Visualizing sound as functional n-grams in Homeric Greek poetry

Christopher W. Forstall
State University of New York at Buffalo

Walter J. Scheirer
University of Colorado at Colorado Springs

Introduction

In this work we are looking for new ways to identify stylistic heterogeneity within the Iliad and Odyssey. As oral-formulaic poetry, the Greek epics contain special evidence of the mutual relationships between poems, composition and creativity. At the same time, scholars of the digital humanities have long recognized that successful digital criticism will find ways to return from statistics to more subjective understanding. Here, we assign n-gram counts to red, green, and blue color components in order to visualize patterns of sound within the poems. The resulting images demonstrate viscerally that several well-known “set-piece” episodes within Homer’s epics have distinct n-gram distributions.

Text, sampling, and controls

- The Iliad and Odyssey were downloaded from the Perseus Project in XML.
- They were concatenated, then broken into 20-line samples:
  - once without alteration, (original series);
  - 10 more times, each time randomly re-ordering the lines of the poems before sampling (series r0…r9)

n-gram distribution

To detect which might be the most interesting n-grams, we calculated s, the number of samples in which a given n-gram occurs.

When n-grams cluster in certain samples, other samples go without, so it then lower in the original than in the r series. This is seen in the graphs at right.

The lower s, the more interesting the n-gram.

We quantify this by

\[ \text{interest} = \frac{1}{s} \text{– mean}(s_{\text{–}}) \]

Exactly how interesting this is depends on the variability of s, so we also consider the standard deviation of \( s_{\text{–}} \).

<table>
<thead>
<tr>
<th>n-gram</th>
<th>Interest</th>
<th>std(s_{\text{–}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>mm</td>
<td>0.146</td>
<td>0.02</td>
</tr>
<tr>
<td>μαχ</td>
<td>0.105</td>
<td>0.09</td>
</tr>
<tr>
<td>αχι</td>
<td>0.101</td>
<td>0.09</td>
</tr>
<tr>
<td>ππο</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>τρω</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>εκτ</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>οδυ</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>δυσ</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>μνκ</td>
<td>0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Example I: Three 2-grams for “horse”

Two of the most interesting 2-grams, h and mm, are part of the word ἵππος, “horse.” To these we added mr, (interest = 10; side\(x\) = 3).

n-gram counts were scaled and translated into color values:

- red: high interest
- green: normal
- blue: low interest

The large bright region in series original corresponds to a set-piece, the chariot race held during the funeral games for Patroclus. Next we consider three 3-grams related to three independent content elements. Each is a component of a Greek hero’s name:

- red: ὦδυssus
- green: Ἀχιllēs
- blue: Ἅδης

Each hero is foregrounded in a different part of the story:

- Ἀχιllēs in Iliad 19 and following
- ὦδυssus in the Odyssey
- Ἅδης in Iliad 5

Not only do the colors show independent thematic elements: they also can represent their interactions.

Quantifying the sound-content relationship

We measure both the number of words containing the n-gram, and the number of times each of those words occurs. The greater the lexical diversity of an n-gram, the less content-driven it is likely to be. For example, compare the 3-grams ἀργ and ὕβρις, both frequent and of high interest values.

The graphs above show the number of times each word containing a given n-gram occurs in the text as a function of that word’s rank. The right is a log-log version of the left. The ὕβρις curve has a shallower slope.

Below are the top 10 words for each. ὅβρις shows a far greater diversity than ἀργ, in large part because it contributes to some common noun and verb inflections.

Example II: Three heroes

Next we consider three 3-grams related to three independent content elements. Each is a component of a Greek hero’s name:

- red: ὦδυssus
- green: Ἀχιllēs
- blue: Ἅδης

Quantifying the sound-content relationship

We measure both the number of words containing the n-gram, and the number of times each of those words occurs. The greater the lexical diversity of an n-gram, the less content-driven it is likely to be. For example, compare the 3-grams ἀργ and ὕβρις, both frequent and of high interest values.

The graphs above show the number of times each word containing a given n-gram occurs in the text as a function of that word’s rank. The right is a log-log version of the left. The ὕβρις curve has a shallower slope.

Below are the top 10 words for each. ὅβρις shows a far greater diversity than ἀργ, in large part because it contributes to some common noun and verb inflections.

Example III: A content-independent pattern?

Here, we used shades of grey to represent counts of a single 3-gram, ὅβρις. The bright region corresponds to the “Catalogue of Ships,” Iliad 2: 484–799, and provides a subjective appreciation of the sound patterns that distinguish this inset piece (beginning with its own invocation to the Muses) from the main poem.

Notes

2. See note 1 above.
5. In composing the n-grams, we transcribe rough breathing as Latin ft initial subscript as ascident (regular ft) and final signs with medial signs as ę.
6. By way of this pattern, "Son of Tydeus."