Motivation

The study of intertextuality, the shaping of a text's meaning by other texts, remains a laborious process for the literary critic. Kristeva suggests that "any text is constructed as a mosaic of quotations, any text is the absorption and transformation of another." The nature of these mosaics is widely varied, from direct quotations representing a simple and overt intertextuality, to more complex transformations that are intentionally or subconsciously absorbed into a text. Sinno, in many cases, the problem is one of pattern recognition, it is a good candidate for automated assistance by computers.

As a case study for our computational analysis of intertextuality, we turn to Paul the Deacon's 8th century poem Angustae Vitae, which we suggest has a strong connection to the poetry of Catullus.

The Clues: Angustae Vitae and Catullus

Although Paul the Deacon posits the classical and monastic worlds as opposites, the use of Catullan diction and models of poetic exchange recalls the paradigm of the Neoteric, proto-elegiac lover, his beloved, and his poetological concerns:

Angustae Vitae is peppered with classical intertexts, but it remains an open question (the indications of influence to this point have been "rare, faint, and probably indirect") as to whether Paul the Deacon had read Catullus. Thus, to build out case, we ask: What similarities exist in sound, word frequency, and metrical form?

The Functional n-gram Analysis

Observation: Sound plays a fundamental role in an author's style, particularly for poets. What similarities exist in sound, word frequency, and metrical form?

Feature: The Functional n-gram:

\[ P(x_{n:1} | x_{1:n-1}) = \frac{C(x_{n:1}, x_{1:n-1})}{C(x_{1:n-1})} \times \frac{\text{freq}(x_{n:1}, x_{1:n-1})}{\text{freq}(x_{1:n-1})} > \theta \]

In this work, we consider primitive sound elements as functional character level bi-grams.

Metrical Analysis

Metrical considerations, including the alternation of heavy and light syllables within a line, are essential in examining poetic style.

For example, Angustae Vitae is composed in elegiac couplets, which regularly take the following form:

\[ \ldots \text{nullaest evans rebus jussit} \ldots \]

\[ \text{Cui dedit luditam aevum libellum (1.1)} \]

Using TEI XML, we add the metrical values to the text. This provides flexibility in extracting hybrid feature sets. We calculate bi- and tri-gram frequencies for the metrical values as for characters in the text. A tiny set of high-frequency features accommodates smaller samples.

The Evidence

For our experiments, we considered source text from Poetae Latini Aevi Carolini², Medieval Latin² (2nd ed.), and The Tesserae Project. Key character-level n-gram considered: "re".

Functional n-gram Analysis

12 training samples (20 lines) most resembling Catullus 1 & 2 (0.06, +0.68). RBF kernel. Angustae Vitae receives a positive classification, indicating stylistic similarity to Catullus.

Raw probability features: Angustae Vitae: 0.486 Catullus 1 & 2: 0.458

For comparison, we also tested other poets, including those inspired by the Latin Neoterics. The table below reflects the number of samples classified as being stylistically similar to Catullus 1 & 2.

<table>
<thead>
<tr>
<th>Author</th>
<th>1/40</th>
<th>2/40</th>
<th>3/40</th>
<th>4/40</th>
<th>5/40</th>
<th>6/40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovid Amores</td>
<td>0.5238</td>
<td>0.3896</td>
<td>0.8696</td>
<td>0.4545</td>
<td>0.5429</td>
<td>0.8636</td>
</tr>
<tr>
<td>Tibullus Elegies</td>
<td>0.4118</td>
<td>0.4191</td>
<td>0.7595</td>
<td>0.4118</td>
<td>0.4191</td>
<td>0.7595</td>
</tr>
<tr>
<td>Propertius Epistles</td>
<td>0.5</td>
<td>0.4071</td>
<td>0.7500</td>
<td>0.5</td>
<td>0.4071</td>
<td>0.7500</td>
</tr>
<tr>
<td>Horace Epistles</td>
<td>0.5</td>
<td>0.4071</td>
<td>0.7500</td>
<td>0.5</td>
<td>0.4071</td>
<td>0.7500</td>
</tr>
<tr>
<td>Vergil Aeneid (book 4)</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
</tr>
<tr>
<td>Catullus 1 &amp; 2</td>
<td>0.458</td>
<td>0.458</td>
<td>0.7564</td>
<td>0.458</td>
<td>0.458</td>
<td>0.7564</td>
</tr>
</tbody>
</table>

Low-Probability Analysis

Serves as a refinement for the base learning results, removing fringe candidates:

<table>
<thead>
<tr>
<th>Author</th>
<th>1/40</th>
<th>2/40</th>
<th>3/40</th>
<th>4/40</th>
<th>5/40</th>
<th>6/40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovid Amores</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
</tr>
<tr>
<td>Tibullus Elegies</td>
<td>0.5</td>
<td>0.4071</td>
<td>0.7500</td>
<td>0.5</td>
<td>0.4071</td>
<td>0.7500</td>
</tr>
<tr>
<td>Propertius Epistles</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
</tr>
<tr>
<td>Horace Epistles</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
</tr>
<tr>
<td>Vergil Aeneid (book 4)</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
<td>0.486</td>
<td>0.458</td>
<td>0.7564</td>
</tr>
<tr>
<td>Catullus 1 &amp; 2</td>
<td>0.458</td>
<td>0.458</td>
<td>0.7564</td>
<td>0.458</td>
<td>0.458</td>
<td>0.7564</td>
</tr>
</tbody>
</table>

Metrical Analysis

Bi-gram frequencies ("**", "***") for Catullus proved to have too much variation for good training:

\[ \begin{align*}
1.0 & : 4.863 \times 10^{-4} & 3.70564 \\
1.0 & : 4.118 \times 10^{-4} & 3.70564 \\
1.0 & : 4.545 \times 10^{-4} & 3.86364 \\
1.0 & : 4.348 \times 10^{-4} & 3.86364 \\
1.0 & : 5.238 \times 10^{-4} & 3.91308 \\
1.0 & : 4.87 \times 10^{-4} & 3.91308 \\
\end{align*} \]

Other Discoveries

While the metrical analysis did not show similarity between Catullus and Angustae Vitae, it does help to show the direction of literary development.

References