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 may contain special evidence of the mutual relationships between poetics, cognition, and creativity.[1] At the same time, scholars of the digital humanities have long recognized that a successful digital criticism will find ways to return from statistics to more subjective understanding.[2]

Here, we assign n-gram counts to red, green, and blue color components in order to visualize patterns of sound within the poems.[3] 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

- Iliad and Odyssey downloaded from the Perseus Project[4] in XML
- concatenated, then broken into 20-line samples.

This was done 11 times: once without alteration, (**original** series); 10 more times, each time randomly re-ordering the lines of the concatenated poems before sampling (series **r0**...**r9**)

n-gram Distribution

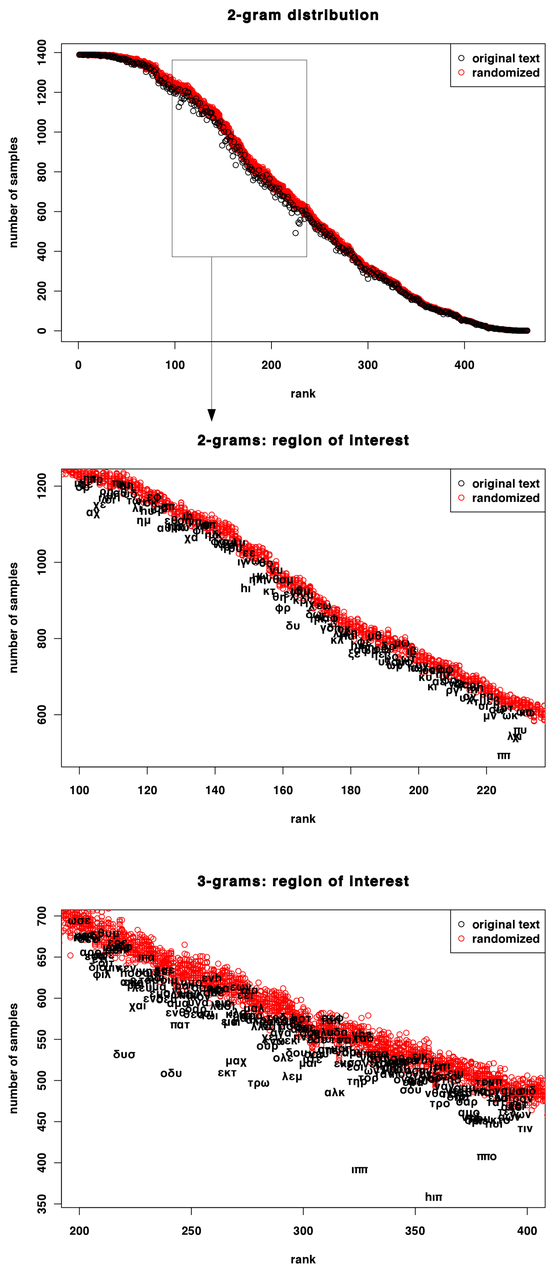
Trying to detect which might be the most interesting n-grams, we calculated *s*, the number of samples in which a given n-gram occurs.

When n-grams cluster in certain samples, other samples go without; *s* is then lower in **the original** than in the **r** series. The lower *s*, the more interesting the n-gram.

We quantify this by

$$\textbf{interest} = s_{\text{original}} - \text{mean}(s_{r0} \ldots s_{r9})$$

Exactly how interesting this is depends on the variability of *s*, so we also consider the standard deviation of (s_{r0} ... s_{r9}).



top 2-grams, by **interest**

n-gram	interest	sdev(s_r)
ττπ	-146	8
δυ	-105	9
αχ	-101	9
hι	-100	10
χι	-90	7
λκ	-86	7
φρ	-75	12
μν	-70	10
κτ	-69	15
πυ	-68	5

top 3-grams, by **interest**

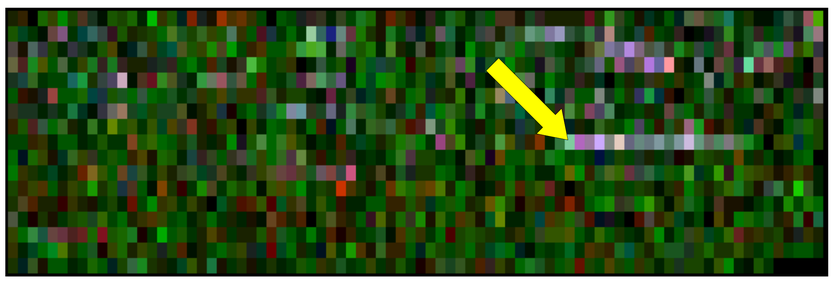
n-gram	interest	sdev(s_r)
ιττπ	-163	7
hιπ	-162	8
δυσ	-139	6
οδυ	-135	11
εκτ	-109	11
τρω	-101	11
πτπο	-100	9
αχι	-98	5
μαχ	-94	5
χιλ	-94	4

Example I: three 2-grams for "horse"

Two of the most interesting 2-grams, **hι** and **ττ**, are part of the word ἵππος, "horse."[6]
To these we added **ιπ**, which had a lower **interest** value.

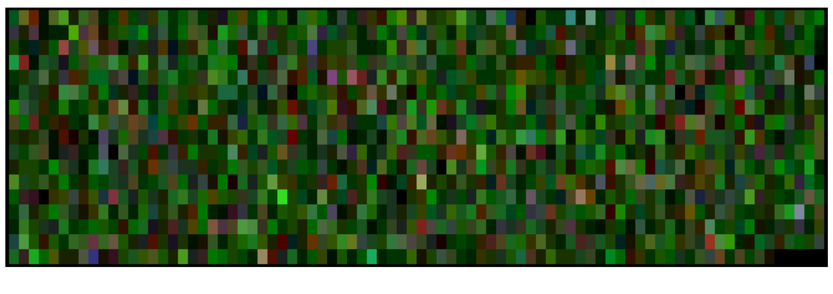
n-gram counts were scaled and translated into color values:

red	**hι**
green	**ιπ**
blue	**ττ**



original

The large bright region in series **original** corresponds to a set-piece, the chariot race held during the funeral games for Patroclus beginning at Iliad 23.259.



r0

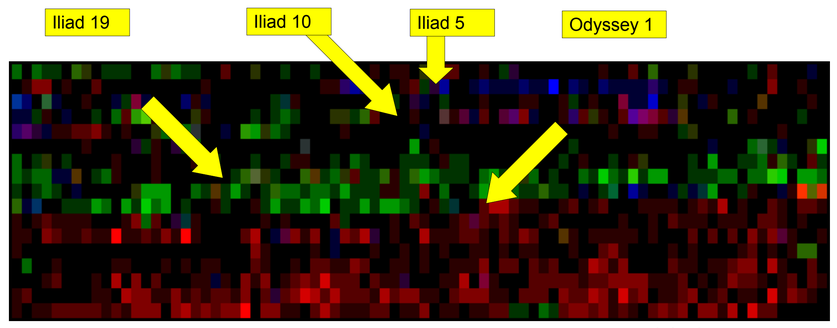
Example II: Three Heroes

Next we consider three 3-grams related to three independent content elements—each is a component of a Greek hero's name:

red	**δυσ**	Odysseus
green	**χιλ**	Achilles
blue	**τυδ**	Diomedes[5]

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

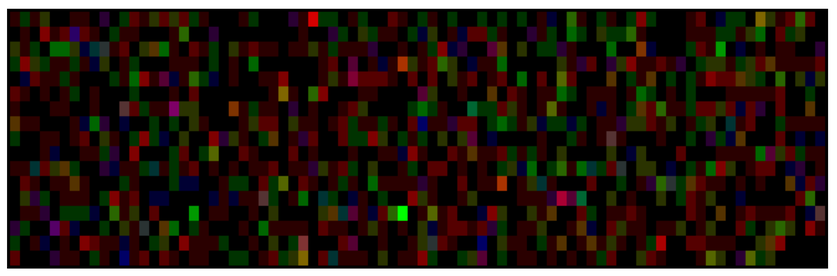
- Achilles in Iliad 19 and following
- Odysseus in the Odyssey
- Diomedes in Iliad 5



original

Note the purple section at Iliad 10. This is the "Night Raid" in which Diomedes and Odysseus worked together.

Not only do the colors show independent story elements; they also can represent their interactions.

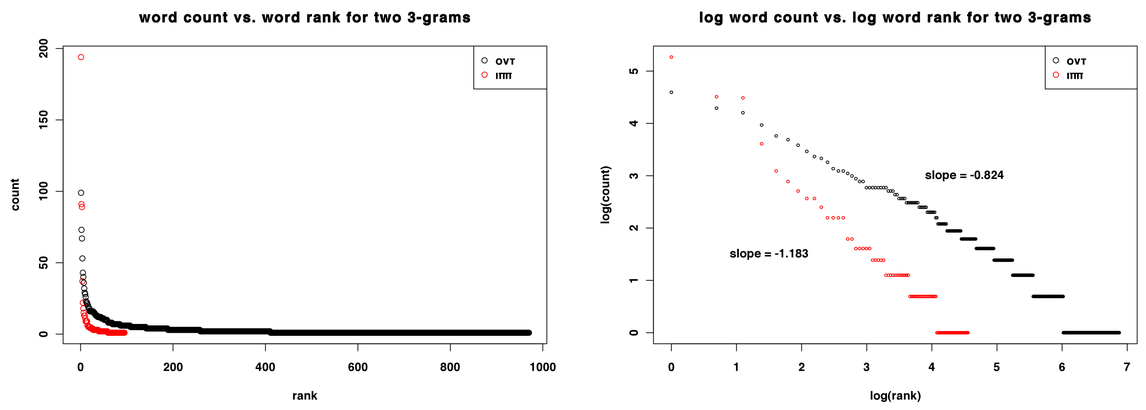


r0

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 **ιπτ** and **ovτ** both frequent and of high **interest** values.



These graphs 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.

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

top words containing **ιπτ**

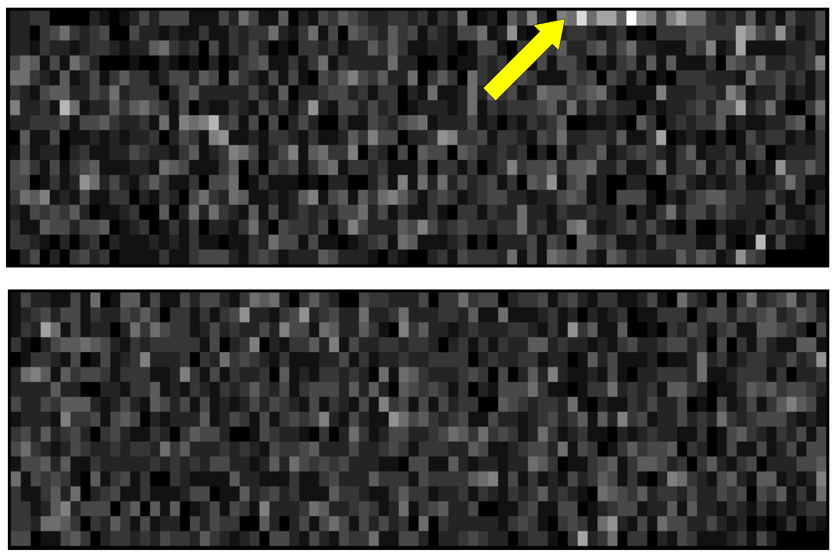
rank	count	word	meaning
1	194	ιπποuσ	"horses" (acc. pl.)
2	91	ιππων	"horses" (gen. pl.)
3	89	ιπποι	"horses" (nom. pl.)
4	37	ιππoτα	"horseman"
5	22	ιπποδαμοιο	"horse-tamer" (gen. s.)
6	18	ιπποισιν	"horses" (dat. pl.)
7	15	ιππον	"horse" (acc. s.)
8	13	ιπποισι	"horses" (dat. pl.)
9	13	ιππω	"horses" (nom./acc. d.)
10	11	ιππηλατα	"pertaining to horse-driving"

top words containing **ovτ**

rank	count	word	meaning
1	99	εοντα	"being"
2	73	ποντον	"sea" (acc. s.)
3	67	ηεπoντο	"they followed"
4	53	εχοντεσ	"having" (nom. pl.)
5	43	ποντωι	"sea" (dat. s.)
6	40	ηικoντο	"they arrived"
7	36	εγενoντο	"they were"
8	32	εχοντα	"having" (acc. s., nom./acc. pl.)
9	29	εοντεσ	"going"
10	28	γεροντοσ	"elder"

Example III: A Content-Independent Pattern?

Here, we used shades of grey to represent counts of a single 3-gram, **ovτ**. The bright region corresponds to the "Catalogue of Ships" inset piece at Iliad 2.484–761.



original

r0

Notes

1. Peabody, B. (1975) *The Winged Word: A Study in the Technique of Ancient Greek Oral Composition as Seen Principally Through Hesiod's "Works and Days."* Albany: SUNY Press.
2. Noted already by Packard, D. W. (1947) "Sound Patterns in Homer," *Transactions of the American Philological Association* 104:239–260.
3. We take our inspiration in part from Plamondon, M. (2009) "Computational Phonostylistics: Computing the Sounds of Poetry," presented at DHCS 2009.
4. Perseus Digital Library Project. Ed. Gregory Crane. http://www.perseus.tufts.edu. Accessed 2/1/2010.
5. By way of his patronym, "Son of Tydeus"
6. In composing the n-grams, we transcribe rough breathing as Latin **h**; iota subscript as adscript (regular **ι**); and final sigma with medial sigma as **σ**.