenta in viridi stipula lactentia turgent.</td>
<td>luc. 3.503 Nec, quamvis viridi lacteae robore, tentas</td>
<td>lactae, uride-uridilae</td>
<td>9</td>
</tr>
</tbody>
</table>

The Tesserae working-definition of allusion

Two-word (bigram) matching by lemma gives first-stage results. These are then ranked by the proximity of the words in the two-word phrases and the rarity of those words.

score = ln \left( \frac{1}{d_t} + \frac{1}{d_s} \right) \ln \left( \frac{f(t)}{f(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.

“Distinctiveness” in allusion is a function of the frequency and proximity of the two words in the matching phrase. When rare words appear close together, we are confident of an allusion.

How do Tesserae results compare to traditional scholarship?

Testing on a sample of Latin epic poetry shows that Tesserae identifies a substantial majority of the allusions important to traditional scholarship.

What about more complex allusions?
Not all allusions are characterized by the re-use of two-word phrases. In order to detect more subtle intertextuality, modules are currently under development to match phrases based on synonyms, sound patterns, and themes. At first these characteristics will be used to help determine the distinctiveness of a two-word match; at a later stage it may be possible to detect matches on these criteria alone.

The figures above represent approximately 3,000 instances of parallel phrases drawn from Lucan Civil War Book 1 and Vergil Aeneid, meeting the minimum criterion for allusion, i.e. two or more words shared between Lucan and Vergil. The reader ratings were produced by graduate student and faculty readers during a seminar on intertextuality at the University at Buffalo. Commentary data was drawn from Heitland and Haskins (1887), Thompson and Bruère (1968), Viansino (1995), and Roche (2009).