# Computational Analysis of Alleged Homeric Texts

Grant Storey
2017
Advised by: Christiane Fellbaum

Submitted to: Princeton University
Department of Computer Science

This Thesis represents my own work in accordance with University Regulations

Date of submission: May 4, 2017

## Abstract

*Questions surrounding the Greek epic poet Homer have fascinated scholars for two and a half millennia. There has been much ink spilled on who he was, what he wrote, or whether he was even a single person, all backed up by hours of painstaking manual work. The types of tools that might allow a computational analysis, like a metrical scanner or dialect analyzer, are currently not accurate enough and do not provide enough feature extraction to be useful for this task. In order to use computational techniques to contribute to the discussion of these questions surrounding Homer, we create two new tools. The first, ᾠδικόν[1], scans dactylic hexameter and extracts relevant features about the meter of the line using two approaches, one inspired by modern students of ancient Greek and the other by ancient Greek speakers themselves. The second, τάμνον[2], uses a rules-based approach to analyze the dialect of ancient Greek words. We show that these tools are highly accurate and improve on previous tools with similar goals, and then use them to analyze a variety of allegedly Homeric and known Non-Homeric texts. These analyses provide support for some existing hypotheses about the relative dating of certain Homeric Hymns, insight into which features characterize the Iliad and the Odyssey, and evidence against common thinking on which of the books of these texts are unusual.*

# 1 Introduction

The epic poet Homer has been a source of fascination for two and a half millennia, from Greek city-states of the 5th century B.C.E. maintaining their own copies of his works, through the Roman Emperor Hadrian asking the Oracle at Delphi about Homer's birth in the second century C.E. [29] and up to continuing scholarship in the modern day. Tied into this interest in Homer, however, are disagreements about which texts he actually authored. The Greeks of the classical period attributed to him both the Epic Cycle, a series of epic poems telling the story of the Trojan War, as well

[1]Pronounced "odikon," this means "the thing that sings" or "the musician" in ancient Greek, as ᾠδικόν extracts the musical rhythm of these ancient poets.

[2]Pronounced "tamnon," this means "the thing that divides" in ancient Greek, as τάμνον divides words into various dialects. Also, τάμνον is the common Greek (Ionic, Aeolic, and Doric) form of the more familiar Attic τέμνον, so the name itself is a reference to the dialects it aims to analyze.

as the *Homeric Hymns*, prayers to specific gods which still bear his name today. The modern consensus leaves him, at best, with authorship of only the *Iliad* and the *Odyssey*. Previous studies on these questions of authorship, however, have generally been achieved through manual work and analysis of the texts rather than leveraging computational techniques. These manual analyses involve choosing a limited set of interesting features, counting the frequency of these features by hand in a few texts, and then considering the results.

The goal of this paper is to contribute to these conversations on authorship by providing a computational analysis of ancient Greek-specific features of texts attributed to Homer as well as a variety of control texts. By using computational techniques, we can examine a far wider number of features and texts than can be reasonably done with a manual analysis. We aim to validate the hypothesis that these features will provide clear differentiation between the pair of the *Iliad* and the *Odyssey* and other texts, both those attributed to Homer and those that undoubtedly had different authors. We also hope to get a grasp on what features are more "Homeric" than others, with three hypotheses: that Homeric texts will contain more dactyls, avoiding consecutive spondees in particular because the stress spondees put on the meter would have been clearer in the oral poetry of the *Iliad* and the *Odyssey*; that they will show a unique mixture of Ionic and Aeolic dialect features; and that they will show more consistent evidence of the digamma (these points will all be explained in Section 2). Lastly, we hope to show that it is difficult to clearly differentiate the books of the *Iliad* from the books of the *Odyssey* and therefore they are unlikely to have been written by two very different authors. There has been much work done on this topic among classical scholars, but no studies so far have utilized computers to analyze a wide variety of features across these texts. Existing computational tools for extracting these features have not been applied to this area, and they suffer from poor performance and a variety of other drawbacks.

This paper makes three contributions. The first, ᾠδικόν, is a tool for scanning lines of dactylic hexameter and extracting metrical features from them. The second, τάμνον, determines various dialect features associated with ancient Greek words. After presenting and evaluating these tools, our final contribution is to use the two tools together to extract metrical and dialect features from

both the alleged Homeric works and other hexameter texts from a variety of authors and time periods, allowing us to create author fingerprints and use these as evidence in the ongoing debates about Homer. Using this analysis of metrical composition and dialect usage, we will provide some evidence against the Homeric authorship of the *Homeric Hymns*, showing in particular that the shorter Hymns and the Hymn to Hermes are clearly non-Homeric, and weigh in on the very existence of Homer as a single individual.

# 2 Background

## 2.1 The Texts and Homer

For the purposes of this work, we divide texts into three main categories.

- *Homeric Texts*: This category contains two poems from the Epic Cycle, the *Iliad* and the *Odyssey*, which contemporary scholars generally attribute to Homer.
- *Pseudo-Homeric Texts*: This category consists of the *Homeric Hymns* and the remaining books of the Epic Cycle, all of which were attributed to Homer in antiquity but to other authors in the modern day. All that remains of the Epic Cycle are fragments from quotations in ancient commentaries, summing to a few dozen lines across all the texts. Since this is such a small sample size, we will not examine any of these fragments as part of this analysis and will focus on the *Homeric Hymns*.
- *Non-Homeric Texts*: These are texts that were never attributed to Homer but are still Greek poems composed in the meter of epic, from the very early works of Hesiod to Nonnus' massive *Dionysiaca*, written in the 4th or 5th century C.E.

One further wrinkle surrounding Homer and authorship is the question of whether Homer existed as a single, individual poet who composed the *Iliad* and *Odyssey* or whether, in fact, these poems were composed incrementally over a long period of time by a series of bards modifying and expanding on the work of their predecessors [19].

## 2.2 Dactylic Hexameter

Ancient Greek poets composed their poetry in specific meters, repeating certain patterns of long and short syllables with slight variations, similar to the way Shakespeare composed his work in iambic pentameter (for example, "In **fair** Ve**ro**na **where** we **lay** our **scene**" with the alternation of unstressed and **stressed** syllables). The meter of epic poetry, including the works of Homer, is dactylic hexameter. It consists of six feet: the first five are either a dactyl (a long syllable followed by two short syllables) or a spondee (two long syllables), and the final foot has two syllables, the first long and the second of either length (See Figure 1). Determining the syllable lengths and the feet present in a line is called "scanning" the line and the pattern for a specific line is the "scansion" of that line.

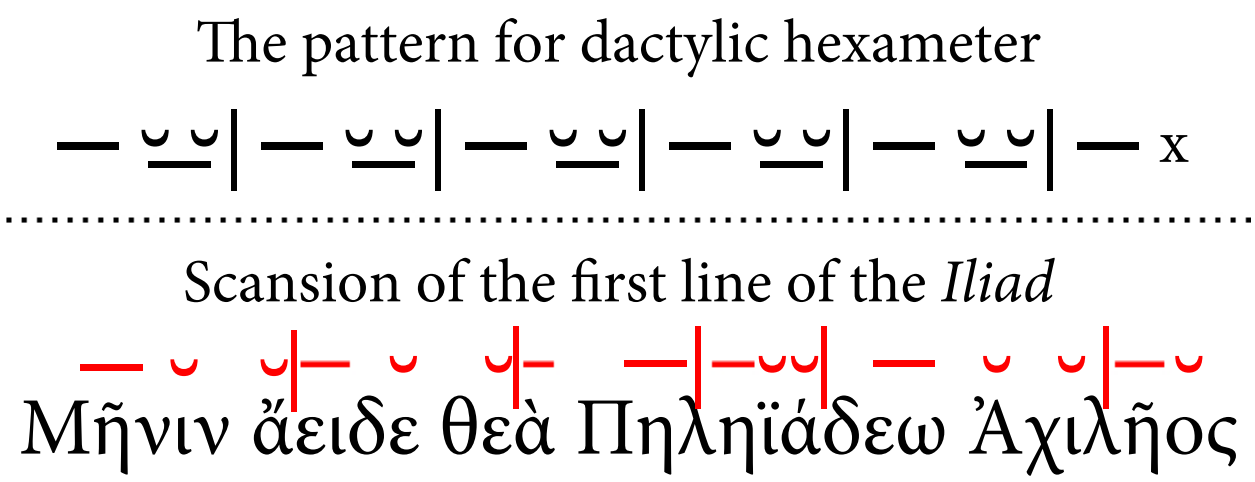


**Figure 1: The pattern for hexameter and an example of a scanned line.**

## 2.3 Ancient Greek Dialects

The ancient Greek world of classical antiquity, before the unification under Alexander of Macedon, spoke a variety of different dialects, which fell into a few broad families. Two in particular interest us here: the Attic-Ionic family, which includes the Attic dialect spoken during the Golden Age of Athens and Ionic dialects spoken all around the Aegean Sea, and the Aeolic family, including the dialect spoken by Sappho on Lesbos and the dialect spoken in Boeotia, just north of Athens. The

works of Homer are written in what is frequently called the "Homeric Dialect," which no person ever spoke as their native dialect. The "Homeric Dialect" consists of a combination of dialect forms, mostly Ionic with some Aeolic and a variety of features that are archaic compared to the Greek of the 6th and 5th centuries B.C.E., which is the time period most commonly considered when discussing dialects. We also briefly note that these broader categories are not universally accepted. Ionic can be divided into East, Central, and Western categories, each with individual features, and the three Aeolic subfamilies (Boeotian, Lesbian, and Thessalian), while they share many characteristics, are also different enough that some argue whether the category of "Aeolic" should even exist at all [36]. For the purposes of this work, we assume that these categories both exist and are useful for classification of texts. For the rest of this work we will use "Ionic" and "Aeolic" under the assumption that they are broad categories with a well-defined set of features common to the various dialects included within them.

For examples of a few differences between the Ionic and Aeolic dialects, see Figure 2.

| Ionic | Aeolic | English |
|---|---|---|
| τέχνης | τέχνας | "of skill" |
| ἡμεῖς | ἄμμές | "we" |
| μία | ἴα | "one" |
| ἔθεσαν | ἔθεν | "he put" |

**Figure 2: A few examples of differences between Ionic and Aeolic.**

## 2.4 Additional Relevant Information and Terminology

### 2.4.1 Ancient Greek

1. When compared to a language like English, ancient Greek is very morphologically complex. An English noun might have two forms, singular and plural, such as "foot/feet," while an ancient Greek noun has an inherent gender (either masculine, feminine, or neuter) and can change form not only between singular, dual ("two feet"), and plural, but also for five cases (nominative, genitive, dative, accusative, and vocative), totaling 15 forms expressed through various endings added to the stem. Ancient Greek verbs can take even more forms. We call the formation of a specific form of a word **conjugation**.
2. When a set of nouns in ancient Greek are all conjugated in a similar manner, we often refer to all of them as part of a single group; for example, the words πόλις, "city," and δύναμις, "power," both have the same set of endings in each case and number (the singular endings are -ις, -εως, -ει, -ιν, -ι in the Attic dialect), and the nouns' stems originally ended in the letter iota, so they are called "iota-stems." Other groups are similarly named after the last letter of their stem, e.g., "sigma-stems" or "digamma-stems."
3. Another important piece of Homeric composition is the digamma (ϝ). This letter corresponds to the "w" sound in "**w**ater," which was present in early Greek but had fallen out of Ionic by the time the poems of Homer were written down. However, it still impacts the meter and perhaps was pronounced in some way [1]. In order to properly scan a variety of lines in the text, this digamma must be "restored." For example, in line 33 of *Iliad* book 1, one must restore the original form ἔδϝεισεν for the written ἔδεισεν to produce proper meter. An initial digamma also prevents elision of short syllables (where a final short vowel is not pronounced when the following word begins with a vowel) and correption of long syllables (where a final long vowel or diphthong is scanned short when the following word begins with a vowel). For example, in ἀγαθῷ ἄναχτι, "to the good leader," the final -ω of ἀγαθῷ would not suffer from correption because of hiatus caused by the invisible digamma at the start of (ϝ)ἄναχτι. However, in works

from later periods the echoes of the digamma were no longer expressed. Using instances of hiatus or changes in the meter, plus comparative evidence from other Indo-European languages, modern scholars are able to reconstruct the digamma in a variety of words, so the dictionary we use for determining vowel lengths in later sections also includes information on the digamma.

### 2.4.2 Natural Language Processing

1. A **token** is an individual occurrence of a word within a text. For instance, line 1 of the *Odyssey*, "ἄνδρα μοι ἔννεπε, μοῦσα, πολύτροπον, ὃς μάλα πολλὰ,"[3] has eight tokens, the first of which is ἄνδρα, "man."
2. Since a single ancient Greek lexicon entry could be conjugated into many different forms, we refer to the base dictionary entry behind a form as the **lemma.** For example, κακός, "bad," is the lemma for the forms κακοί, "bad (men)," and κακῇ, "to a bad (woman)," as well as any other form of the word.
3. A **parse** of a token consists of that token's lemma and relevant morphological information about the token. For example, given the token "ἐμή," "mine (feminine)," one valid parse consists of the lemma ἐμός, "mine," and the fact that ἐμή is the "singular feminine nominative adjective form of ἐμός." A given token can have multiple parses; ἐμή, for example, could be one of two cases (the nominative or the vocative).

# 3 Related Work

There is a large body of work in the classical scholarship concerning the question of Homer's existence. Graziosi and Burgess analyze the reception of Homer in antiquity [21][10]. Fowler outlines the history of scholarly discussions of Homer and current conceptions of his identity [19]. Jensen argues that the *Iliad* and the *Odyssey* were oral compositions of sixth century B.C.E. Athens [33]. Sherratt uses references to armor, weapons, and other items that can be dated in the

[3] "Tell me, muse, of the crafty man who [wandered through] many places..."

archaeological record to show that the *Iliad* contains elements from a variety of time periods from perhaps as early as the 16th century to the late eighth century B.C.E., arguing for a multi-step composition [45]. This is, of course, a small sample of the volumes written on this question.

Some features we examine in this work have also been used as part of manual analyses in the past. For metrical analysis, there are a variety of papers analyzing specific metrical features within the text of Homer and other Greek authors. Beekes breaks down a variety of common structural patterns found within hexameter in the *Iliad* [4]. Mojena examines the behavior of the so-called "Mute + Liquid" rule in Theocritus to show that prepositives were closely considered with their following word, unlike true word breaks [37]. Bulloch proposes a new "law" for the hexameter of Callimachus [9]. Clayman and Nortwick use a random sample of 13 works to show that enjambment does not allow accurate dating of various hexameter works, while Barnes uses a different methodology to challenge their results and interpretation [13][3]. Greenberg analyzes word breaks, metrical feet, and the placement of common words to find key composition differences between book 1 of the *Iliad* and the hexameter portions of the works of Theognis and Solon [22]. Companions to Homer also frequently include detailed explanations of the specifics of Homeric hexameter [52][5].

The works specifically examining dialect features in Homer generally take the form of grammars or comprehensive works on the subject [38][36]. Fick attempts to translate the entirety of the Aeolic forms in the *Iliad* and the *Odyssey* into Ionic forms to recover what he believes to be the original form of the text [18]. We also find dialect used in arguments surrounding hypotheses on the chronology of the Homeric poems, such as Bolling's work on the introduction of Ionic dialect forms [7]. Again, companions to Homer are a valuable source for information on the complexities and specifics of dialects in the works of Homer [30][5]. Buck's *The Greek Dialects* contains a comprehensive breakdown of the dialect differences in the classical period, with occasional mention of the features in Homer as well [8].

Janko also analyzes morphological features of Homeric texts in order to establish their relative dating, including the usage of computers to assist with creating tables of this data, though much of this data collection was done in the early 1980s with programs written on punch cards [31, 32].

Jones analyzes Janko's data from a slightly different perspective to argue against the existence of an "Aeolic phase" of epic [34].

From a computational standpoint, there are a variety of previous attempts at both scansion and dialect analysis of texts. Eder uses spectral density analysis of hand-scanned data from 10 samples of hexameter text (7 Greek, 3 Latin) to provide evidence for his argument that older hexameter is more strongly rhythmic [16]. Fusi creates a general-purpose scanner, which aims to scan a variety of types of Indo-European poetry, including not only Greek and Latin but also languages like Sanskrit, losing out on some accuracy on Greek texts for this broader applicability [20]. There is a Greek hexameter-specific analyzer by Papakitsos, but it makes a few simplifying assumptions that do not properly cover the range of oddities and difficulties of hexameter [43]. The Classical Language Tool Kit includes a scansion module, but it is designed to work on all Greek text and does not account for many of the complexities of dactylic hexameter [17]. There is also a very good scanner from Vilnius University, but even this scanner is not perfect, missing a variety of corner cases and overzealously applying others [49].

The only major tool for analyzing Ancient Greek forms and potentially determining their dialect is the Perseus Project's Morpheus [15]. This tool has many important features, especially its morphological parsing of Ancient Greek tokens, but in terms of analyzing the dialect of given tokens it suffers from some serious limitations. Its dialect marking is rather inconsistent: it generally marks non-Attic forms with the appropriate dialect and provides no dialect for canonically Attic forms, but where the form only shows peculiarities in Attic it marks the general form with no dialect and the Attic form as "Attic." In some cases, it gives no dialect marking to universal forms, but for other forms it marks them as part of every dialect. Morpheus also provides no reasons for a given dialect choice, leaving it up to the user to determine why a form might be considered "Doric" or "Poetic." To address many of these issues, we previously created a version of the τάμνον tool for differentiating between the Attic and Doric dialects for analysis of the plays of Euripides [48].

# 4 Approach

In broad strokes, our process consists of three stages:

- ᾠδικόν scans the lines and extracts metrical features from them.
- τάμνον analyzes the tokens of the text and extracts dialect features from them.
- We run analyses on the features extracted by these two tools, comparing the Homeric texts against the Pseudo-Homeric texts, Non-Homeric texts, and even against themselves to look for differences between the two poems and between books of each individual poem. The specifics of this are left for Section 8.

## 4.1 Metrical Features

Feature extraction from a fully scanned line follows a set of well-defined rules for each specific feature and is reasonably straightforward. The greater technical challenge with this process is properly scanning the line. In order to provide the most useful feature extraction, we aim to improve on previous scansion techniques and reach 99% of lines successfully scanned[4]. We take two broad and potentially complementary approaches to scansion.

- The first approach is based on the technique of a modern student of Greek poems and is therefore termed the *Student Approach*. This approach divides the line into syllables, marks obvious lengths (an eta in the middle of a word is always long, an omicron followed by an eta is always short) and then uses this partial set of lengths and knowledge of the structure of hexameter to determine the remaining lengths, though in rare cases this produces multiple possible scansions.
- The second approach is based on the technique an ancient Greek speaker would have used and is therefore termed the *Native Speaker Approach*. This approach uses a dictionary of lemmas to determine the lengths of every vowel in the line, which a native speaker would have known inherently when speaking the poem out loud. With this information, the composition of the entire line should be known without the need for any guessing.

[4]To look ahead for a moment, we manage to reach 98.4%.

## 4.2 Dialect Features

The key idea for this approach is to use a rules-based method for characterizing the dialect of a specific token. Although there are often benefits to a probabilistic model for classification, in this specific application there are variety of reasons that a rules-based method is more appropriate.

The first big issue with a probabilistic method is training data: there is a comparatively small corpus of Aeolic and Ionic texts, and many of the texts will be used as Non-Homeric comparisons, so the need to separate training and test data would lead to even smaller corpora. It would be quite difficult to come up with a reliable classifier given such a small corpus, and the program's definition of the dialects would be based on the choice of authors used. On the other hand, a rules-based approach does not require any training corpora, and Carl Buck's book *The Greek Dialects* [8] includes a list of features of every dialect. So there is a proper list of rules, making a rules-based approach appealing.

A major benefit of the rules-based approach is that it provides a ready-made set of features to examine. Instead of extracting a large variety of features and trying to see which ones identify the dialects, then re-using those for the dialect feature analysis, the rules-based approach allows us to use the frequency of each rule as its own feature in fingerprinting authorship.

Lastly, a rules-based method can be converted to a probabilistic method by converting each of the rules into features, so this tool can always be extended from being rules-based to being probabilistic.

We also note that there are some inherent difficulties in classifying the true dialect of tokens in this text due to the combinations of different dialects and the presence of archaic and classical forms. For example, from a classical perspective the form τόσος is Ionic and τόσσος is Aeolic, but τόσσος is also the ancient Ionic form of the word, so it may simply be the preservation of an older form of the text rather than a sign of "Aeolic" influence [30]. We recognize this difficulty and address it in two ways. First, where possible, we mark such a feature as *both* Aeolic and Homeric (Archaic) to cover both possibilities. Second, we include an analysis of the individual rule outcomes rather than just the overall dialect footprints. This will allow us to differentiate between an author using classical Ionic and one using Archaic Ionic or Aeolic, which provides some useful information.

# 5 Implementation

This section is divided into three parts; the first describes the general preprocessing of the texts, the second ᾠδικόν, and the third τάμνον.

## 5.1 Preprocessing

The two tools share a common preprocessing step. This preprocessing cleans the data and runs a morphological parse on each token, determining information about the individual tokens and lemmas behind those tokens. This information is then used by both tools to assist in generating metrical and dialect features. The individual steps are described below:

1. **Cleaner**: This runs over the input data and cleans it up, removing punctuation and certain features of the text, like capitalization, which interfere with morphological parsing. It produces a version of the input file with all of the tokens cleaned.
2. **Morphological Parser**: In order to determine more information about a token, we must know the lemma of that token and information about the lemma. For example, is the token πολιτῶν a verb, adjective, noun, or exclamation? What is the associated lemma, and how does one conjugate lemmas of this sort? In this case, πολιτῶν is "the genitive plural form of the masculine α-stem noun πολίτης, 'citizen.'" The way for a computer to determine this information is with a morphological parser. Designing our own Ancient Greek morphological parser is beyond the scope of this project, so we take advantage of Perseus' Morpheus, an online morphological parser [15]. Programs can send requests to the Perseus server for the parse of a given token, and the server will return all parses provided by Morpheus, with the appropriate lemma and morphological information. The problem with relying on Morpheus is that there is no simple fallback for when it fails. This means that without a massive hand-maintained dictionary of exceptions, the tools must ignore tokens that Morpheus cannot parse. The most common errors occur on names, like Γοργονες, "Gorgons," and compounds, like βαρύδουπον, "heavy-thud." In

some rare cases, however, the errors occur on tokens that clearly show Aeolic features, like αὕτα, "herself," with its non-Ionic long α ending. Nevertheless, Morpheus returns parses for the vast majority of tokens[5], so these occasional failures are acceptable.

In order to get all the necessary morphological parse information, we first run a Morpheus query for every unique token within the input text and store all of the parse results in the Form Info file. However, there is extra conjugation information about some nouns and adjectives that is necessary for proper form analysis, so after the first round of queries we run through lemmas matching the profile of these certain types. For each match we run a second query to determine whether it is actually this type. For example, the noun with lemma πόλ-ις, "city," is an iota-stem whose genitive singular is πόλ-εως, while the noun with lemma χάρ-ις, "grace," is not an iota-stem and its genitive singular is χάρ-ιτος, but the lemma alone may not tell us which of these types a noun ending in -ις is. Morpheus does not provide stem-type information, so we assume that every token ending in -ις is an iota-stem, which means that its Attic genitive singular would end in -εως. We then query Morpheus again with the -εως form to determine if our hypothesis about the lemma is correct; πόλεως returns a match, so we know πόλις is an iota-stem, while χάρεως returns no valid parses, so we know that χάρις is not an iota-stem. We output this supplementary information to the Lemma Info file.

## 5.2 Implementation of Metrical Analysis

The metrical analysis tool is called ᾠδικόν. The flow of control for both the Student Approach and the Native Speaker Approach is as follows (see Figure 3 for the overall flow and Figure 4 for an example of the steps on a single line):

1. **Phoneme Division**: The line is divided into individual phonemes, preserving diphthongs and the like. For example, οἰωνοῖσι ("by the birds of prey") is divided into οι.ω.ν.οι.ς.ι. The Native Speaker Approach also analyzes the lemma of each token and uses the dictionary entry for that lemma to add known inherent lengths (necessary for alphas, iotas, and upsilons) to both the stem

[5] e.g., for the *Iliad*, Morpheus successfully parses 20,253 out of 20,433 tokens for a success rate of 99.1%

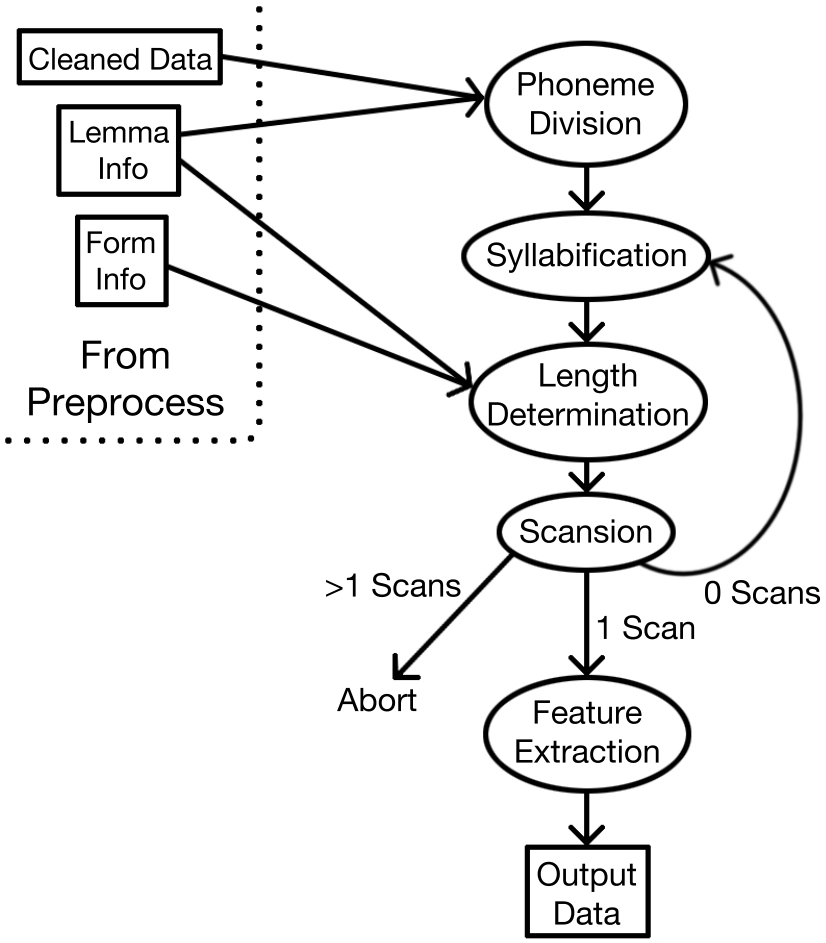


**Figure 3: The control flow for ᾠδικόν's Student and Native Speaker Approaches.**

and the ending, so ριγήσω (I will shudder) becomes ρ.ι[long].γ.η.ς.ω. It also adds digammas, so ἔργον becomes ϝ.ε.ρ.γ.ο.ν (with the initial digamma restored).

2. **Syllabification**: The line is divided into syllables using a two-pass system. First, the phonemes are grouped into consonants-vowel sets, so each group consists of the 0 or more consonants between the previous vowel and the next vowel, plus that next vowel. For the second pass, if a set begins with multiple consonants, the first is moved to the end of the previous syllable, making it a closed syllable. For an example, see the progression from steps (2) to (4) in Figure 4, and note that the lambda has moved from the start of the 12th consonant-vowel grouping in (3) to the end of the 11th syllable in (4). If an earlier attempt at scansion of this line failed and the program returns to this step, it will apply exceptional rules, like combining -εων clusters into a single long syllable, to attempt to produce a valid scansion.

## Example of the flow of ᾠδικόν for the second line of *The Iliad*

(1) Original Line:

οὐλομένην, ἣ μυρί᾽ Ἀχαιοῖς ἄλγε᾽ ἔθηκε,

(2) Phonetically Divided:

ου.λο.με.ν.η.ν.η.μυ.ρι.α.χαι.οι.ς.αλ.γ.ε.ε.θ.η.κ.ε

(3) Consonant-Vowel Groupings:

ου.λο.με.νη.νη.μυ.ρι.α.χαι.οι.ςα.λγε.ε.θη.κε

(4) Syllabificated Line:

ου-λο-με-νη-νη-μυ-ρι-α-χαι-οι-ςαλ-γε-ε-θη-κε

(5a) Length Determination (Student Approach):

— ◡ ◡ — — ? ? ? — — — ◡ ◡ — ◡

ου-λο-με-νη-νη-μυ-ρι-α-χαι-οι-ςαλ-γε-ε-θη-κε

(5b) Length Determination (Native Speaker Approach):

— ◡ ◡ — — — ◡ ◡ — — — ◡ ◡ — ◡

ου-λο-με-νη-νη-μυ-ρι-α-χαι-οι-ςαλ-γε-ε-θη-κε

(6) Proper Scansion:

— ◡ ◡|— —|— ◡ ◡|— —|— ◡ ◡|— ◡

ου-λο-με-νη-νη-μυ-ρι-α-χαι-οι-ςαλ-γε-ε-θη-κε

**Figure 4: Example of ᾠδικόν's process on the second line of the *Iliad***

3. **Length Determination**: This module attempts to determine a length for each syllable using a variety of rules. For example, a closed syllable is always long, and an epsilon (ε) in an open syllable is always short (see Appendix A.2 for the full explanation). The Native Speaker Approach also uses lemma data to determine the natural length of alphas, iotas, and upsilons in given tokens. If a length cannot be determined, it is left unknown. For example, in Figure 4 (5a), the Student Approach can determine lengths for a variety of the syllables, but is unable to determine the lengths of the upsilon, iota, and alpha in the middle of the line. The Native

Speaker Approach, with its additional dictionary knowledge, would be able to fill in all three of those lengths using the dictionary entries for μυρίος and Ἀχαιός in Figure 4 (5b).

4. **Scansion**: Given the list of syllables and their known/unknown lengths, an attempt is made to fill in the missing syllables such that they fit the valid patterns of hexameter, using a dynamic programming method to build up valid partial scans. Ideally, the Native Speaker Approach would have a known length for every syllable, but a variety of features like hiatus, correption, and ictus lengthening (see Appendix A.2) can occasionally lead to confusion in this area. If a single valid scansion is found, the program proceeds to the next step. If multiple possible scansions are found, the program aborts as the correct scansion cannot be determined. If no possible scansions are found, the program returns to the Syllabification step with the knowledge that this is a later pass and more exceptions must be applied.
5. **Feature Extraction**: Once a single valid scansion has been found for the line, the program uses the original text and the scanned line to extract a variety of relevant features. Some features are simple: for example, is the fifth foot a dactyl or a spondee? Some are more complicated: for example, is Meyer's first law observed (that is, if the second foot is a dactyl, there is no word break between the two short syllables)? See Appendix A.3 for a detailed list of metrical features extracted.

## 5.3 Implementation of Dialect Analysis

Our dialect analyzer is called τάμνον. For an overview of the control flow, see Figure 5.

The dialect analyzer takes as input the cleaned data, form and lemma information for that data, and the list of rules, then applies the rules to each token to determine its dialect. The main challenge for the dialect analyzer is that a specific token can have multiple parses with different dialect analyses for each parse. Since determining the proper parse for a token is an area of open research and there is no simple way to access resources that have this data, we keep track of the minimum and maximum counts for each piece of information tracked. For example, the token ἦν could be the Ionic form of a conjunction meaning "if," the Attic/Ionic third singular imperfect of

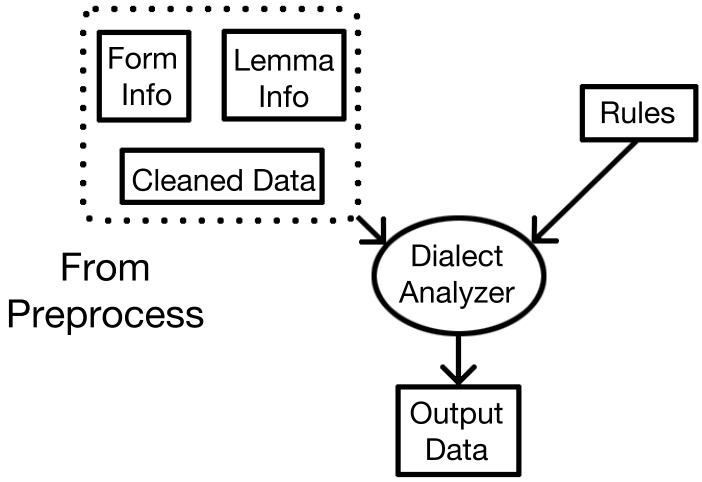


**Figure 5: The control flow for** τάμνον**.**

the verb εἰμί meaning "he/she/it was," or the Aeolic/Doric/Archaic third *plural* imperfect of the same verb, meaning "they were." This token would increase the maximum possible number of Ionic forms by one, but would not increase the minimum number. The token ϑερσει, "courage," on the other hand, has many parses, but all display the Aeolic ϑ[ε]ρσ- (for Ionic ϑ[α]ρσ-), so this token definitely exhibits an Aeolic dialect feature and would increase both the maximum and minimum count of Aeolic features by one.

The analyzer needs to produce four major data sets, using rules that identify Ionic features, Aeolic features, and "Homeric" features[6]:

1. The maximum and minimum number of tokens that have features of each possible combination of dialects (that is, "Definitely Ionic and definitely not Aeolic" is different from "Definitely Ionic, possibly Aeolic").
2. The maximum and minimum number of tokens that have features of each dialect, as well as the tokens associated with each of these categories.
3. The number of possible and definite matches of each dialect for each of the rules in the list.
4. The parses, with the dialect for that parse and reasons for the choice, for each of the tokens.

To create these datasets, we use a 3-level loop, which runs through each token, each rule, and then

[6]The "Homeric" dialect refers to archaic features like the genitive singular ending -οιο that appear in older works like those attributed to Homer.

each potential parse for the token, keeping track of the necessary information for each dialect, rule, token, and parse. At the end, τάμνον recombines the aggregate information into the necessary datasets.

While it would be more efficient to examine each unique token rather than every occurrence of each token, we include this capability so that future versions of τάμνον can potentially include information based on the token's context.

# 6 Data

## 6.1 The Texts

The Perseus Digital Library has online copies of both the *Iliad* and the *Odyssey*, the *Homeric Hymns*, and 15 other Greek hexameter works from various time periods with a total of 50,820 lines among the 15 texts. For more information on the texts analyzed, see Appendix A.1.

It is important to note that there are two major sources for potential differences between the digitized text and what was actually composed or written by the original author. The first source of errors is in the transmitted manuscripts: ancient scholars could have improperly attempted to correct a form they saw as incorrect and the copyists who wrote the manuscripts passed down to us may have copied forms incorrectly. There are a variety of modern modifications and small disputed sections, but for the texts we are analyzing these are generally just a small section within one line, and are unlikely to drastically change the meter, so we leave them as is.

The second source of errors is modern editors themselves modifying the manuscripts to include a form they believe to be correct. An example of this is editors who "restore" the earlier genitive singular -οο to replace the later genitive singular -ου.

Both of these types of changes could lead to digital texts that do not match their originals in the distant past, and if any of these changes showed a bias towards one type of metrical or dialect feature it could lead to a corresponding bias in the results. Because we cannot go into the past and determine the original forms of texts before the manuscript errors, and there are no digital version

of the transmitted manuscripts, we cannot fix these errors within this project. Despite these caveats, we make the assumption that the digital texts represent the ground truth. This approach may not capture all the intricacies of the real world, but at least it's clean and allows simple analysis without introducing our own complex assumptions. Our conclusions will hold true concerning the input text presented, and the tools could be used on the original manuscript texts or later, corrected versions of the text if they become available in digital form.

## 6.2 Scansion Rules

The various rules and exceptions (e.g., the double-consonants at the start of Σκάμανδρος are counted as a single consonant for metrical purposes) are taken from Benner's *Short Homeric Grammar* and West's *Homeric Meter* [5][52]. See Appendix A.2 for a description of the metrical rules observed and Appendix A.3 for the list of metrical features extracted (also based on explanations in Benner and West).

## 6.3 Dialect Rules

The rules for determining whether a token shows Ionic or Aeolic dialect features were taken from Carl Buck's book *The Greek Dialects* [8], and rules for determining archaic features were taken from Benner's *Selections from Homer's Iliad* [5]. Specifically, we created rules for each of Buck's features of Ionic and the general Aeolic dialects and Benner's features of the archaic Greek forms found in older works. Rules involving lemmas and forms that were not recognized by the Morpheus parser were removed, as were some rules that required recognizing the correct Proto-Greek form of the stem. For a list of rules, see Appendix A.4.

For the rules involving tokens that Morpheus could not parse, we can assume that they do not appear very often, and therefore excluding them from the set of rules will not have a significant impact on the results. While it would be possible to build a supplementary handler for these features, that is beyond the scope of this project. The rules involving the Proto-Greek form of stems were mostly cases where the Ionic form of a stem contains an η that was a long α in Proto-Greek; however,

it is difficult to programmatically differentiate these new Ionic ηs from ηs that were originally ηs in Proto-Greek. The best way to determine this would be to run through a digital copy of a large Ancient Greek dictionary looking for words whose dictionary entries include an alternative Ionic/Aeolic form, but due to time constraints these rules were not included as part of this paper. While this is a limitation of the tool in its current form, it would be simple to extend the tool to include these additional rules. The issue lies in the rules list rather than the tool itself. Even without the ability to detect these specific forms, the tool can still provide insight about the prevalence of the rules included and provide a general overview of the presence of these types of Ionic and Aeolic forms.

# 7 Evaluation

Before looking at the results, it is important to understand how accurate ᾠδικόν and τάμνον are at their respective tasks. We perform this evaluation here.

## 7.1 Scansion Accuracy

We compare ᾠδικόν's output scansions to ground-truth scansions for the first 100 lines of the *Iliad*, the *Odyssey*, and Callimachus' *Hymns*. For each text, we categorize the number of times that ᾠδικόν's scansion matches the ground truth (Agreement), the number of times that ᾠδικόν's scansion differs from the ground truth (Disagreement), and the number of times ᾠδικόν fails to produce a parse (Failure). The evaluation results for the Student Approach and Native Speaker Approach can be found in Table 1 and Table 2 respectively.

### 7.1.1 Scansion Issues

The Student Approach disagrees with the ground truth on only line 40 of the *Odyssey*, because ignoring the natural lengths of the words produces a valid scansion. However, when one considers the true lengths of the words, it is clear that a slightly more complicated scansion of the line, with

| Text | Agreement | Disagreement | Failure |
|---|---|---|---|
| *Iliad* | 98 | 0 | 2 |
| *Odyssey* | 99 | 1 | 0 |
| Callimachus' *Hymns* | 97 | 0 | 3 |

**Table 1: Comparison of agreement, disagreement, and failures when ᾠδικόν's Student Approach is compared against ground truth data for the first 100 lines of various texts.**

| Text | Agreement | Disagreement | Failure |
|---|---|---|---|
| *Iliad* | 100 | 0 | 0 |
| *Odyssey* | 100 | 0 | 0 |
| Callimachus' *Hymns* | 97 | 0 | 3 |

**Table 2: Comparison of agreement, disagreement, and failures when ᾠδικόν's Native Speaker Approach is compared against ground truth data for the first 100 lines of various texts.**

ictus lengthening, is correct. Since the Student Approach only applies ictus lengthening if it can find no other solution, it assumes that its initial scansion is correct.

The Student Approach fails on 5 lines (1.6%). It fails on line 3 of the *Iliad* and line 83 of Callimachus' *Hymns* because there are multiple possible scansions, depending on what lengths are assigned to various alphas, iotas, and upsilons. It fails on line 33 of the *Iliad* because the word ἔδεισεν is actually to be scanned as ἔδϝεισεν, with a digamma.

The Native Speaker Approach fails on 3 lines (1.0%). It fails on one line that the Student Approach does not: Callimachus's line 33, because Morpheus parses the token ὦνα as a form of ὦνος by the dictionary, giving it a long final vowel, when in fact the token in a contraction of ὦ ἀνα, with a short final vowel.

Both approaches fail on lines 53 and 75 of Callimachus' Hymns because those lines require rather unusual licenses with the scansion, taking both τεύχεα and ἰοκύν as two syllables.

### 7.1.2 Comparison to the Vilnius Scanner

We also compare ᾠδικόν's approaches to the scanner from Vilnius University [49] on the 611 lines of book 1 of the *Iliad*.

Both the Student Approach and the Vilnius scanner fail on 7 lines: lines, 3, 33, 277, 406, 489, 548, and 568. The Native Speaker Approach only fails on lines 277 and 548. The Vilnius scanner fails on an additional 22 lines. Both approaches succeed for every line on which the Vilnius scanner succeeds.

In addition the two scansions differed on 78 of the lines. For every one of these lines, this happens because the Vilnius scanner defaults to applying the rule that εω, εων, εοι, εο, εου, εαι and εα should be scanned as single longs, while by default the Student Approach assumes they should not. In these lines, there are enough vowels of unknown lengths to allow two different valid interpretations. Comparing whether the Student Approach or the Vilnius scanner is correct in each these cases is not particularly valuable, because it is more or less chance whether the inherent vowel lengths support one approach or the other. However, we are confident in making the assumption that the endings should not be scanned as a single long by default because this appears to be the case more generally [5]. For example, in line 472 there is no reason to scan ϑεὸν as a single long unless nothing else will allow the line to scan. The Native Speaker Approach, which uses the natural lengths, picks the correct scansion for these lines.

### 7.1.3 Overall Performance

The Student Approach successfully scans 79,370 out of 80,952 lines (98.05%) across the whole set of texts. The Native Speaker Approach successfully scans 79,207 out of 80,952 lines (97.84%) across the whole set of texts. This means that, on the whole, the Native Speaker Approach actually performs more poorly than the Student Approach. The reason for this is the limitations of the dictionary of lengths used in conjunction with the Native Speaker Approach. Because such a dictionary did not previously exist in digital form, we created our own by processing a digitized copy of Liddell and Scott's dictionary (LSJ), available from Perseus [35]. However, the format of LSJ does not lend itself to easy extraction of lengths; for example, a length may be specified as "short," with a note in English mentioning it is sometimes long in Homer (e.g., the entry for ἵμηι), or there may be oddities like ἐρίηρος, which has a long ι as an adjective but a short ι when used as a

substantive (again this note is buried inside the entry). Hand curating 95,390 different dictionary entries is obviously out of the scope of this project, but we have confidence that the accuracy would increase to around 99.5% with a better dictionary. For example, the first 100 lines of the *Iliad* and the *Odyssey* were successfully scanned with the addition of only 1 hand-curated form (the digamma in δ(ϝ)εισω) added as part of an attempt to include a few common words with digammas. The final issue is a problem of form recognition: if we have the form like εἴσεται, this could either be a form of εἶμι, with no initial digamma, or a form of (ϝ)οἶδα, with a digamma. For now, the program does not assume digammas when it is not sure there is one, but in line 548 of *Iliad* 1 it is in fact a form of οἶδα and the digamma is necessary to scan the line. Using treebanks or similar data to select a most likely form would allow us to circumvent this problem and scan even more accurately, but that is also out of the scope of this project.

Because of the limitation of both techniques, for the results section, we attempt the Native Speaker Approach, and if it fails we fall back on the Student Approach. This technique successfully scans 79,648 out of 80,952 lines (98.39%).

## 7.2 Dialect Identification Accuracy

To evaluate τάμνον, we check τάμνον's results against the dialect results from Perseus' Morpheus [15]. There are two main challenges to evaluating τάμνον's effectiveness: first, because τάμνον only reports dialect differences based on the provided rules, τάμνον cannot identify tokens whose dialect features are not part of τάμνον's rule set (e.g., long alphas in word stems); second, part of the purpose of τάμνον is to fix problems in Morpheus's dialect analyzer, so a match in 100% of the tokens would actually mean we have failed. We address the first problem by not testing on dialect forms which are not in τάμνον's rules. We address the second problem by not simply comparing the results of τάμνον and Morpheus but by doing a more complicated analysis.

We examine all possible parses (each token may have multiple parses) in *Iliad* book 1. For each of the three dialects, we compare whether τάμνον and Morpheus marked each parse as that dialect. See Table 3 for the evaluation results.

| | **Ionic** | **Aeolic** | **Homeric** |
|---|---|---|---|
| **Both**: | 499 parses | 431 parses | 416 parses |
| τάμνον: | 548 parses | 241 parses | 159 parses |
| **Morpheus**: | 1,978 parses | 1,634 parses | 3,001 parses |
| **Neither**: | 15,651 parses | 16,370 parses | 15,100 parses |
| **Match on (Total: 18,676)**: | 16,150 parses (86.47%) | 16,801 parses (89.96%) | 15,516 parses (83.08%) |

**Table 3: Number of parses in each category for each dialect.**

### 7.2.1 Issues

A manual analysis of all 7,561 instances of disagreement shows that τάμνον is always correct when it assigns a dialect and Morpheus does not. For example, κεῖνοι is an Ionic form, not "poetic" as per Morpheus.

For tokens where Morpheus indicates the dialect and τάμνον does not, it is for one of three reasons:

1. **Unknown Rule**: The dialect assignment is based on a rule not included in τάμνον. Often the reason for this dialect assignment is unclear, one of the weaknesses of Morpheus that τάμνον attempted to address. Since none of these rules appear to be included in Buck and are poorly defined, we do not find it useful to include them in the list of rules for the current project.
2. **Parse Failure**: The morphological parsing of the token was incorrect or the attempts by the preprocessor to determine type information about the lemma failed. For example, ἀλή is not properly recognized as an alpha-stem. This happens on very few tokens (in the tens for each dialect), and is thus not considered a major issue for the goals of this project.
3. **Labelling Difference**: Morpheus and τάμνον have slightly different labelling criteria. For example, Morpheus identifies some parses as Ionic, Aeolic, and Homeric at the same time. Since τάμνον seeks to differentiate between only these dialects, it does not mark these types of tokens as all three dialects, but as not showing a distinct feature of any of them. For example, the second parse of ἄλτο is identified by Morpheus as "homeric doric ionic and aeolic." The other instance of this is where Morpheus identifies a feature as Homeric-Ionic (because it appears in

both Homer and Ionic), whereas τάμνον identifies it as only Ionic because it shows a feature present in both older Ionic (that of Homer) and later Ionic, and is therefore best termed simply "Ionic." For example, the feminine first declension endings in -η are a feature of Ionic in both time periods that is identified as both "Ionic" and "Epic" by Morpheus. Since these identifications are functionally equivalent from a feature standpoint, this is not an issue for τάμνον's accuracy.

These cases represent a potential area for future work with the addition of more rules (for issue 1), additional work to supplement Morpheus' parsing (for issue 2), or additional dialects (for issue 3). However, these issues are all currently acceptable for the reasons mentioned above.

# 8 Results and Discussion

We begin with an overview of the feature results before examining individual results in depth.

## 8.1 An Overview of the Data

Our first step is to perform a Principal Component Analysis on the results produced by ᾠδικόν and τάμνον for the various texts. A Principal Component Analysis (PCA) takes the entire list of 2,631 features and uses a linear combination of these to create a new *component*, which is a single feature that describes as much of the difference between the texts as possible. It then repeats the process to create a second component designed to differentiate the texts as much as possible, given the differentiation already present in the first component. It continues creating more components that maximize the difference given the previous components until the components can describe all of the difference between the texts. By taking the first two, three, or four components of a Principal Component Analysis, we can gain insight into the sources of greatest differentiation between texts and provide a visual representation of which texts are naturally more similar or different within this feature space, as a two-dimensional space is easier to visualize than a 2,631-dimensional space. Another important part of the PCA is that it is done in an *unsupervised* way, so it has no idea which texts are supposed to be Homeric or not; it just tries to make sure texts that were different across the

2,631 features are now similarly different across the new features.

For our first data set, we run the PCA on only the texts as a whole, then apply the transformations to individual books of the larger texts as well. In this way, we create a space of maximum difference between the many overall texts and then observe which of the overall texts the individual books are most similar to. The results can be viewed in Figure 6. We note that in the first two dimensions, the texts generally fall into two clusters. One contains the *Dionysiaca*, the *Taking of Ilios*, and the *Abduction of Helen*. The other contains most of the remaining texts, with a few key outliers to be discussed later (*Fall of Troy* and the works of Bion fall somewhere in between). One can then see that the third and fourth components separate the *Iliad* and the *Odyssey* from the *Homeric Hymns* and other texts that were clustered with it in the first two dimensions. Even in this simple case, we can see that there is reason to believe the *Iliad* and *Odyssey* differ from the other texts.

Next, we run an analysis with all of the books considered as individual texts, visible in Figure 7, to get a view of the clusters when all of the books are considered in the fitting. We note that this graph shows the first and *third* components, as the second component groups a few of the Hymns and the *Shield of Heracles* together with the *Iliad*/*Odyssey* while the third component shows clear differentiation. The *Iliad* and the *Odyssey* are in a nice cluster in the bottom left corner, three of the longer Hymns are nearby, and the short Hymns and Hymn to Hermes are clearly separate.

Finally, we run an analysis considering only the *Iliad*/*Odyssey*, the *Homeric Hymns*, and the works of Hesiod, visible in Figure 8, to get a closer look at the Homeric and Pseudo-Homeric texts. This analysis shows clear differentiation between the *Iliad*/*Odyssey* and the *Homeric Hymns* plus the work of Hesiod, as well as providing some intuition that the Hymn to Hermes and Short Hymns are again outliers. Callimachus' *Hymns* are plotted to give a context in terms of where an Alexandrian text falls within this space.

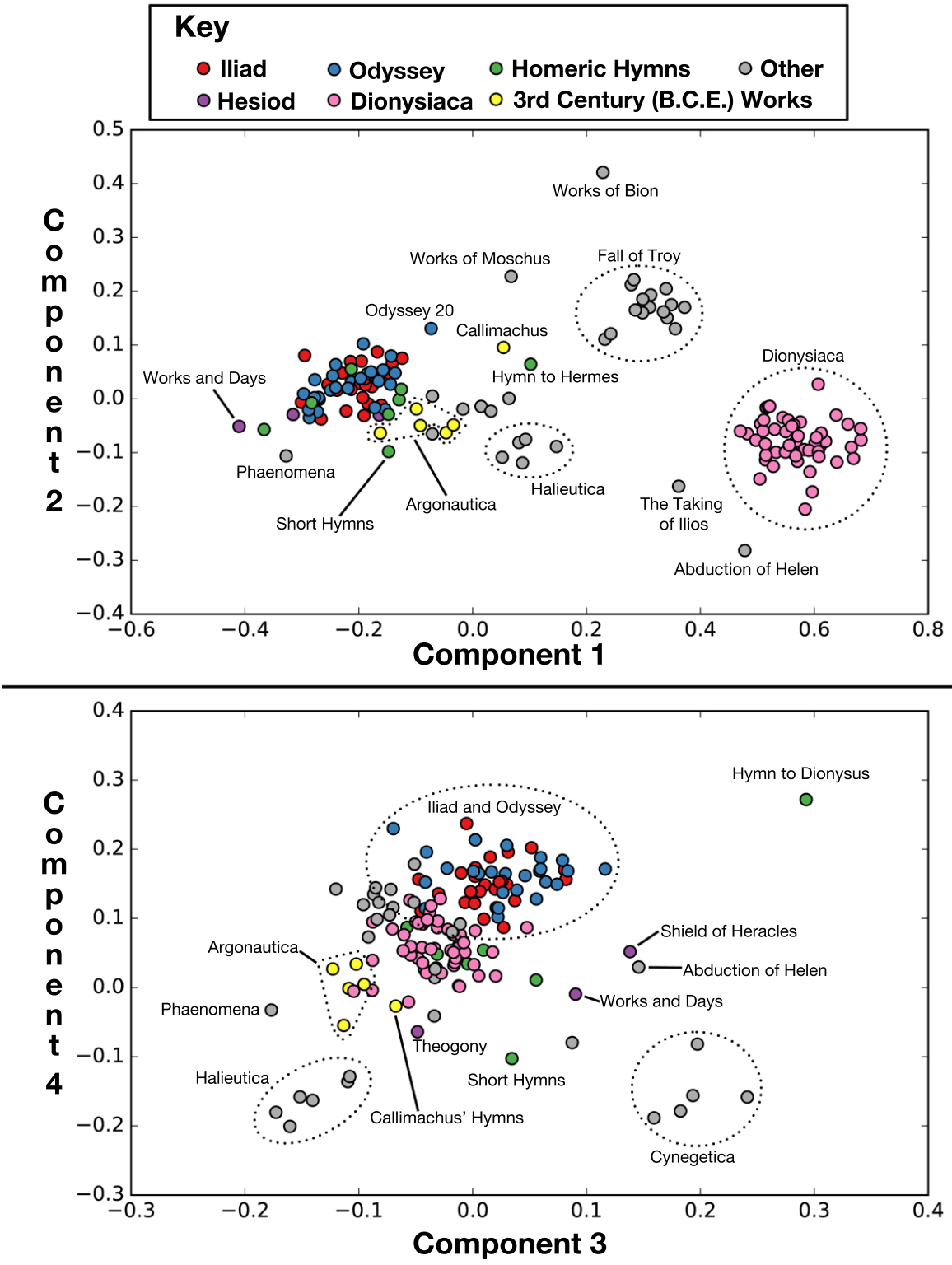


**Figure 6: The first four components of a PCA trained only on entire texts. Note that the third and fourth components separate out the *Iliad* and the *Odyssey*, minus a few interlopers from the *Dionysiaca* and *Fall of Troy* clearly differentiated by component 1.**

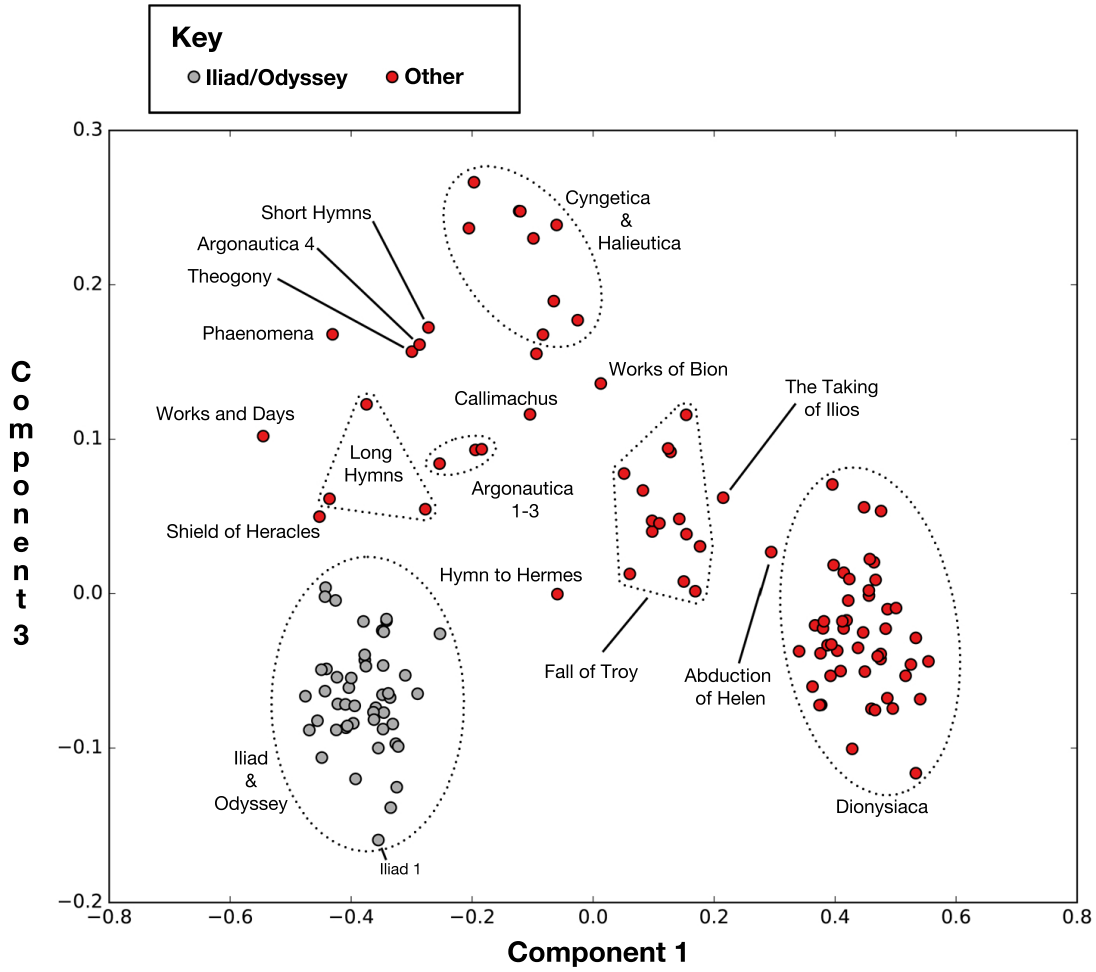


**Figure 7: The first and third components of a PCA performed on all of the individual books and texts. Books of the *Iliad* and *Odyssey* in gray, other texts in red.**

To assist our further discussions below, we also perform the following analyses:

1. Before Principal Component Analysis, we use a cross-validated lasso to select a smaller subset of features that preserve differentiation between the Homeric and Non-Homeric texts. This allows us to analyze which features are more predictive and have better interpretability.
2. We cross-validate a variety of classifiers on a randomized data set designed to include some Homeric, Pseudo-Homeric, and Non-Homeric works in every test set.
3. We test these same classifiers on a hold-one-out basis, where the classifiers are trained on all of the texts/books except for one and then used to predict whether that single book is Homeric or Non-Homeric.

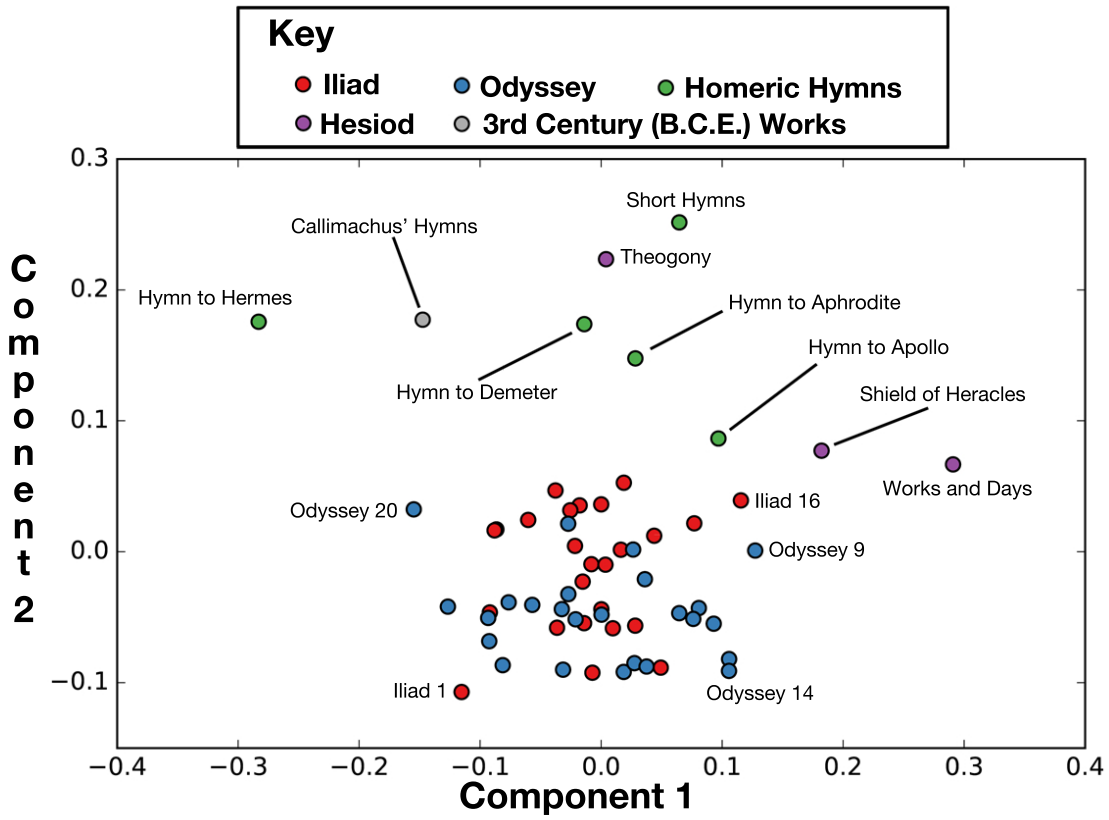


**Figure 8: The first two components of a PCA performed only on the *Iliad*, the *Odyssey*, the *Homeric Hymns*, and the works of Hesiod, with Callimachus' *Hymns* plotted for context.**

The "variety of classifiers" mentioned above involve running every valid permutation of feature selection, preprocessing, and classifier described below. The feature selection algorithms are either no feature selection or feature selection using a cross-validated lasso with one of three thresholds yielding between 27 and 31 resulting features. The preprocessing algorithms are a Standard Scaler, which just normalizes the features, and Principal Component Analysis, keeping the best 2, 3, and 4 components. The classifiers are:

- Logistic Regression.
- Support Vector Machine (Linear Kernel).
- Support Vector Machine (RBF Kernel).
- Random Forest.
- K Nearest Neighbors classifiers using 2, 3, 4, and 5 neighbors both weighted and non-weighted by distance.

We include Logistic Regression and the SVM with a Linear Kernel because they are both simple linear classifiers that provide a clear boundary between the two types of texts. We include the SVM with RBF Kernel and the K Nearest Neighbors classifiers because the Homeric texts seem to cluster into ellipsoid shapes, and these classifiers work well on that type of data. We include the Random Forest for a more complex ensemble classifier's perspective. For a reader unfamiliar with machine learning, knowing the specifics of these classifiers is not essential to understanding the following results; it suffices to say that we have chosen a few standard machine learning algorithms to help us understand how clear (or unclear) the differences between these texts are.

## 8.2 Question 1: Can We Define "Homeric"?

From a qualitative standpoint, Figures 7 and 8 show that in these various spaces, the books of the *Iliad* and *Odyssey* can be contained in a reasonably simple shell without including any other texts. While the longer *Homeric Hymns* typically fall close to the edge of this shell, a clear dividing line can be drawn between them and the books of the *Iliad* and *Odyssey*. They also quite clearly show that the shorter *Homeric Hymns* are rather different from the longer ones, and that the Fourth *Homeric Hymn*, to Hermes, is as similar to the *Fall of Troy* or works of Callimachus as to the other Hymns and the *Iliad* and *Odyssey*. These texts, the Hymn to Hermes and shorter Hymns, are believed to be later compositions, so our method provides more evidence for previously identified differences within *The Homeric Hymns* [12]. We also note that the first *Homeric Hymn*, to Dionysus, is quite different from the rest of the texts (as seen in Figure 6); unfortunately we only have a small fragment of the original text and therefore the sample is too small to conclusively say whether it is actually different or just an artifact of the limited data. For this reason we have excluded it from the later analyses.

We would also like a more quantitative confirmation of this qualitative eye test. Can this data be used to conclusively differentiate between Homeric and Non-Homeric texts? For the current data, the answer is unfortunately no. We present the results of running the classifiers mentioned above on all of the texts, either using 5-fold cross-validation or hold-one-out, in Tables 4 and 5. Even when

we perform the hold-one-out analysis, only three analyzers out of 189 correctly identify every text (the Lasso with a threshold of .15, .25, and .40, Standard Scaling preprocessing, and Linear Kernel SVM), and none of the classifiers are 100% accurate in the 5-fold situation.

The classifiers, whether in the 5-fold cross-validation case or hold one out case, never misidentify books of the *Dionysiaca*, the *Halieutica*, or the fall *Fall of Troy*, nor do they misidentify the texts *Abduction of Helen*, Callimachus' *Hymns*, or the works of Bion. In general, they also correctly identify books of the *Iliad* and *Odyssey*, though Random Forests of various sorts misidentify *Iliad* 6, 8, 16, 16, 21 and *Odyssey* 3, 7, 9, 10, 20. With a small data set like this one it is reasonable to believe that the Random Forest simply overfits the data. *Iliad* 21 is also misidentified twice by Logistic Regression. For a fuller breakdown of the classifier results, see Tables 4 and 5.

However, things are slightly less clear when we analyze the Pseudo-Homeric texts. Less than 20% of the classifiers misidentify the Hymn to Hermes and the shorter *Homeric Hymns*, a number roughly comparable to the *Argonautica* book 3, which is indisputably Non-Homeric (though of course Apollonius would have drawn inspiration from Homer). In this case, it is safe to say that these texts are Non-Homeric. For the other Hymns, to Demeter, Apollo, and Aphrodite, generally more than half of the classifiers identify them as "Homeric." We could set the threshold to say a text is Homeric if greater than 90% of the classifiers identify it as Homeric, and we would have a nice barrier between the Hymn to Apollo, identified as Homeric by 86% of classifiers, and book 21 of the *Iliad* identified as Homeric by 94%, but this would be a little bit arbitrary and is not the conclusive evidence we desire. We also note that the *Shield of Heracles* would be identified as "Homeric" by this criteria; in the 5-fold cross-validation case, it is considered more Homeric than *Iliad* 21! However, since this text is an homage to the description of Achilles' shield in the *Iliad*, this can be explained as Hesiod doing a very good job of imitating the source material.

We are left with the following situation: from a qualitative standpoint, it can be seen that the *Iliad* and the *Odyssey* differ from the *Homeric Hymns*. From a quantitative standpoint, we can also say with confidence that the shorter Hymns and the Hymn to Hermes are very different from the Homeric and Pseudo-Homeric texts, as we would expect. In all of the tests, using most of the better

classifiers or running a properly calibrated voting algorithm would show that they are non-Homeric. However, with the current features and limited data, our set of classifiers cannot reliably say, from a quantitative standpoint, that the large Hymns (To Demeter, Apollo, and Aphrodite) are *definitely not* Homeric. On the other hand, the data does not show that they clearly *are* Homeric, so there is hope that further analysis can help crystallize the differences hinted at in these analyses.

| **Text** | **5-Fold** | **Hold-1-Out** | **Text** | **5-Fold** | **Hold-1-Out** |
|---|---|---|---|---|---|
| *Iliad*: Book 1 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 1 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 2 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 2 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 3 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 3 | 0% (0/189) | 1% (1/189) |
| *Iliad*: Book 4 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 4 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 5 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 5 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 6 | 0% (0/189) | 1% (1/189) | *Odyssey*: Book 6 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 7 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 7 | 0% (0/189) | 1% (1/189) |
| *Iliad*: Book 8 | 2% (4/189) | 1% (1/189) | *Odyssey*: Book 8 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 9 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 9 | 0% (0/189) | 2% (4/189) |
| *Iliad*: Book 10 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 10 | 1% (1/189) | 0% (0/189) |
| *Iliad*: Book 11 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 11 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 12 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 12 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 13 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 13 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 14 | 0% (0/189) | 1% (1/189) | *Odyssey*: Book 14 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 15 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 15 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 16 | 1% (1/189) | 1% (2/189) | *Odyssey*: Book 16 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 17 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 17 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 18 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 18 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 19 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 19 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 20 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 20 | 3% (5/189) | 6% (11/189) |
| *Iliad*: Book 21 | 6% (11/189) | 2% (4/189) | *Odyssey*: Book 21 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 22 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 22 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 23 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 23 | 0% (0/189) | 0% (0/189) |
| *Iliad*: Book 24 | 0% (0/189) | 0% (0/189) | *Odyssey*: Book 24 | 0% (0/189) | 0% (0/189) |

**Table 4: This table shows the number of classifiers that incorrectly identified each book of the *Iliad* and the *Odyssey* when they were 5-fold cross-validated and trained on every other text and then used to evaluate the given book.**

| Text | 5-Fold | Hold-1-Out |
|---|---|---|
| Hymn to Demeter | 67% (126/189) | 44% (83/189) |
| Hymn to Apollo | 86% (162/189) | 81% (153/189) |
| Hymn to Hermes | 19% (36/189) | 19% (36/189) |
| Hymn to Aphrodite | 80% (152/189) | 51% (96/189) |
| Short Hymns | 14% (26/189) | 14% (26/189) |
| Callimachus' *Hymns* | 0% (0/189) | 0% (0/189) |
| *Abduction of Helen* | 0% (0/189) | 0% (0/189) |
| *Shield of Heracles* | 99% (188/189) | 92% (173/189) |
| *Theogony* | 19% (36/189) | 21% (39/189) |
| *Works and Days* | 90% (171/189) | 77% (145/189) |
| *Cynegetica*: Book 1 | 7% (13/189) | 0% (0/189) |
| *Cynegetica*: Book 2 | 0% (0/189) | 0% (0/189) |
| *Cynegetica*: Book 3 | 0% (0/189) | 0% (0/189) |
| *Cynegetica*: Book 4 | 0% (0/189) | 0% (0/189) |
| *The Taking of Ilios* | 0% (0/189) | 0% (0/189) |
| *Argonautica*: Book 1 | 6% (12/189) | 3% (6/189) |
| *Argonautica*: Book 2 | 5% (9/189) | 3% (6/189) |
| *Argonautica*: Book 3 | 34% (64/189) | 13% (25/189) |
| *Argonautica*: Book 4 | 14% (27/189) | 10% (18/189) |
| *Phaenomena* | 61% (116/189) | 19% (36/189) |
| Works of Moschus | 4% (8/189) | 4% (8/189) |
| Works of Bion | 0% (0/189) | 0% (0/189) |

**Table 5: This table shows the number of classifiers that incorrectly identified our Pseudo-Homeric and Non-Homeric books and texts, when the classifiers were 5-fold cross-validated and trained on every other text and then used to evaluate the given book. The books of the *Dionysiaca*, *Fall of Troy*, and the *Halieutica* are left out because there were 0 failures on all books in all cases.**

## 8.3 Question 2: What Features are Homeric?

Given that there are these clusters of Homeric and Non-Homeric texts when we perform a Principal Component Analysis, what features are contributing to these clusters? Since the Principal Component Analysis has all 2,631 features to choose from, its components are quite complex linear combinations. For example, the second component in Figure 8, differentiating the Homeric

Texts from the Pseudo- and Non-Homeric texts, has 422 features, and many of them contribute significantly to the differentiation. This makes interpretability rather difficult. The five features given the most weight are the following, with a "yes" answer meaning the text is more Homeric (see Appendix A.3 items 3 and 4 for explanations of what these features are):

- Is there any caesura in the first foot?
- Is there a feminine caesura in the fifth foot?
- Did correption fail to happen when it could have?
- Is there a masculine caesura in the first foot?
- Is there any caesura in the fifth foot?

The reader will notice that this really boils down to three features, a caesura in the first foot, a caesura in the fifth foot, and correption frequency. The differences are also reinforced by other less heavily weighted features including Ionic frequency (less Homeric) and Mute + Liquids being pronounced together (less Homeric). The frequency of caesura in the first and fifth feet may support the hypothesis that the poems of Homer were oral compositions using repeated formulae: for example, in many formulae the name Ἀχιλλεύς (Achilles) is placed at the end of the line, creating a feminine caesura in the fifth foot, and in general this perhaps reflects a comfort with placing shorter words at the start and end of lines as part of these formulae. The lack of correption (hiatus) present in the Homeric texts may also reflect a higher frequency of digammas, which cause hiatus, but this hypothesis would be best confirmed by using sense disambiguation to actually count the number of observed digammas explicitly. The lower frequency of Ionic forms fits well with the fact that the works of Homer are composed in a mixed Ionic/Aeolic style compared to the pure Ionic of many contemporary works like those of Hesiod. The Mute + Liquid rule being utilized less frequently in Homer may also reflect oral versus written composition, as this change puts stress on the spoken meter[7] which would have been more noticeable to a poet composing the works orally.

In order to reduce the number of features and duplication, as well as increasing interpretability, we can also perform feature reduction to get a more manageable feature set. We run such a feature

[7] Try saying "ap-la" and compare it to "a-pla".

reduction to produce a set of 24 features, then run a Principal Component Analysis on these 24 features. Figure 9 shows the first and third components of this result. Because we have reduced the number of features available, it no longer provides a perfect boundary between the Homeric and Non-Homeric texts, as the Shield of Heracles (which the reader will recall is an imitation of Homer) is more "Homeric" than *Iliad* 8, and one of the longer hymns is quite similar as well. However, it still generally provides us with a "Homeric" cluster, and since our goal in this section is to determine what features are Homeric rather than to perfectly differentiate texts, it is not a major concern. The 24 features chosen are:

- How frequently are each of the 5 feet spondaic?
- How many lines have 3 spondees in a row?
- How frequently is there word break in foot 1?
- How frequently is there word break in foot 5?
- How frequently is there word break after the arsis (masculine caesura) of foot 2 and/or 4?
- How frequently is there word break in the thesis (feminine caesura) of foot 1, 2, and/or 3?
- How frequently is there a word break after foot 1 and/or 2?
- How frequently is there ictus lengthening?
- How frequently are mutes and liquids pronounced together, both overall and relative to the number of instances it could occur?
- How frequently do we see tokens showing an Ionic but not an Aeolic feature?
- How frequently do we see tokens showing an Aeolic but not an Ionic feature?
- How frequently do we see tokens possibly showing an Ionic feature?
- How frequently do we see tokens possibly showing a Homeric feature?
- How frequently do we see tokens definitely showing a Homeric-only feature?
- How frequently do we see Homeric -οιο endings?

Without even looking at which of these features are more or less Homeric, we note that the set includes both metrical features and whether a text shows Ionic/Aeolic/Homeric features, which means that both of our tools have relevant information for differentiating these texts.

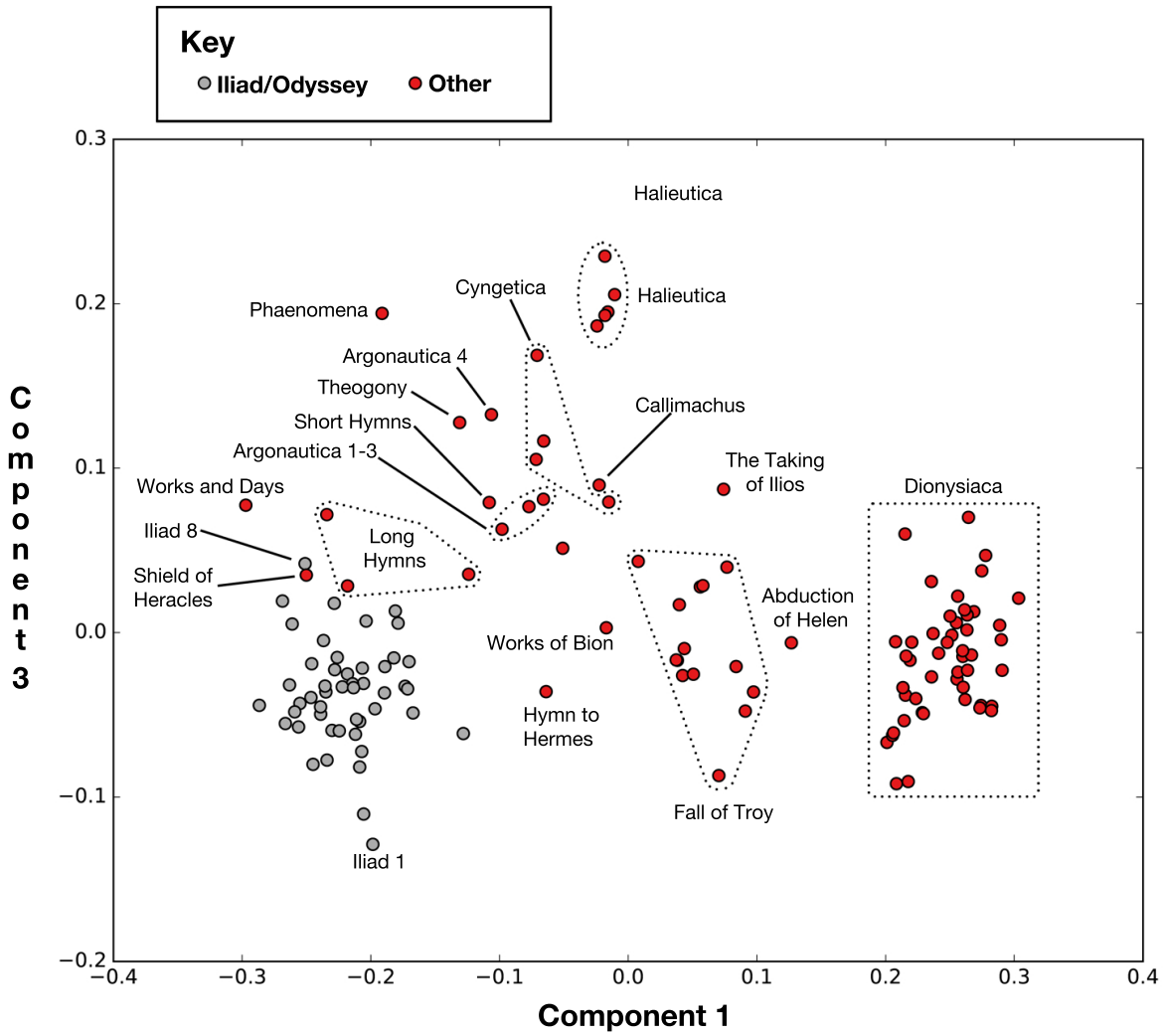


**Figure 9: The first two components of a PCA performed on a reduced set of 24 features for all of the books and texts. Books of the *Iliad* and *Odyssey* in gray, other texts in red.**

We now consider the first and third components of a PCA run on every individual book using this reduced feature set, as seen in Figure 9. The first component, on the x-axis, ranges from the *Iliad*/*Odyssey* on the one hand to the *Dionysiaca* on the other, and appears to very roughly approximate time of composition, with later works to the right. The ten most significant features, ordered from most to least significance, with greater frequency meaning more Homeric (that is, less Dionysiacan), are:

1. How frequently is there no word break in the thesis of the third foot (no feminine caesura)?
2. How frequently is foot 1 spondaic?

3. How frequently is there no word break after the thesis of the fourth foot (no masculine caesura)?
4. How frequently is foot 3 spondaic?
5. How frequently is foot 2 spondaic?
6. How frequently is there a word break in the first foot?
7. How frequently is there a word break after the second foot?
8. How frequently is there a word break after the first foot?
9. How infrequent are Ionic forms?
10. How frequently is there a word break in the fifth foot?

We note that the *Dionysiaca*, which dates from the 4th or 5th century C.E. is generally more dactylic than the other texts, which perhaps represents a standardization of preference for dactyls. We also note the importance of "Ionic forms," which may seem counter-intuitive but actually has a logical explanation: the Koine Greek which survived to the time period of the *Dionysiaca* was based on Attic Greek, which is quite similar to Ionic. If τάμνον separately identified Attic and Ionic features, we would expect it to be the Attic features that were chosen for this differentiation.

The third component, on the y-axis, helps differentiate the Homeric texts from the other non-*Dionysiaca* texts (for the most part). Its 10 most significant features, in order of most to least significant, with greater frequency being more Homeric, are:

1. How frequently is there a word break in the first foot?
2. How frequently is foot 2 dactylic?
3. How frequently is there a word break after the second foot?
4. How frequently is foot 1 dactylic?
5. How frequently is foot 4 spondaic?
6. How frequently is foot 5 dactylic?
7. How frequently is there a word break in the thesis of the second foot (feminine caesura)?
8. How frequently is there a word break in the fifth foot?
9. How frequently is there a word break in the thesis of the first foot (feminine caesura)?
10. How frequently is foot 3 dactylic?

There is a significant drop-off in importance before the 10th and final feature in the list above, but we also note with amusement that the 12th most important feature is that having more "Homeric" dialect features actually correlates with the *Non*-Homeric texts. Returning to the list at hand, Homer is generally more dactylic than the other texts in feet 1, 2, and 5. Since spondees placed stress on the meter when spoken aloud, this could perhaps be seen as evidence for the works of Homer to be oral compositions compared to the other texts as textual compositions. We also see again the tendency for word breaks in the first and fifth foot, as mentioned earlier. Lastly, the fact that these texts show *more* Homeric forms can be explained by the idea of "false archaism," where later poets, attempting to sound more like Homer, imitated some obvious features of his work (like the masculine genitive ending in -οιο). This attempt at imitation may in fact lead to greater usage of these types of features than the original in an attempt to sound like an epic is "supposed to."

## 8.4 Question 3: How Different Are the *Iliad* and *Odyssey*?

Can we reasonably conclude that the *Iliad* and *Odyssey*, taken as whole texts, had different authors? From a qualitative standpoint, after performing a PCA on the books of these two texts as seen in Figure 10, the books of the two texts look quite well mixed. Perhaps along the second component, one sees the *Iliad* generally trending positive and the *Odyssey* trending negative, but it is not particularly clear or well-defined.

Moving to a more quantitative view, in order to analyze the texts we perform a 5-fold cross validation of every valid permutation of the following feature selection, preprocessors, and classifiers: The feature selection algorithms are:

- No feature selection.
- Lasso feature selection yielding 30 features.

The preprocessing algorithms are:

- A Standard Scaler, which just normalizes the features.
- PCA with 2, 3, and 4 components.

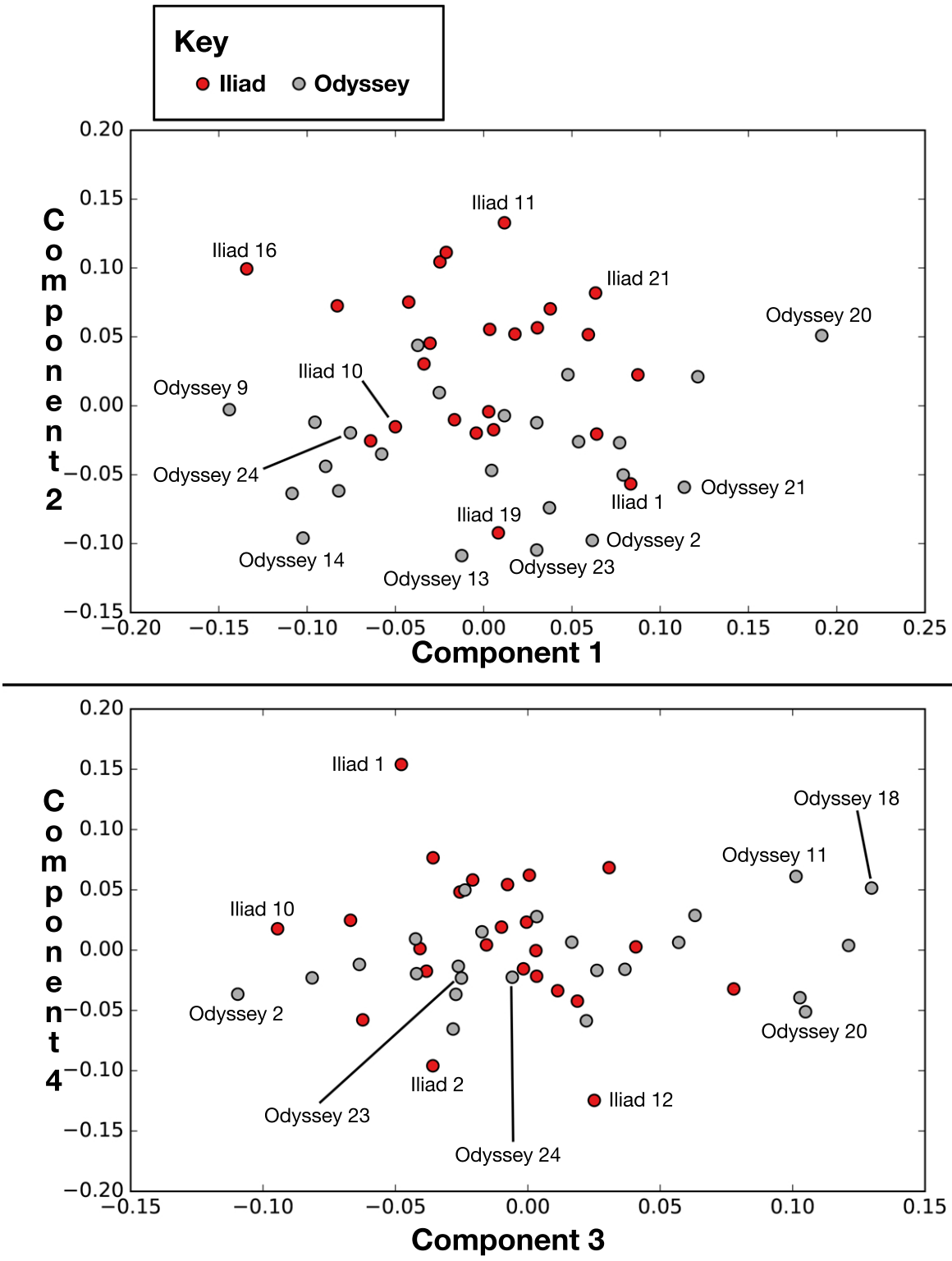


**Figure 10: The first four principal components when analyzing only the *Iliad* (with books in red) and the *Odyssey* (with books in gray).**

The classifiers are:

- Logistic Regression.
- Support Vector Machine (Linear Kernel).
- Support Vector Machine (RBF Kernel).
- Random Forest.
- K Nearest Neighbors classifiers using 2, 3, 4, and 5 neighbors, both weighted and non-weighted by distance.

The average accuracy of these 95 classifiers is 78%, with the best classifier, Lasso feature selection to Standard Scaler to 2 Nearest Neighbors weighted by distance, having 100% accuracy, and the two worst classifiers, lasso feature selection to PCA (with 2 or 3 features) to SVM with RBF kernel, having 24% accuracy. 78% accuracy is well above the 50% we would expect if the two were indistinguishable; as a comparison, when splitting the books of the *Dionysiaca* into the first 24 and second 24, the average accuracy of this same technique is 44%. On the other hand, comparing the books of *Fall of Troy* to the *Dionysiaca*, two texts by clearly different authors, yields an average accuracy of 98%. So the *Iliad* and *Odyssey* fall somewhere in the middle.

We also use all 189 feature selection/preprocessing/classifier combinations mentioned in earlier sections, train them on either the *Iliad* or *Odyssey* and all the Non-Homeric texts, then test them on the other Homeric text (the *Odyssey* and the *Iliad*, respectively), and determine how many fail. The results can be viewed in Table 6. We note that even the book with the least accuracy (*Iliad* 8) is incorrectly identified by only 28% of classifiers, and only seven books are incorrectly identified by more than 15%. This makes it clear that the *Iliad* and the *Odyssey* are very similar texts within the space of texts we are examining, as "Iliad" classifiers generally correctly identify books of the *Odyssey* and the same is true of "Odyssey" classifiers identifying the *Iliad*.

Overall, the two texts seem similar but not identical. They certainly do not display the unity of *Dionysiaca*, but they are not starkly different enough to clearly be of different authorship, and the differences within the two texts are far more significant than the differences between them. This middle ground could be explained by a composition process involving long lines of oral poets

passing the works on and modifying them in similar ways, two texts by the same author in slightly different genres, or a single author writing one text earlier in life and the other later, as Janko argues [31]. A closer differentiation between these (and other) possibilities will have to be the focus of a later work.

| **Text** | **Classifier Failures** | **Text** | **Classifier Failures** |
|---|---|---|---|
| *Iliad*: Book 1 | 10% (19/189) | *Odyssey*: Book 1 | 10% (19/189) |
| *Iliad*: Book 2 | 11% (21/189) | *Odyssey*: Book 2 | 10% (18/189) |
| *Iliad*: Book 3 | 7% (13/189) | *Odyssey*: Book 3 | 11% (20/189) |
| *Iliad*: Book 4 | 12% (23/189) | *Odyssey*: Book 4 | 11% (21/189) |
| *Iliad*: Book 5 | 15% (29/189) | *Odyssey*: Book 5 | 6% (12/189) |
| *Iliad*: Book 6 | 12% (23/189) | *Odyssey*: Book 6 | 7% (14/189) |
| *Iliad*: Book 7 | 10% (19/189) | *Odyssey*: Book 7 | 17% (32/189) |
| *Iliad*: Book 8 | 28% (52/189) | *Odyssey*: Book 8 | 10% (18/189) |
| *Iliad*: Book 9 | 10% (18/189) | *Odyssey*: Book 9 | 12% (23/189) |
| *Iliad*: Book 10 | 10% (19/189) | *Odyssey*: Book 10 | 9% (17/189) |
| *Iliad*: Book 11 | 15% (29/189) | *Odyssey*: Book 11 | 7% (13/189) |
| *Iliad*: Book 12 | 10% (19/189) | *Odyssey*: Book 12 | 4% (7/189) |
| *Iliad*: Book 13 | 13% (24/189) | *Odyssey*: Book 13 | 7% (13/189) |
| *Iliad*: Book 14 | 15% (28/189) | *Odyssey*: Book 14 | 7% (13/189) |
| *Iliad*: Book 15 | 11% (21/189) | *Odyssey*: Book 15 | 10% (18/189) |
| *Iliad*: Book 16 | 15% (28/189) | *Odyssey*: Book 16 | 10% (18/189) |
| *Iliad*: Book 17 | 11% (20/189) | *Odyssey*: Book 17 | 2% (3/189) |
| *Iliad*: Book 18 | 11% (21/189) | *Odyssey*: Book 18 | 10% (19/189) |
| *Iliad*: Book 19 | 10% (18/189) | *Odyssey*: Book 19 | 9% (17/189) |
| *Iliad*: Book 20 | 13% (24/189) | *Odyssey*: Book 20 | 19% (35/189) |
| *Iliad*: Book 21 | 13% (25/189) | *Odyssey*: Book 21 | 13% (24/189) |
| *Iliad*: Book 22 | 10% (18/189) | *Odyssey*: Book 22 | 12% (22/189) |
| *Iliad*: Book 23 | 10% (19/189) | *Odyssey*: Book 23 | 7% (14/189) |
| *Iliad*: Book 24 | 6% (12/189) | *Odyssey*: Book 24 | 7% (13/189) |

**Table 6: We trained 189 different classifier pipelines on the *Iliad* or the *Odyssey* and all the Non-Homeric texts, then analyzed their accuracy on the books of the other text (e.g., if we trained on the *Iliad* we test on the *Odyssey*). The failure percentages and numbers are shown in this table.**

## 8.5 Question 4: Are Certain Books of the *Iliad* and the *Odyssey* Outliers?

This book-by-book analysis performed above and visible in Table 6 brings us to the question of unusual books in the *Iliad* and *Odyssey*. The seven that are incorrectly identified by more than 15% of the classifiers as mentioned above are *Iliad* books 5, 8, 11, 14, and 17, and *Odyssey* books 7 and 20. Looking at the first four components of a Principal Component Analysis of the two texts in Figure 10, we can see that *Odyssey* 20 and *Iliad* 1 both end up heavily differentiated from the others (along components 1 and 4, respectively). We also see *Iliad* 1 at the bottom of the Homeric works in Figures 9 and 7. In Table 4, *Iliad* 21 and *Odyssey* 20 are the only books misidentified by more than 2% of classifiers. The only texts that seems to be marked as odd by multiple analyses are *Iliad* 1 and *Odyssey* 20. The main differentiating factors for *Odyssey* 20 appear to be that, when word breaks appear in the third foot, they occur in the middle of the thesis (feminine caesura) rather than between the arsis and the thesis (masculine caesura). However, the literature does not generally flag book 20 as strange so this does not seem to reflect a massive difference. *Iliad* book 1, on the other hand, is the first book of the first text, and therefore may have been maintained more carefully through ages of oral transmission than other texts, preserving archaic forms and metrical patterns.

Notably absent from any of the unusual books identified by this analysis are the books that are actually identified by scholars as potentially late additions to the texts. Book 10 of the *Iliad* is frequently considered late, and parts of book 11, 23, and 24 of the *Odyssey* are also sometimes argued as late [19]. However, these texts do not appear particularly unusual by most of the metrics we examine. In Figure 10, for example, they are generally near the center of the cluster, and on the occasional component where they are near an edge they are not the biggest outlier. For example, *Iliad* 10 is the second furthest to the left by component 3, but not as far as *Odyssey* 2. These four books are not any of the few texts misidentified by classifiers in the 5-fold or hold-one-out analyses in Table 4 and in fact they are some of the books that are *most* accurately identified by the classifiers trained only on other texts, as seen in Table 6. In general, it is remarkable how little these features support viewing these books as late. Thus, based on our current feature results, we are not able to support any existing hypotheses about certain books being later interpolations.

# 9 Future Work

There are a variety of improvements and future directions that could build on this project.

## 9.1 Improvements to the tools

**Improvements to** ᾠδικόν:

There are many available improvements for ᾠδικόν. First, a few ways to improve the accuracy of the Native Speaker Approach:

- Expand the dictionary behind the Native Speaker Approach by hand-checking every dictionary item to ensure it has the proper lengths.
- Add a list of exceptions to the morphological parser to catch correct parses that it misses.
- Add the ability to use treebank data to select a most likely parse and use only that parse's lengths (rather than taking the union of all possible parses). This would also allow us to collect more certain data on digammas directly, since at the moment parse ambiguity makes the statistics rather unreliable.

We could also add some small tweaks and improvements to make it run more quickly and more accurately in very rare corner cases. There are more features it could extract, like explicitly detecting hiatus, the use of repeated formulae, and similarities between the scansion of various lines (during the manual scansion of the *Iliad* and *Odyssey*, we noticed that there were frequently "runs" of similar scansion patterns). Another major improvement is a third approach we dub the *Scholar Approach*. This replicates the method used by scholars to determine the natural lengths vowels in these archaic words (and the presence of digammas) by using the Student Approach to scan the text and using this scansion to discover the natural lengths of the words in the line. Using the lengths from lines that the Student Approach can scan, we would then create a dictionary for the Native Speaker approach, thus re-deriving this knowledge from the texts themselves, and perhaps adding new lengths not included in the dictionary. This dictionary could then be used to scan the lines that could not be scanned in the initial run through of the Student Approach.

**Improvements to** τάμνον:

τάμνον also has many avenues for future improvement. The current list of rules is limited to what was accessible using the combination of Buck, Benner, and Morpheus' morphological analysis, but could be improved by researching rules not included in Buck or Benner and by taking advantage of a dictionary of words to determine Ionic/Aeolic/Attic/Doric variants of the stems of many different words, not just the few dozen we took from Buck. We would also like to extend the number of dialects to include not just Ionic, Aeolic, and archaic features, but Attic, Doric, Arcado-Cypriot, and even more specific dialects within these groups. Also, instead of just keeping track of the maximum and minimum counts for each dialect (that is, how many tokens *could* be Ionic and how many tokens are *definitely* Ionic), we could use treebank datasets from Perseus to determine the most likely parse for a given token and therefore assign a probability of the token showing features of each dialect rather than the binary maximum/minimum setup.

We would also like to publish both of these tools, perhaps as part of the Classical Language Toolkit, for others to use in their own analyses.

## 9.2 Morphological Parsing

We use Morpheus as a basis for many of our analyses, but this has a limited dictionary and fails to parse a variety of forms. Though we mentioned earlier that building a morphological parser is beyond the scope of this project, between the dictionary and stemmer built for the Native Speaker Approach, it would not be a huge step to build an actual morphological parser, with a larger dictionary to cover some of the tokens that Morpheus fails on. We could also go through our created dictionary with a fine-tooth comb and publish a "Greek dictionary for computational analysis" including the lengths of the vowels, digammas, various dialect forms, and other features so that future scholars working in this area can benefit from our work and have an easier path forward.

## 9.3 Further analyses of Homeric texts

Beyond analyzing the texts with improved versions of ὠδικόν and τάμνον, one could also look at a further set of preprocessing techniques and classifiers, as well as doing a more exhaustive search of the hyperparameter space, to find classifiers that can reliably differentiate between the *Iliad*/*Odyssey* and the large *Homeric Hymns* besides the Hymn to Hermes. We could also look more closely at every feature to see which ones are more Homeric and draw conclusions about the texts in that way. It would also be interesting to compare our feature extraction results to the features tabulated by Janko, and perhaps track the features he analyzed more closely in our analyses.

Finally, we would like to analyze the fragments of the Epic Cycle, but since there are only about two dozen lines from a variety of texts, there is not enough data to really make any conclusive claims about the fragments from a computational standpoint. However, one could still use the features identified as "Homeric" by a classifier to assist in a manual analysis.

## 9.4 Further applications of these tools

These tools can also be applied to analyzing many other Greek texts. τάμνον, for example, was originally designed for analyzing the plays of Euripides and could be applied to Greek texts beyond just hexameter for a variety of tasks. One could imagine expanding ὠδικόν to apply the techniques to Latin hexameter scansion, or even expanding its range to allow scansion of the other types of meter employed by ancient Greek and Latin authors, of which there were many. There are also interesting analyses to be done on the Non-Homeric texts used in this work. For example, the features seem to categorize book 4 of *The Argonautica* as slightly different from the first three books, and ictus lengthening appears in only a few of the books of the *Dionysiaca*, including many of those that seem to attempt an archaic style. These and other interesting observations could be explored in more depth.

# 10 Conclusion

In this paper, we used computation techniques to analyze allegedly Homeric texts and contribute to discussions surrounding these texts in a new way. We determined that previous research and tools in this area did not fully address our needs for extracting features from the text, so we created our own tools. ᾠδικόν uses one of two slightly different approaches to scan hexameter with more accuracy than previous tools, then extracts metrical features from the scanned texts. τάμνον analyzes the dialect features of various tokens within the text and provides information about which rules specifically contributed to that dialect, a feature lacking from the only other existing metrical analyzer. We evaluated these two tools to show that they are very accurate and represent a clear improvement over previous tools in the area. We then analyzed a variety of Greek hexameter texts, from the ancient works of Hesiod to Nonnus' work halfway through the first millennium C.E., and discovered that the *Iliad* and *Odyssey* are very similar and appear qualitatively to be a distinct group separate from the *Homeric Hymns* and other texts. Although we were unable to conclusively show stark differences between the *Iliad*/*Odyssey* and three of the *Homeric Hymns* from a quantitative standpoint, we did manage to show that the shorter Hymns and the Hymn to Hermes *are* quite different from the *Iliad*/*Odyssey* and the other three extant large Hymns. We also showed that books of the *Iliad* and *Odyssey* that are thought to be possible later additions do not appear particularly unusual within this feature space. Lastly, we present a variety of future improvements that could allow a different analysis of these works or other texts as scholars continue to gain new insight into these old texts using new computational techniques.

# 11 Acknowledgments

First and foremost, I would like to thank my adviser, Professor Christiane Fellbaum, for her support and guidance from when I was a student in her seminar as a freshman to the completion of this thesis as a senior. She has always encouraged my research and helped me become a better scholar.

Professor Joshua Katz also played a key role in the creation of this thesis, pointing me in the right direction on everything from which grammars to consult for obscure scansion rules to which texts are thought to be outliers and which resources would help provide background on similar work done by classicists.

I owe many thanks to Professor Timothy Barnes, who first introduced me to the works of Homer and the Homeric Question, guided me as I worked on my junior independent work, and continued to help with any questions I had during the creation of this work. Without him I would never have even thought to ask the questions that led to this thesis.

This thesis is built on the foundation provided by the work of Gregory Crane and the Perseus Digital Library team. Their past efforts provided the infrastructure which made my tools and analyses possible and saved me countless headaches.

To my Colonial Club friends old and new, from semi-casual petitioners and promise reigniters to the new regimen: thank you for all the love, support, laughter, and friendship. Colonial has been my second home for two and a half years, and it may not have always been clean, but it *was* always a good time. Of all the titles I have been given at Princeton, I am proudest to wear the name "Fool."

To Chef Gil and the entire Colonial kitchen staff, whose food kept my spirits high and prevented me from withering into nothingness.

To all my wonderful friends in the classics department, who welcomed me as one of their own even though I am a fake. To Erynn, who always believed I could get this done and made sure I knew I would, even when I wasn't sure myself. To Selena, who always listened to my complaints with grace and kindness, even when she was dealing with thesis crises of far greater magnitude. The best answer to "why would you study Latin?" is that it lets me make friends like you two, and it has been a true pleasure being in all your classes every single semester.

To Elliot, who stuck with me through the Good Friday betrayal, freezing midnight walk from Quaker Bridge, and countless other things best not committed to writing. Being randomly placed in a room with you freshman year was the best decision I made at Princeton, and none of my achievements would have been possible without you "driving me into solitude."

To Elaine, who was here to welcome me to the East Coast with open arms all four years.

To Judy, who always made sure I was okay and never let me forget to take care of myself.

To Grandma, who has always been as excited as I am about everything I do at Princeton. I hope I have made you proud.

To my wonderful parents, who have provided me with unwavering love and support, whether I was a spoiled brat, angsty teen, or college student living a continent away. Thank you for always believing in me and fostering my love of learning.

To Owen, whose hard work and dedication is a continual source of inspiration for me. I couldn't have asked for a better brother.

And to Ann, whose laughter, enthusiasm, and courage made my world a brighter place.

# References


[1] W. S. Allen, *Vox Graeca: The Pronunciation of Classical Greek*. Cambridge University Press, 1987.

[2] Anonymous, *Homeric Hymns*, H. Evelyn-White, Ed. Harvard University Press, 1914. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0137

[3] H. R. Barnes, "Enjambement and oral composition," *Transactions of the American Philological Association (1974-)*, vol. 109, pp. 1–10, 1979.

[4] R. Beekes, "On the structure of the greek hexameter:'o'neill'interpreted," *Glotta*, vol. 50, no. 1./2. H, pp. 1–10, 1972.

[5] A. R. Benner, *Selections from Homer's Iliad*. Oklahoma Press, 1903.

[6] Bion, "Works of bion," in *The Greek Bucolic Poets*, J. M. Edmonds, Ed. William Heinemann, 1919. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0647

[7] G. M. Bolling, "Ionic forms in homer," *Language*, vol. 31, no. 2, pp. 223–231, 1955.

[8] C. D. Buck, *The Greek Dialects*. Chicago: The University of Chicago Press, 1995.

[9] A. W. Bulloch, "A callimachean refinement to the greek hexameter," *The Classical Quarterly (New Series)*, vol. 20, no. 02, pp. 258–268, 1970.

[10] J. S. Burgess, *The tradition of the Trojan War in Homer and the epic cycle*. JHU Press, 2001.

[11] Callimachus, *Works*, A. Mair, Ed. William Heinemann, 1921. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0481

[12] J. Clay, "The homeric hymns," in *A New Companion to Homer*, I. Morris and B. Powell, Eds. Brill, 1997.

[13] D. L. Clayman and T. Van Nortwick, "Enjambement in greek hexameter poetry," *Transactions of the American Philological Association (1974-)*, vol. 107, pp. 85–92, 1977.

[14] Colluthus, "Rape of helen," in *Oppian, Colluthus, Tryphiodorus with an English Translation*, A. Mair, Ed. William Heinemann, 1928. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0495

[15] G. Crane, "Generating and parsing classical greek," *Literary and Linguistic Computing*, vol. 6, no. 4, pp. 243–245, 1991. [Online]. Available: http://llc.oxfordjournals.org/content/6/4/243.abstract

[16] M. Eder, "How rhythmical is hexameter: A statistical approach to ancient epic poetry," *Digital Humanities 2008*, p. 112, 2008.

[17] K. P. J. et al., "Cltk: The classical language toolkit," https://github.com/cltk/cltk, 2014–2017, DOI 10.5281/zenodo.v0.1.46.

[18] A. Fick *et al.*, *Die Homerische Ilias: nach ihrer Entstehung betrachtet und in der ursprünglichen Sprachform widerhergestellt*. Vandenhoeck und Ruprecht, 1886, vol. 1.

[19] R. Fowler, "The homeric question," *The Cambridge Companion to Homer*, pp. 220–32, 2004.

[20] D. Fusi, "A multilanguage, modular framework for metrical analysis: It patterns and theorical issues," *Langages*, no. 3, pp. 41–66, 2015.

[21] B. Graziosi, *Inventing Homer: the early reception of epic*. Cambridge University Press, 2002.

[22] N. A. Greenberg, "A statistical comparison of the hexameter verse in" iliad" i, theognis, and solon," *Quaderni Urbinati di Cultura Classica*, vol. 20, no. 2, pp. 63–75, 1985.

[23] H. Hansen and G. M. Quinn, *Greek, an intensive course*. Fordham Univ Press, 1992, vol. 1.

[24] Hesiod, "Shield of heracles," in *The Homeric Hymns and Homerica*, H. Evelyn-White, Ed. Harvard University Press, 1914. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0127

[25] ——, "Theogony," in *The Homeric Hymns and Homerica*, H. Evelyn-White, Ed. Harvard University Press, 1914. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0129

[26] ——, "Works and days," in *The Homeric Hymns and Homerica*, H. Evelyn-White, Ed. Harvard University Press, 1914. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0131

[27] Homer, *The Odyssey with an English Translation*, A. T. Murray, Ed. Harvard University Press, 1919. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0135

[28] ——, *Homeri Opera in five volumes*. Oxford University Press, 1920. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0133

[29] K. Hopkins, *Conquerors and Slaves*. Cambridge University Press, 1978.

[30] G. Horrocks, "Homer's dialect," *MNEMOSYNE-LEIDEN-SUPPLEMENTUM-*, pp. 193–217, 1997.

[31] R. Janko, *Homer, Hesiod and the Hymns: diachronic development in epic diction*. Cambridge University Press, 1982.

[32] ——, "πρῶτον τε καὶ ὕστατον αἰὲν ἀείδειν: relative chronology and the literary history of the greek epos," in *Relative chronology in early Greek epic poetry*, D. T. Haug, Ed. Cambridge University Press, 2012.

[33] M. S. Jensen, *The Homeric question and the oral-formulaic theory*. Museum Tusculanum Press, 1980, vol. 20.

[34] B. Jones, "Relative chronology in an 'aeolic phase' of epic," in *Relative chronology in early Greek epic poetry*, D. T. Haug, Ed. Cambridge University Press, 2012.

[35] H. Liddell and R. Scott, *A Greek-English Lexicon*. Clarendon Press, 1940. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.04.0057

[36] D. G. Miller, *Ancient Greek Dialects and Early Authors: Introduction to the Dialect Mixture in Homer, with Notes on Lyric and Herodotus*. Walter de Gruyter, 2014.

[37] A. Mojena, "The behaviour of prepositives in theocritus' hexameter," *Glotta*, vol. 70, no. 1./2. H, pp. 55–60, 1992.

[38] D. B. Monro, *A grammar of the Homeric dialect*. Oxford, Clarendon P, 1891.

[39] Moschus, "Works of moschus," in *The Greek Bucolic Poets*, J. M. Edmonds, Ed. William Heinemann, 1919. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0647

[40] N. of Panopolis, *Dionysiaca*, W. Rouse, Ed. Harvard University Press, 1942. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0485

[41] Oppian, "Cynegitica," in *Oppian, Colluthus, Tryphiodorus with an English Translation*, A. Mair, Ed. William Heinemann, 1928. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0489

[42] ——, "Halieutica," in *Oppian, Colluthus, Tryphiodorus with an English Translation*, A. Mair, Ed. William Heinemann, 1928. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0488

[43] E. C. Papakitsos, "Computerized scansion of ancient greek hexameter," *Literary and Linguistic Computing*, vol. 26, no. 1, pp. 57–69, 2011.

[44] A. Rhodius, *Argonautica*, G. W. Mooney, Ed. Longmans, 1912. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0228%3Atext%3DId.

[45] E. S. Sherratt, "'reading the texts': archaeology and the homeric question," *Antiquity*, vol. 64, no. 245, pp. 807–824, 1990.

[46] Q. Smyrnaeus, *The Fall of Troy*, A. S. Way, Ed. William Heinemann, 1913. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0490

[47] A. Solensis, *Phaenomena*, G. R. Mair, Ed. William Heinemann, 1921. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0483

[48] G. Storey, "Dialect markers in the lyric sections of the plays of euripides," unpublished manuscript, Computer Science Department, Princeton University.

[49] M. Strockis, "Greek hexameter analysis," http://www.thesaurus.flf.vu.lt/eiledara/index.php, accessed: 2017-02-23.

[50] Theocritus, *Idylls*, R. J. Cholmeley, Ed. George Bell & Sons, 1901. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0228%3Atext%3DId.

[51] Tryphiodorus, "The taking of ilios," in *Oppian, Colluthus, Tryphiodorus with an English Translation*, A. Mair, Ed. William Heinemann, 1928. [Online]. Available: http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0491

[52] M. West, "Homer's meter," *MNEMOSYNE-LEIDEN-SUPPLEMENTUM-*, pp. 218–237, 1997.

# A Appendices

## A.1 Texts Included

See Table 7 for a list of the hexameter texts analyzed. All texts are from the Perseus Digital Library.

## A.2 Scansion Rules

For scansion, in general a syllable is long if

1. The syllable's core is a naturally long vowel (η, ω, or long α,ι,υ), e.g., -τη-, or
2. The syllable's core is a diphthong (like αι, ευ), e.g., -σαι-, or
3. The syllable is closed (i.e. the vowel is followed by two consonants or a double consonant like ξ) e.g., -δεν-.

If a syllable is not long, it is short.

However, there are a variety exceptions that must be considered:

- A long vowel or diphthong at the end of a word is usually short when the next word begins with a vowel; e.g., in *Iliad* book 1 line 17, the αι in κραὶ ἄλλοι is short.
- Two vowels can run together to form a single long syllable; this happens regularly with the genitive singular ending -εω. For example, in the very first line of the *Iliad*, the final -εω in Πηληϊάδεω scans as a single long.
- This can also happen with a vowel and a diphthong, e.g., in *Iliad* book 1 line 18, θεοὶ is scanned as a single syllable.
- A mute (π, β, φ, τ, δ, θ, κ, γ, χ) and a liquid (λ, ρ and later μ, ν) are occasionally pronounced together as a single consonant, so the previous syllable is not closed. For example, in the formula ἔπεα πτερόεντα προσηύδα the πρ at the start of προσηύδα is pronounced together and the previous alpha is short.
- A short vowel before the words Σκάμανδρος, Σκαμάνδριος, σκέπαρνον, Ζάκυνθος, and Ζέλεις is not closed and remains short.
- A short vowel that falls before a word break in the middle of a foot can be lengthened (generally

by a pause in sense, but sometimes by the ictus alone). For example, in *Iliad* book 2 line 39, the γὰρ in γὰρ ἔτ᾽ ἔμελλεν is scanned long due to this ictus lengthening.

- A syllable may count as closed even if one of the following "consonants" is an unwritten digamma.

## A.3 The Metrical Features

The metrical features analyzed are as follows:

1. For each of the first five feet, is the foot dactylic (long short short, e.g., οὔνεκα, the first foot of *Iliad* book 1 line 11) or spondaic (long long, e.g., πολλὰς, the first foot of *Iliad* book 1 line 3)?
2. Spondaic runs: are there 2, 3, 4, or even 5 spondees consecutive in a single line? The natural element of dactylic hexameter was the dactyl, so use of spondees, especially consecutive spondees, put stress on the meter and was avoided [52].
3. Information on caesuras and diaereses: where in the meter do the word breaks occur? Caesuras are word breaks in the middle of feet, and diaereses are word breaks between feet. We also differentiate between "masculine caesura" (caesura between the arsis and thesis of the foot, i.e. after the first long) and "feminine caesura" (caesura within the thesis of the foot, i.e. between the two shorts).
4. Presence of correption. Correption occurs when a long vowel at the end of a word is followed by a vowel at the start of the next word, causing the initial vowel to be scanned as short (e.g., *Iliad* book 1 line 14, ἑκηβόλου Ἀπόλλωνος, where one would generally expect the finally syllable of ἑκηβόλου, as a diphthong, to be long, but is here short).
5. Presence of ictus lengthening: a short syllable can be scanned as long if it sits in the first syllable of a foot and is followed by a pause (or even if it is not followed by a pause).
6. Presence of the "mute + liquid" rule: In some cases, a "mute" consonant followed by a "liquid" consonant can be pronounced together as a single syllable (e.g., τά πρῶτα could be divided ταπ-ρω-τα (long long short) or τα-πρω-τα (short long short). We include three subfeatures: the frequency of all mute+liquids, mute+ρ/λ, and mute+μ/ν, as the μ/ν type is slightly rarer, especially in earlier works.

7. Meyer's laws of Alexandrian hexameter: these are three "laws" said to hold for all of Alexandrian hexameter.
   (a) Words which begin in the first foot do not end between the shorts or at the end of the second foot.
   (b) Disyllables scanning short-long do not occur before the primary caesura.
   (c) Words that begin in the middle of the 3rd foot and end in the middle of the 5th foot are avoided.

   Note that these laws do not always hold true for Homeric verse. The very first line of the *Iliad*, Μῆνιν ἄειδε, θεά, Πηληιάδεω Ἀχιλῆος for example, breaks all three. ἄειδε begins in the first foot and ends between the two shorts of the second foot, breaking the first law; θεά is a short-long disyllable just before the caesura, breaking the second law; and Πηληιάδεω leaves a word break in the middle of the 3rd and 5th feet simultaneously, breaking the third law.

## A.4 The Dialect Rules

See Table 8 and Table 9 for a list of the included dialect rules. The Ionic and Aeolic rules are from Buck's *The Greek Dialects* and the Homeric archaism rules are from the Homeric Grammar in Benner's *Selections from Homer's Iliad* [8][5]. Some information necessary for rule determination was taken from Hansen and Quinn's *Introduction to Ancient Greek* [23].

## A.5 The Code

You can find the code associated with this project, including both τάμνον and ᾠδικόν, at `https://github.com/storey/seniorThesisCode`.

| Work | Attributed to | Lines | Source |
|---|---|---|---|
| The *Iliad* (24 books) | Homer | 15,683 | [28] |
| The *Odyssey* (24 books) | Homer | 12,107 | [27] |
| The *Homeric Hymns* | Anonymous | 2,342 | [2] |
| *Hymns* | Callimachus | 1,012 | [11] |
| *Abduction of Helen* | Colluthus | 394 | [14] |
| *Shield of Heracles* | Hesiod | 479 | [24] |
| *Theogony* | Hesiod | 1,046 | [25] |
| *Works and Days* | Hesiod | 831 | [26] |
| Assorted Works | Moschus | 446 | [39] |
| *Dionysiaca* (48 books) | Nonnus of Panopolis | 21,261 | [40] |
| *Cynegetica* (4 books) | Oppian | 2,144 | [41] |
| *Halieutica* (5 books) | Oppian | 3,506 | [42] |
| *Fall of Troy* (14 books) | Quintus Smyrnaeus | 8,800 | [46] |
| *Idylls* | Theocritus | 2,975 | [50] |
| *The Taking of Ilios* | Tryphiodorus | 691 | [51] |
| *Argonautica* (4 books) | Apollonius of Rhodes | 5,835 | [44] |
| *Phaenomena* | Aratus Solensis | 1,155 | [47] |
| Assorted Works | Bion of Phlossa | 246 | [6] |

**Table 7: The Texts Analyzed. We further break the Homeric Hymns into the first five "long" hymns and the 28 "short" hymns.**

| Shorthand | Rule Description | Section |
|---|---|---|
| SW.1 | ἄς = ἕως | Buck 41.4 |
| SW.2 | -αος vs -εως | Buck 41.4 |
| SW.3 | -σ-/-σσ-/-ττ- variants | Buck 82 |
| SW.4 | Forms of the plural personal pronoun | Buck 119.2, .5 |
| SW.5 | The conjunction εἰ | Buck 134.1 |
| SW.6 | The particle ἐάν | Buck 134.1b |
| SW.7 | The particle ἄν | Buck 134.2 |
| SW.8 | ἄτερος = ἕτερος | Buck 13a |
| SW.9 | δέκομαι = δέχομαι | Buck 66 |
| SW.10 | ὄνυμα = ὄνομαι | Buck 22c |
| SW.11 | Forms of ἕνικα | Buck 144a |
| SW.12a | Adverbs ending in -ου | Buck 132.1 |
| SW.12b | Adverbs ending in -ει | Buck 132.2 |
| SW.12c | Adverbs ending in -θεν | Buck 132.9 |
| SW.12d | Adverbs ending in -κα vs -τε | Buck 132.11 |
| SW.13 | Ionic -ει- for attic -ε- | Buck 54 |
| SW.14 | δείλομαι = βούλομαι | Buck 75.b |
| SW.15 | ἰαρός = ἱερός | Buck 13.1 |
| SW.16 | Forms of ἐκεῖνος | Buck 125.1 |
| SW.17 | Forms of κοινος | Buck 137.7 |
| SW.18 | Forms of κρατερός | Buck 49.2a |
| SW.19 | Forms of δημιουργός | Buck 167 |
| SW.20 | Forms of εὐθύς | Buck Glossary |
| SW.21 | Forms of μία | Buck 114.1 |
| SW.22 | Homeric forms of γονυ, δορυ, ζευς, ναυς | Benner 97, 98, 101 |
| SW.23 | Homeric forms of πολυς | Benner 105-106 |
| SW.24 | Homeric πτολις | Buck 104 |

**Table 8: A list of single-word rules included in τάμνον.**

| Shorthand | Rule Description | Section |
|---|---|---|
| NE.1a | Endings of singular feminine long alpha-stems | Buck 8 |
| NE.1b | Endings of singular feminine short alpha-stems | Buck 8 |
| NE.2 | Singulars of masculine alpha stems | Buck 41.4, Benner 65 |
| NE.3 | Plurals of alpha stems | Buck 41.4, Benner 65 |
| NE.4 | Forms of digammma stems | Buck Buck 43, 111.3, Benner 86 |
| NE.5 | Forms of iota stems | Buck 109, Benner 103 |
| NE.6 | Dative plural in -εσσι | Buck 107.3 |
| NE.7 | Homeric second declension endings | Benner 73-74 |
| VE.1 | Ionic μι-verbs inflected like contracts | Buck 160 |
| VE.2 | Third person middle forms | Buck 139.2 |
| VE.3 | Alpha contract endings | Buck 41.1 |
| VE.4 | Athematic 3rd plural secondary ending | Buck 138.5 |
| VE.5 | Active infinitive endings | Buck 154.1 |
| VE.6 | Homeric verb endings | Buck 136, 137, 142 |
| NM.1 | Nu Movable (simple) | Buck 102 |

**Table 9: A list of generally applicable rules included in τάμνον.**