Proceedings of the 32nd Annual UCLA Indo-European Conference

November 5th, 6th and 7th, 2021

Edited by

David M. Goldstein
Stephanie W. Jamison
Brent Vine

with the assistance of

Angelo Mercado

BUSKE

Contents

One or Many Homers?
Using Quantitative Authorship Analysis
to Study the Homeric Question

CHIARA BOZZONE   RYAN SANDELL

Ludwig-Maximilians-Universität München

This paper applies techniques of *quantitative authorship analysis* (QAA) to the Homeric corpus (*Iliad* and *Odyssey*) to attempt to shed light on the composition and internal structure of these works. The primary objectives are to demonstrate a) that QAA can replicate the modern *communis opinio* on major structural divisions within the corpus (e.g., that the *Iliad* and the *Odyssey* should be ascribed to minimally two different authors and that *Iliad* 10 stands out within the Homeric corpus), and b) that QAA can be used to evaluate which among existing models of the textualization of Homer's epics appears more likely. Specifically, results obtained using hierarchical clustering techniques indicate a) that each of the two Homeric epics admits of groupings that appear independently credible in terms of language and content, and b) that a multi-event model of textualization involving multiple authors is overall more plausible than a single-event model.

1 Introduction

One of the most enduring puzzles in the study of Western literature is the so-called Homeric question, i.e., the set of interconnected problems concerning how and when the *Iliad* and the *Odyssey* were first composed and how they came to assume their current form. Numerous solutions have been explored since antiquity: already in Alexandria one could pit the lumpers (who thought both the *Iliad* and the *Odyssey* where the work of a single poet)[1] against the χωρίζοντες 'splitters' (who thought the *Iliad* and the *Odyssey* were the work of different authors); their nineteenth- and twentieth-century counterparts are the Unitarians (who believe each poem reflects the work of a single poet) and the Analysts (who aim to uncover "different hands" within each poem). The modern understanding of the technique

---

1   The most famous Unitarian in antiquity is perhaps the author of the treatise *On the Sublime* (IX.11–5), who believed that Homer composed the *Iliad* in his youth and the *Odyssey* in his old age.

of Homer as oral and traditional in nature (cf. Lord 1960, Parry 1971) has further complicated this issue, and produced new variations on the usual themes.[2]

Nowadays, based on a series of qualitative arguments, most scholars of Homer subscribe to the following views:

- The *Iliad* and the *Odyssey* were likely composed by at least two different authors[3] (though a few radical Unitarians remain, such as, for instance, Wachter 2007 and Janko 2012).

- The *Iliad* was composed prior to the *Odyssey* (for an overview of the arguments, see Andersen and Haug 2012:1–19).

- Oral tradition played some part in the composition of the poems, either simply as a necessary premise, as a means of transmission, or as an actual means of composition.[4]

- The poems were not immune from later interpolations.[5] While in most cases these should be understood as small additions or subtractions of individual lines, most scholars agree that *Iliad* 10 (the *Doloneia*) in its entirety is such an addition.[6]

Beyond these points, disagreements are sharp, and the theories are many. A simplistic division can be set up between two different models of *textualization* of the poems. The first, which we will call the *Single-event Hypothesis* (e.g., West 2010, 2014), posits an individual author for each poem, who would have composed (and re-composed) the text over several decades, or even a lifetime, and committed it to writing. Few alterations would have happened afterwards. Some of these models tie the textualization to a special occasion and see the poems as *oral-dictated*

---

2   For a short history of the Homeric question, see Nesselrath 2011 and Turner 2011.
3   As Wilamowitz (cited in Passa 2016:165) puts it: "whoever puts the language, religion, and customs of the *Iliad* and the *Odyssey* on the same plane does not deserve scientific consideration."
4   Foley 2007 is a good basic introduction to how the theory of orality impacts our understanding of the poems.
5   As West 1998:v writes: "ab aliis interpolatum esse poema manifestum est, mirumque esset, si aliter se res haberet" [That the poem contained interpolations from other sources is evident, and it would be remarkable if this had not been the case].
6   "The *Doloneia* is the only single extended passage within the *Iliad* which has been labelled a 'late addition' or 'not authentic' by most Homeric scholars, starting from the famous note in the scholia" (Danek 2012:106). A classic treatment is Danek 1988.

texts.[7] Some (such as West 2010, 2014) envision a writing poet. In any event, all of these models assume that each poem as we have it reflects a single grammar (possibly over the course of a few decades) and a grand design by an exceptionally gifted poetic mind.

The second model, which we will term the *Multiple-event Hypothesis* (Nagy 1996:52–4, 2009:4–6, 2020), views the *Iliad* and the *Odyssey* as the result of a long tradition, with both poems having reached their current forms gradually over the course of several centuries through the cumulative work of many individuals within that tradition, and not as the products of an individual author. In this scenario our texts (which represent the serendipitous results of a long textualization process, rather than a pre-existing grand design) would reflect the output of multiple grammars, over the span of many decades and perhaps even centuries. Under this model, no real difference holds between "original" and "interpolations," and all the multiple forms and variants of the text as we can garner from the surviving evidence are of equal value.[8]

In this divided landscape our contribution to the problem of Homeric authorship is to introduce a new method, very much distinct from previous efforts. Our method is quantitative and, in many ways, "dumb": it does not bring to the question all of the knowledge (and thus prejudices) that a trained philologist would, and it does not look at the same features that philologists have previously been taking into consideration. It is, however, a well-established method that has been successfully applied to numerous problems of authorship analysis over the past few decades. As such, we hope that it may provide *external support* for some existing theories and help to choose among them.

Our goals in this paper are modest: we aim to establish that techniques of quantitative authorship analysis can be profitably applied to the Homeric poems, and specifically to show:

- that they can replicate some of the main conclusions that scholars have converged upon using qualitative methods;

---

7   Viewing the poems as oral-dictated texts (following Lord 1991) usually implies positing a shorter time frame for composition, and no opportunity for the poet to revise and re-compose their work after dictation; these theories also bring attention to the role of the scribe(s) in the process of textualization. Recent studies bringing typological and comparative considerations to this issue are Jensen 2011 (who proposes a very specific time and place for the textualization of both epics, namely, the late summer of 522 BCE in Athens) and Ready 2019.

8   An editorial project reflecting this model of the Homeric text is being implemented in the Homer Multitext Project (https://www.homermultitext.org).

- that they can help to decide which of different existing models of the textualization of Homer's epics is perhaps most likely.

In what follows, we first provide a brief introduction to Quantitative Authorship Analysis (§2). In Section 3 we address the question as to whether a single author of the Homeric epics is likely. In Section 4 we turn to evaluating likely groupings within books of the *Iliad* in isolation from the *Odyssey*.

2 A short introduction to Quantitative Authorship Analysis

The core assumption underlying *Quantitative Authorship Analysis* (henceforth QAA) is that *linguistic style* present in a text ("document") can be used to determine likely *metalinguistic properties* of that text (e.g., genre, date of composition, the specific author) in opposition to other documents with different such properties (e.g., different genre, different author). For a general introduction to methods of QAA, see Juola 2006, Statamatos 2009, and Juola 2012.

The Homeric question belongs to the most difficult class of authorship problems, namely, unsupervised authorship analysis, which attempts to answer the following questions: do two or more documents of unknown authorship have the same author (*verification*)? How many distinct authors are likely present in a pool of anonymous documents? In contrast to *attribution* problems, in which documents belonging to plausible known authors are available, neither the *Iliad* nor the *Odyssey* can be attributed to a known author, nor do we know many authors we should seek to find in the corpus. No fail-safe methodology exists for this type of problem, i.e., to conclusively identify how many "distinct" authors are to be found among a set of documents.[9] What we present in this paper is a reasonable first attempt at attacking the issue.

Performing an unsupervised authorship analysis can be broken down into three steps: 1) establishing a linguistic corpus, 2) choosing and gathering features therein, and finally 3) assessing the similarity of the documents under consideration and discovering plausible groupings (*clusters*) of the documents. For the studies in Sections 3 and 4, all documents were extracted from digital editions provided by the *Thesaurus Linguae Graecae* (*TLG*), where each document corresponds to a traditional book (e.g., Book 1 of the *Iliad* or Book 7 of *Herodotus*).[10] The features

---

9   See Stover and Kestemont 2016 for one approach to a *verification* problem in Latin literature similar in nature to the Homeric question.

10   We are aware that the traditional book divisions employed for the Homeric epics may not correspond to their actual historical textualization; see Heiden 1998 and 2000 for discussion concerning the validity of the book divisions, with references to earlier literature.

employed consisted of *character trigrams* (i.e., sequences of three consecutive typographic characters, ignoring whitespace) and *word bigrams* (i.e., sequences of two orthographic words). Examples of some of the most frequent word bigrams and character trigrams are given in Tables 1 and 2; note there that these features are essentially devoid of topical content, consisting largely of sequences of function words or recognizable inflectional material. In both cases, the corpus was cleaned by removing all personal and place names and converting all characters to lowercase; for the bigrams, a small number of formulaic epithets assessed as too "contentful" were manually removed.

Table 1. Top 100 word bigrams, cleaned (Homer)

τε καὶ | τὸν δ | οἳ δ | δ ἄρα | ὃ δ | δ ἄρ | δέ οἱ | δ αὖτε | ἣ δ | δὲ καὶ | δ ἐν | τὴν δ | ὣς φάτο | αὐτὰρ ἐπεὶ | δέ μιν | ὣς ἔφατ | ἀλλ ὅτε | ὅτε δὴ | οὔ τι | οἱ μὲν | ἀλλ ἄγε | δ αὖ | δ ἐπὶ | οὐ γὰρ | δέ τοι | δ ἀπαμειβόμενος | ὣς ἔφαθ | ἔπεα πτερόεντα | ὃ μὲν | ὣς εἰπὼν | πτερόεντα προσηύδα | αὐτὰρ ὃ | καί μιν | νῦν δ | ὣς ἄρα | ἦ τοι | οὐ γάρ | οἳ δὲ | τὸν δὲ | ἀπαμειβόμενος προσέφη | ὅ γ | ἐν δ | αὖτε προσέειπε | τῷ δ | ἔφαθ οἱ | δ οὔ | δέ τε | ὅ γε | δ αὖτ | ὃ δὲ | οὐδέ τι | ὡς δ | δ οὐκ | ἐν δὲ | ἀντίον ηὔδα | ἐκ δ | δ ὅτε | δ ἠμείβετ | τὸν μὲν | τοῖσι δὲ | ἠμείβετ ἔπειτα | οὐ μὲν | εἰ δ | δ ἐς | μὲν ἄρ | τοῦ δ | τοὶ δ | καὶ τότε | δ ἐκ | ἀμφὶ δὲ | σὺ δ | μὲν γὰρ | δ ἐπ | δ ἐπεὶ | τοῖσιν δ | εἰ δέ | δ ὥς | δ ἔπειτα | αἳ δ | ἔκ τ | ἐνὶ φρεσὶ | ῥα καὶ | τὼ δ | αὐτίκα δ | ἄρα πάντες | ἐπεὶ δὴ | ἐπεὶ οὔ | ἐπὶ νῆας | μὲν γάρ | ἦ ῥα | εἰς ὅ | ἣ μὲν | ὅς τις | ἦ μάλα | δ οὐ | οὐδέ τις | φωνήσας ἔπεα | ἐς πατρίδα | σὺ δὲ | ὣς ἄρ

Table 2. Top 500 character trigrams, cleaned (Homer)

καὶ|μεν|σιν|οντ|ντε|αὐτ|ντα|μέν|νος|ντο|ενο|μὲν|ἀλλ|οιο|
πολ|περ|ται|νκα|τοι|κατ|ρον|εκα|οισ|μοι|σθα|ισι|νον|προ|
εσσ|όντ|τες|σαν|ατο|οὐδ|ςκα|θαι|ικα|ετο|τὸν|ναι|τος|τον|
άντ|λλο|λον|οςἀ|σσι|ειν|τερ|ους|ἐπε|ένο|αντ|τεκ|νἐπ|γὰρ|
ονἐ|ροι|πάν|ακα|θεν|σσε|ἄλλ|νδρ|πει|ἀμφ|αιο|όμε|δυσ|
τἀρ|νοι|παρ|ατα|ὀδυ|θυμ|ιστ|ουσ|μετ|ῖσι|οῖσ|στο|ςτε|τις|
ἐπὶ|νἐν|ςἐπ|λοι|ενα|ὐτὰ|κεν|μον|ονἀ|ὀνδ|ρος|ετα|εσθ|
ματ|φίλ|ἔπε|νἀλ|στα|γάρ|χαι|μιν|νἀν|ἀχα|οςπ|οςἐ|ήσα|
τέρ|ινὰ|ρος|ςἀν|δἄρ|νδὲ|ιπε|ινἐ|λλὰ|ςἐν|ωντ|νοὐ|ειτ|ιον|
πεὶ|τεπ|σας|πατ|ενἀ|καλ|ναὐ|ἄρα|ρὸς|μεγ|μάλ|ολλ|ρατ|
έμε|ἄρο|τορ|λος|ομέ|πόλ|ποι|οςδ|υσσ|θεο|τοῖ|ονα|μέγ|
νῦν|ποτ|σεν|σσα|σει|οὔτ|ατὰ|ςἀλ|τετ|μαι|ἔνθ|ονκ|οιτ|
ονδ|ονο|εὺς|ρων|ατε|σον|ἐνὶ|ἀνδ|ντι|αἰἀ|σαι|τοῦ|ησι|δὲπ|

Table 2 continued

δέτ\|σικ\|ἐγὼ\|κακ\|εντ\|άλα\|ατρ\|αμέ\|ειπ\|καί\|ωνὰ\|χος\|ταπ\|
ονἔ\|ειρ\|ἀρἐ\|αῦτ\|ρας\|δὲκ\|αςὰ\|ἐπι\|έει\|κον\|ερο\|ομε\|ίλο\|θον\|
πον\|ενέ\|οικ\|νπο\|νων\|τοκ\|εῖν\|υσι\|σιδ\|ταμ\|εγά\|ιτο\|οὐκ\|νπρ\|
τῶν\|ότε\|χον\|αιτ\|τοδ\|ονπ\|έγα\|ενε\|ιος\|οςκ\|κτο\|ιτα\|ἀντ\|οιμ\|
ινο\|αιὰ\|σσο\|άμε\|νμε\|ρεσ\|ςοὐ\|νὰμ\|δέμ\|σιπ\|νὰπ\|δὲπ\|οκα\|
τεδ\|χαλ\|τὸς\|που\|αἰὲ\|οσέ\|ςπρ\|νει\|νδα\|αικ\|λλε\|εύς\|ένη\|έλε\|
δὲμ\|ρῶν\|τας\|εςὰ\|φρο\|άων\|τατ\|νὰγ\|έμο\|ρου\|τοπ\|οῖο\|σιμ\|
τιν\|πεί\|νησ\|ίην\|των\|αμε\|ςὰχ\|ςαὺ\|ατέ\|τακ\|έων\|ἀργ\|αίν\|οςτ\|
νεἰ\|ίων\|ἀγα\|σατ\|ωνέ\|έρο\|νπε\|ίον\|ππο\|ςπε\|ςεἰ\|τρε\|σετ\|ασι\|
ονε\|στῆ\|στή\|ησε\|ἔκτ\|τελ\|αστ\|ςὰπ\|λέμ\|ὅτε\|δαῦ\|λων\|οςἔ\|
οὺς\|ςδὲ\|τεμ\|ενέ\|αιν\|ντί\|ςὰγ\|έσθ\|αιὲ\|ταδ\|άτο\|ετε\|έντ\|αςὲ\|
απά\|μαχ\|ινα\|ιῶν\|ἵππ\|ιμε\|νθα\|λει\|μεθ\|μάχ\|ὐτο\|ωνπ\|έτι\|
εὐρ\|έεσ\|ομα\|τηλ\|σθε\|αἰτ\|ςμε\|εις\|ροτ\|άτε\|τὴν\|ίοι\|ειδ\|υμὸ\|
ὅφρ\|ὸςὲ\|φρα\|ήσε\|στε\|ὑδέ\|νεκ\|απο\|μὸν\|αλκ\|όνο\|θαλ\|ιοὺ\|
ἀπο\|απρ\|ίης\|νατ\|αιῶ\|τρώ\|ενδ\|γεν\|σὶν\|νδε\|μος\|αἰπ\|μνη\|
ινε\|εθε\|λου\|υσε\|εςὲ\|έλα\|τρο\|οιδ\|ιθε\|ἔφα\|ὑτὸ\|είν\|δρῶ\|είρ\|
νῶς\|τῆρ\|αιδ\|ἐκα\|ρισ\|ςτο\|ςὸδ\|ρες\|ρέπ\|ιμέ\|δον\|μην\|ερὶ\|ἐπέ\|
αιπ\|αλέ\|χιλ\|εια\|φέρ\|ςπο\|ωνκ\|υσα\|ἀχι\|ιπο\|τοτ\|έον\|ὸςὰ\|
δέο\|επο\|τότ\|λεύ\|οςο\|ασθ\|ίνο\|ηστ\|μοῖ\|κρα\|λευ\|την\|χει\|μήτ\|
\|μελ\|στι\|ραδ\|νὰρ\|αἰδ\|ῆος\|ἐστ\|ετέ\|ραν\|οςμ\|νὰλ\|νου\|ἠδὲ\|
όεν\|αίμ\|εἰδ\|κεῖ\|ὸνὰ\|ινἔ\|αθε\|πεπ\|ανα\|οςα\|σιτ\|δος\|νἔπ\|αςδ\|
πρὸ\|αἰμ\|ιςὰ\|νπα\|λισ\|ήνη\|άσσ\|έμα\|εῖδ\|έρω\|λεμ\|ωνο\|ταν\|
οἰδ\|νύπ\|μητ\|αλλ\|δης\|ςᾶρ\|ὸςδ\|νήσ\|ονᾰ\|επα\|ικε\|αια\|δει\|
ντω\|ελε\|ργε\|ρὸν\|νδέ\|ιπα\|νας\|σκε\|πεσ\|εικ\|νῆα\|εἰς\|δὰπ

Features were extracted using the R package *stylo* (Section 3) or functions written by the second author, Sandell, supported by the R package *ngram* (Schmidt and Heckendorf 2021); this work and all subsequent analyses were conducted in R Version 4.1.2 (R Core Team 2021). From a *feature matrix*, which contains the frequency of each feature in each document, normalized for the document's length, the distance between each pair of documents may be calculated, imagining the documents in an *n*-dimensional space (where *n* = the number of documents).[11] Intuitively, the more similar the values in the feature matrix, the smaller the distance between two documents. Finally, the resulting *similarity matrix* may be passed to a *clustering algorithm*, which arranges objects into groups based on their similarity. Hierarchical clustering methods, such as the average linkage algorithm (which we

---

11   Specifically, we employed Burrow's Classic Delta (Burrows 2002) for analyses in Section 3 and standard cosine distance for analyses in Section 4.

have employed throughout), generate *dendrograms* (see Figure 1), where each
node in the tree may be interpreted as a potential group.[12]
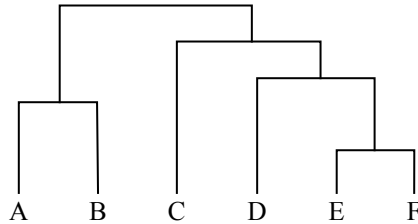

Fig. 1. A schematic cluster dendrogram

For more detailed exposition and technical details concerning the methodology
outlined above, as well as code for reproducing the analyses and visualizations in
Sections 3 and 4, we refer the reader to files available at https://github.com/
rpsandell/WeCIEC32.

3 Is a single Homer likely?

The objective of this first study is to test whether a single "Homeric" authorial
signal can be identified to the exclusion of other known ancient Greek authors,
following the criterion of a *common root node*. This criterion is relatively weak,
and potentially easy to satisfy; it merely proposes that, if all documents belonging
to the *Iliad* and *Odyssey* are to be assumed to share a single author, a hierarchical
clustering algorithm ought to create a node in the tree (cf. Figure 1) that includes
precisely those documents and excludes all documents known (or typically be-
lieved) to have a different author.

Concretely, we fed the works of several known ancient Greek authors to a clus-
tering algorithm, in order to test whether the features and algorithm could success-
fully group the works of known authors under a single root node for each author to
the exclusion of documents by other known authors (thus picking up on some sort
of *authorial signal*).[13] Since that indeed turned out to be the case (with one single
small but interesting exception), we were interested in verifying whether such an
authorial signal could be identified for the Homeric corpus or parts thereof.

---

12 The dendrogram in Figure 1 should be read from the bottom up and was constructed as follows:
   documents E and F were the closest, and were grouped (i.e., clustered) together first; then A and
   B were the closest remaining, and were grouped together; then the average of E and F was closest
   to D; the average of DEF was then closest to C; finally, the two nodes AB and CDEF were joined
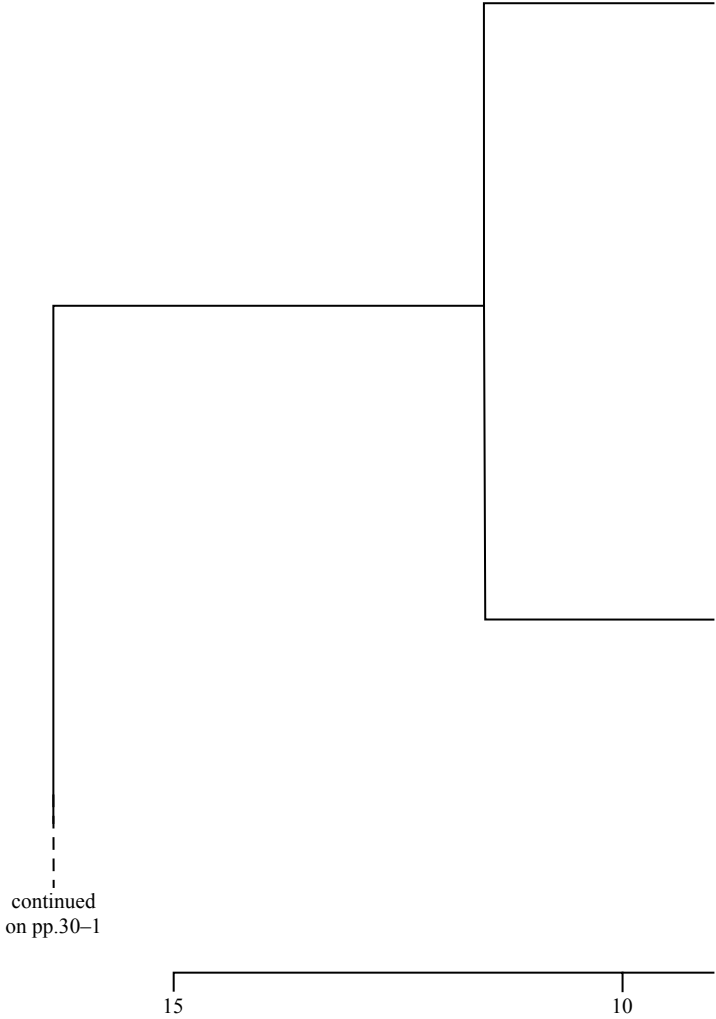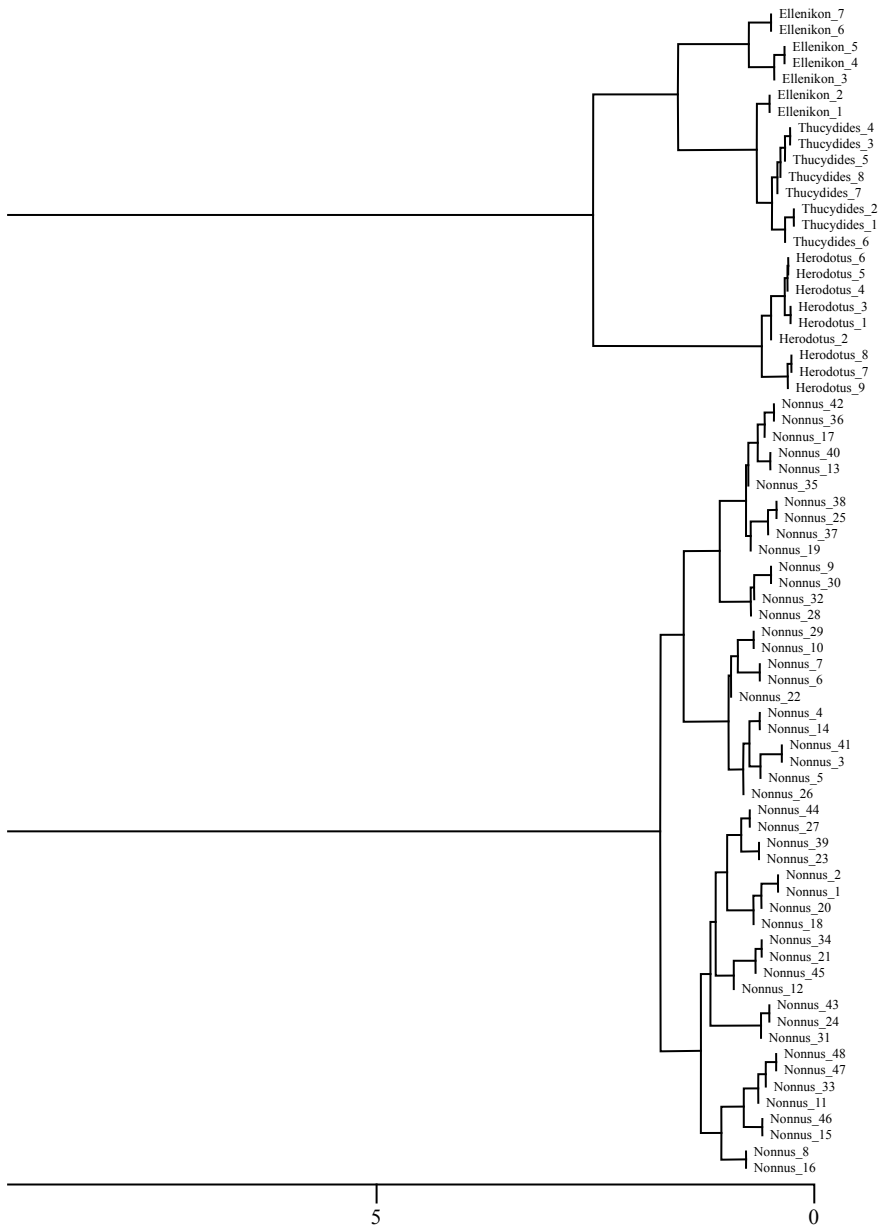   at the root.

15   10

Figure 2. Dendrogram using top 53 word bigrams,

```
                                        Ellenikon_7
                                        Ellenikon_6
                                        Ellenikon_5
                                        Ellenikon_4
                                        Ellenikon_3
                                        Ellenikon_2
                                        Ellenikon_1
                                        Thucydides_4
                                        Thucydides_3
                                        Thucydides_5
                                        Thucydides_8
                                        Thucydides_7
                                        Thucydides_2
                                        Thucydides_1
                                        Thucydides_6
                                        Herodotus_6
                                        Herodotus_5
                                        Herodotus_4
                                        Herodotus_3
                                        Herodotus_1
                                        Herodotus_2
                                        Herodotus_8
                                        Herodotus_7
                                        Herodotus_9
                                        Nonnus_42
                                        Nonnus_36
                                        Nonnus_17
                                        Nonnus_40
                                        Nonnus_13
                                        Nonnus_35
                                        Nonnus_38
                                        Nonnus_25
                                        Nonnus_37
                                        Nonnus_19
                                        Nonnus_9
                                        Nonnus_30
                                        Nonnus_32
                                        Nonnus_28
                                        Nonnus_29
                                        Nonnus_10
                                        Nonnus_7
                                        Nonnus_6
                                        Nonnus_22
                                        Nonnus_4
                                        Nonnus_14
                                        Nonnus_41
                                        Nonnus_3
                                        Nonnus_5
                                        Nonnus_26
                                        Nonnus_44
                                        Nonnus_27
                                        Nonnus_39
                                        Nonnus_23
                                        Nonnus_2
                                        Nonnus_1
                                        Nonnus_20
                                        Nonnus_18
                                        Nonnus_34
                                        Nonnus_21
                                        Nonnus_45
                                        Nonnus_12
                                        Nonnus_43
                                        Nonnus_24
                                        Nonnus_31
                                        Nonnus_48
                                        Nonnus_47
                                        Nonnus_33
                                        Nonnus_11
                                        Nonnus_46
                                        Nonnus_15
                                        Nonnus_8
                                        Nonnus_16
        5                               0
```

53 MFW Culled @ 0%
Classic Delta distance

Burrow's Classic Delta, average-linkage clustering

continued
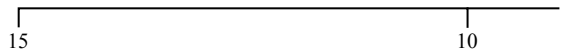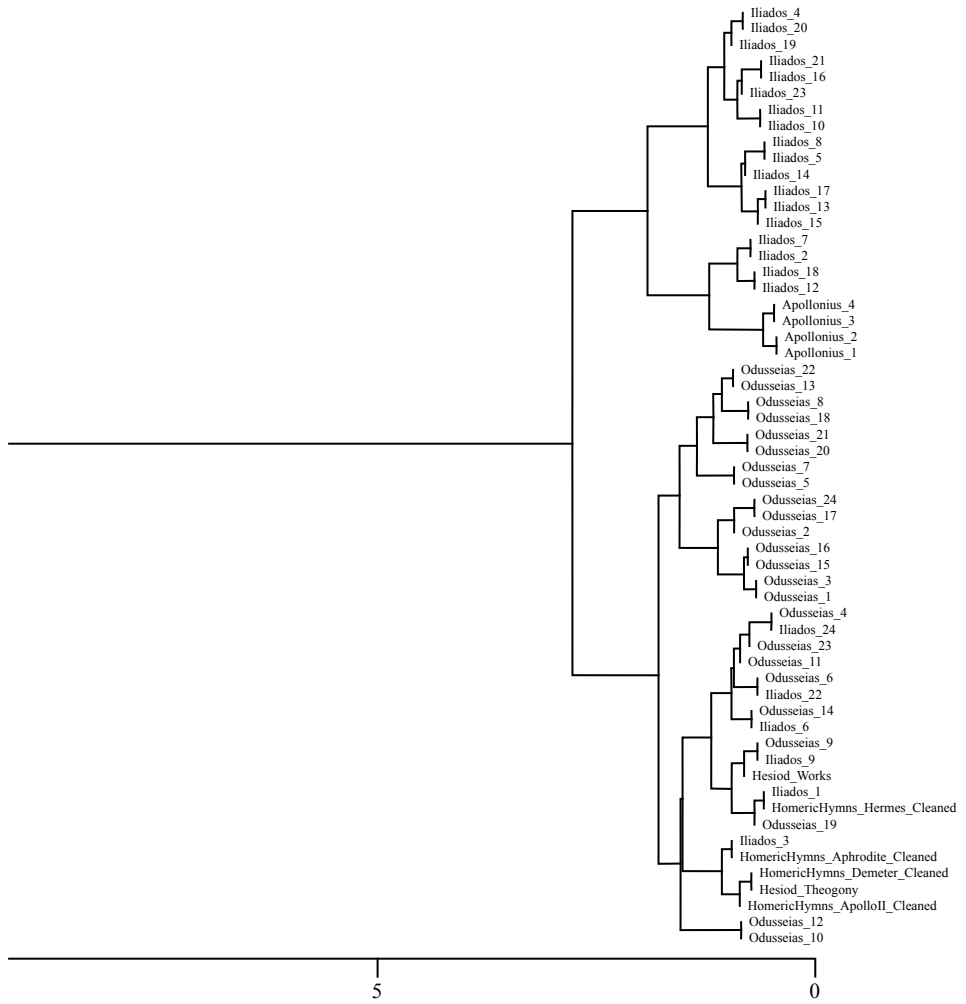from pp.28–9

15   10

Figure 2

Iliados_4
Iliados_20
Iliados_19
Iliados_21
Iliados_16
Iliados_23
Iliados_11
Iliados_10
Iliados_8
Iliados_5
Iliados_14
Iliados_17
Iliados_13
Iliados_15
Iliados_7
Iliados_2
Iliados_18
Iliados_12
Apollonius_4
Apollonius_3
Apollonius_2
Apollonius_1
Odusseias_22
Odusseias_13
Odusseias_8
Odusseias_18
Odusseias_21
Odusseias_20
Odusseias_7
Odusseias_5
Odusseias_24
Odusseias_17
Odusseias_2
Odusseias_16
Odusseias_15
Odusseias_3
Odusseias_1
Odusseias_4
Iliados_24
Odusseias_23
Odusseias_11
Odusseias_6
Iliados_22
Odusseias_14
Iliados_6
Odusseias_9
Iliados_9
Hesiod_Works
Iliados_1
HomericHymns_Hermes_Cleaned
Odusseias_19
Iliados_3
HomericHymns_Aphrodite_Cleaned
HomericHymns_Demeter_Cleaned
Hesiod_Theogony
HomericHymns_ApolloII_Cleaned
Odusseias_12
Odusseias_10

5   0

53 MFW Culled @ 0%
Classic Delta distance

continued

To these ends, we compiled a large corpus of ancient Greek hexametric poetry and historical prose, comprising a total of 130 documents and including the *Iliad* and *Odyssey* (48 documents in total); *Theogony* and *Works and Days* (Hesiod; two documents); four longer *Homeric Hymns* (Aphrodite, Demeter, Hermes, and Apollo; four documents); *Argonautica* (Apollonius Rhodius; four documents); *Dionysiaca* (Nonnus; 48 documents); *Historiae* (Herodotus; nine documents); *History of the Peloponnesian War* (Thucydides; eight documents); and *Hellenica* (Xenophon; seven documents). All texts were also purged of diacritics for reasons of known discrepancies (cf. n.23 below) in editorial practices.

The similarity between the documents was calculated by considering the 53 most frequent word bigrams,[14] using Burrow's Classic Delta as a distance measure; clusters were then assembled using average-linkage clustering. The results are visualized in Figures 2 through 4.[15]

In Figure 2 on pp.28–31 we can observe a clear top node split between "older" hexametric poetry (Homer, Hesiod, Hymns, Apollonius Rhodius in a large node at the bottom), versus Nonnus (documents in the lower node on p.29) and the historians (under the node at the top). For each distinct known author, a single node dominates all documents that should with certainty be ascribed to that author: an authorial signal is thus clearly identifiable for Herodotus, Thucydides, Xenophon *Hellenica* 3–7 (see further below), Nonnus, and Apollonius Rhodius. Note, for instance, that all 48 of the documents pertaining to the *Dionysiaca* (in the continuation of Figure 2 on p.29) build a cluster that excludes documents pertaining to any other (known) author. Thus, for each known author, we can find a common root node at some depth in the tree for the documents to be ascribed certainly to that author; meanwhile, we may interpret higher nodes (e.g., the common node dominating all documents of Herodotus, Thucydides, and Xenophon) as detecting similarity due to genre and/or chronology (in this case, these are all of the prose

---

13   In other words, each node in a dendrogram can be understood as a potential author. When all the documents belonging to a known or attributed to some hypothesized author are contained under a single node, to the exclusion of documents by other known authors, then an "authorial signal" may be recognized.

14   These are the bigrams that all occur in at least ninety percent of the documents. By using only features that are attested in most of the documents, we limit similarity that would be discovered by the simple "absence" of a feature.

15   We present the results of this particular feature set since it allows for a clear visualization of the results, given that trials with other feature sets (e.g., 500 most frequent character trigrams) or treatments of the documents (e.g., more documents of all equal size) yielded qualitatively similar results as concerns the primary question.

documents in our sample). In Figure 3 we zoom in on the top node, pertaining to the historians.
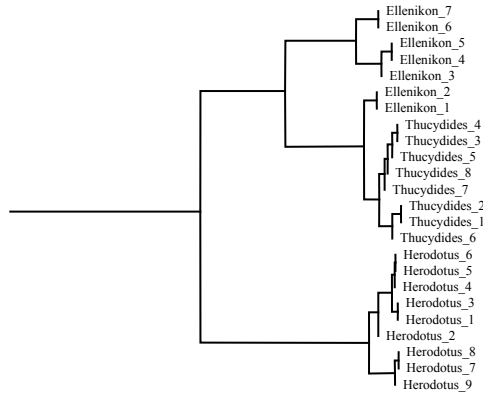
Fig. 3. The historians
(detail from the continuation of Figure 2 on p.29)

Here, we see an interesting wrinkle in our results: Books 1 and 2 of the *Hellenica* share a common root node with Thucydides. This corresponds to a known authorship problem, and it is widely believed that the first two books of the *Hellenica* may be based on original source material from Thucydides (cf. Thomson 1969). This combination of feature set and clustering algorithm thus seems to have correctly recognized Thucydides' authorial signal even "disguised" among Xenophon's writings.[16]

Emboldened by these results, we can now turn to the node comprising older hexametric poetry (Figure 4). Here our grounds for optimism seem to vanish—at least if we were interested in finding a clear authorial signal for Homer. The method successfully identifies a single authorial signal corresponding to Apollonius Rhodius, though the books of the *Argonautica* are in turn clustered with some (but not all) books of the *Iliad*. For Homer the situation is dire: no single node dominates all books of the *Iliad* and *Odyssey*. Their "last common ancestor" includes the *Argonautica*, Hesiod, and the four *Homeric Hymns*. As far as the individual poems are concerned, the books of the *Iliad* are split between two different top nodes (all including other materials); and even the *Odyssey*, which is grouped under a single

---

16   Other feature sets, meanwhile, instead cluster Books 1 and 2 of the *Hellenica* more closely with the remainder of the *Hellenica*. The most plausible interpretation, in accord with the existing literature on the problem, is that the first two books of the *Hellenica* indeed reflect a case of mixed authorship.

lower node, shares this node with parts of the *Iliad*, Hesiod, and the *Homeric Hymns*.[17]
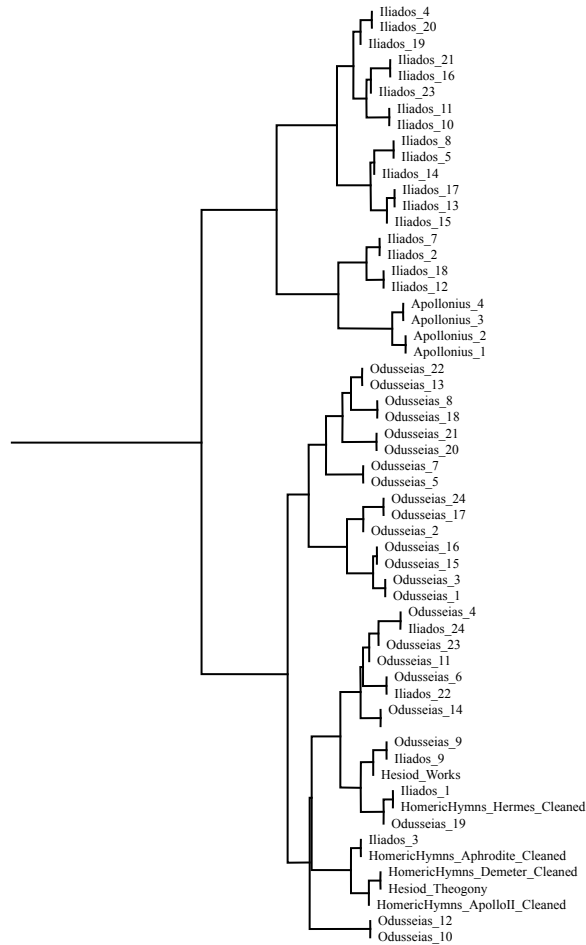
Fig. 4. Older hexametric poetry
(detail from the continuation of Figure 2 on p.31)

17 Interestingly, the works attributed to Hesiod seem to exhibit the same problem as Homer: the *Works and Days* and *Theogony* are not clustered together. This is consistent with the ancient Boeotian testimony that only the *Works and Days* were Hesiod's genuine work (Pausanias 9.31.4). Note that this skepticism is not shared by the modern critics, who usually see *Works and Days* and *Theogony* as stemming from a single author. See Cingano 2009 for the ancient evidence on the composition of the Hesiodic corpus.

A few comments are in order: our method seems to have yielded different results for texts created using oral formulaic technique (Homer, Hesiod, the *Homeric Hymns*), as opposed to texts composed in writing and by a known author (whether in verse or prose). The former yield no clear authorial signal, while the latter consistently do. Why should this be the case?

An oral-dictated text should in principle be no different from one that is written down by its author, in that it ought to reflect a single grammar. The peculiarities of such a single grammar ought to yield a particular authorial signature, which our method in turn attempts to detect.[18] If a single author were responsible for the *Iliad* (or most of the *Iliad*) as we have it, and if their text had been committed to writing, why does the signal of this author seem irrecoverable? To be sure, oral-formulaic texts (such as Homer, Hesiod, and the longer *Homeric Hymns*) share some general similarities in style (due to their reliance on traditional phraseology), but it is unclear that the traditionality of the style should make any authorial signal impossible to recover (we know, for instance, that there are many and clear differences in language and technique between Homer and Hesiod[19]).[20] At first sight, these results appear more compatible with a multiple-event scenario, or, at the very least, with a multiple-author-for-each-epic scenario. If this is indeed the case, could our methods tell us more about the internal structure of the epics, and how the different parts came to be assembled? It is to these questions that we next turn.

---

18 See Bozzone 2014:68–82 for the concept of an individual poetic grammar within an oral-formulaic tradition.

19 For the language and style of Hesiod, see Cassio 2009 and Hunter 2009 respectively.

20 We should not be tempted to believe that oral-formulaic composition in performance would have allowed for the verbatim faithful transmission of a patchwork of pieces composed by various poets, thereby preserving the individual signature of many different authors. As put in Bozzone 2014:78:

> I do not believe that there could be *verbatim* oral transmission of the poems after they were composed: at every new performance, the poems had to be generated anew through the I-Language [i.e., the individual grammar] of the poet who was performing them: they will then bear the constructional signature of this last individual in the chain of transmission. This is the classic oralist position (Lord 1960). While *verbatim* oral transmission is documented for some oral traditions (like that of the Rigveda), nothing leads us to believe that such methods of transmission were employed for Greek epic poetry; such transmission would require a kind of training and technique completely different from what Lord has described for the Serbo-Croatian tradition.

4 Internal groupings in the Homeric corpus

Can QAA establish any reliable internal groupings for the Homeric corpus, and do these groupings seem to align in some way with the scholarly consensus described above? In this section, we first examine the *Iliad* and the *Odyssey* taken together and then the *Iliad* individually.[21] In order to deliver results that are as informative and fine-grained as possible, some extensions of the basic method described in Section 2 are employed.

As far as features are concerned, we used both the top 100-word bigrams and the top 500-character trigrams, with analyses of the latter employing the Source Code Author Profiling (SCAP) method of Frantezkou et al. 2007.[22] For the analyses using word bigrams, we manually removed any personal and place names and other potentially contentful words (such as unique epithets) from the feature set. This was done to avoid topical effects (for example, that all books featuring Achilles might be grouped together on that basis). We also recombined some "duplicate features" that were due to different editorial practices in the respective texts of *Iliad* and the *Odyssey*.[23]

Instead of simply relying on a similarity matrix to quantify the distance between our documents, we applied *k-means clustering* prior to generating dendrograms, in order to obtain a clearer signal from our data. *K*-means clustering is a heuristic algorithm that sorts objects into groups, provided it knows ahead of time how many groups should be assumed to be present. Since the correct number of groups for Homer is unknown, the best approach here is to run *k*-means clustering hundreds of times, each time specifying a different number of groups (i.e., clusters), and see what groupings appear most robust (i.e., which documents are clustered together with greatest frequency). In this way one can build a *co-association matrix* (Layton, Wetters, and Dazeley 2011). This is essentially a spreadsheet with one

---

21 A similar individual treatment of the *Odyssey* is available in the online appendices at https:// github.com/rpsandell/WeCIEC32.

22 Under SCAP, the feature set is converted to a distance between each pair of two documents, where distance is measured as the proportion of overlap between the N most frequent character trigrams in each document.

23 Specifically, the edition of the *Odyssey* of von der Mühll (1962) frequently prints instances of the definite article or demonstrative pronoun ὅ, ἥ, τό without a grave, whereas Allen's (1931) edition of the *Iliad* prints a grave in most cases. For example, Allen's text contains exclusively οἳ δ᾽, whereas von der Mühll's contains only οἱ δ᾽ (without grave). The number of such highly frequent words that would be recognized by the computer as distinct would artificially inflate the degree of dissimilarity between documents belonging to the *Iliad* and *Odyssey*, respectively, when left uncorrected.

column and one row for each of the documents under study (an example is shown in Figure 5). Specifically, we applied *k*-means clustering one thousand times, each time selecting a different random value for the number of clusters, k ($10 \leq k \leq 24$, where the number of documents is 48); the number of times that two documents were assigned to the same cluster was then recorded in the spreadsheet (in the cell corresponding to that specific combination of books). Precisely because the *k*-means algorithm is heuristic, somewhat different results may be produced by each run of the algorithm; intuitively, objects that consistently fall into a cluster together when different values for *k* are selected may be regarded as constituting a more robust grouping.

	Iliados_1.txt	Iliados_10.txt	Iliados_11.txt	Iliados_12.txt	Iliados_13.txt	Iliados_14.txt	Iliados_15.txt	Iliados_16.txt	Iliados_17.txt
Iliados_1.txt	1000	3	0	0	0	0	0	0	0
Iliados_10.txt	3	1000	2	0	0	0	0	0	0
Iliados_11.txt	0	2	1000	173	804	163	564	819	602
Iliados_12.txt	0	0	173	1000	180	0	15	195	26
Iliados_13.txt	0	0	804	180	1000	214	746	983	712
Iliados_14.txt	0	0	163	0	214	1000	431	202	432
Iliados_15.txt	0	0	564	15	746	431	1000	729	944
Iliados_16.txt	0	0	819	195	983	202	729	1000	705
Iliados_17.txt	0	0	602	26	712	432	944	705	1000
Iliados_18.txt	448	5	7	0	0	0	0	0	0
Iliados_19.txt	368	3	0	0	0	0	0	0	0
Iliados_2.txt	0	0	2	630	1	0	0	2	2

Fig. 5. Example of co-association matrix

The co-association matrix was then used to calculate the cosine distance between the objects, and the result was in turn used to generate a dendrogram using average linkage clustering.[24] Summing up, the individual steps for QAA for each type of feature set were as follows:

- Word Bigrams: features (manual cleaning) > co-association > cosine distance > clustering

- SCAP (Character Trigrams): features > SCAP distance > co-association > cosine distance > clustering

---

24 Cosine distance was employed here for two reasons: 1) Burrow's Delta does not appropriately apply as a distance measure where the inputs are not frequencies of linguistic objects; 2) cosine distance has shown better performance on problems of authorship analysis as opposed to other standard measures of distance (e.g., Euclidean or Manhattan distance); cf. generally Evert et al. 2017. All of the figures below that have been generated using cosine distance will represent this distance as "Height," with values ranging from 0 to 1.

4.1 *The* Iliad *and the* Odyssey *taken together*

Our first goal was to test whether our methodology would be able to replicate some of the most agreed-upon conclusions concerning the Homeric question, namely, that the *Iliad* and the *Odyssey* stem from different times (and arguably different authors), and that *Iliad* 10 is an outlier in our corpus.
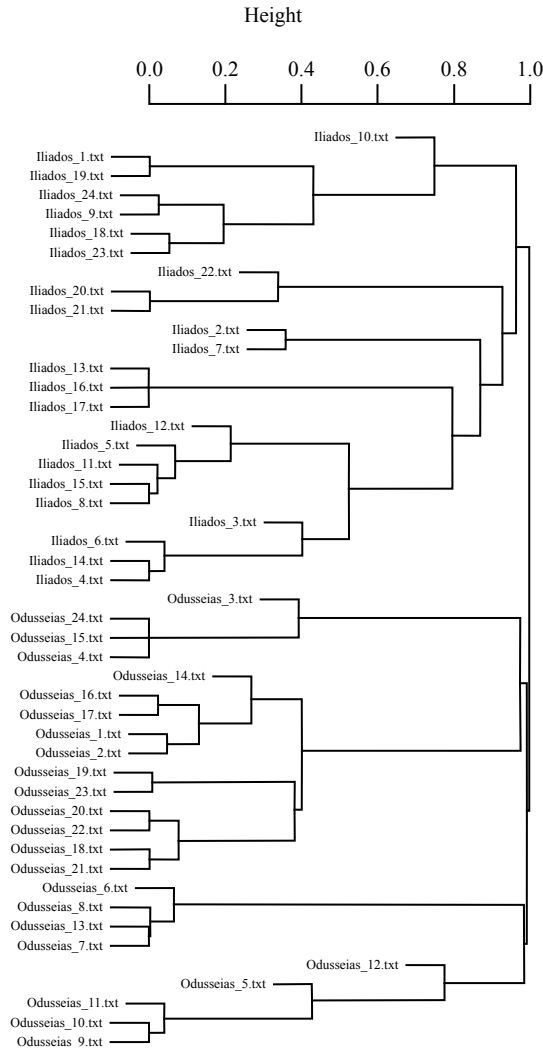

Fig. 6. SCAP 500 co-association matrix; cosine average

The results shown in Figure 6 seem to easily capture both statements: the dendrogram shows a clear split between the *Iliad* and *Odyssey* at the top, where each

poem falls under a single node; *Iliad* 10, moreover, is clearly isolated within the *Iliad*, being the very last document added to its cluster, and much higher up than the rest (where height on the tree corresponds to distance).[25] In addition, one may observe some promising internal structures for the individual epics themselves,[26] to which we will return in the following sections. While Figure 6 exhibits a clear split between the *Iliad* and *Odyssey*, it is important to stress that these methods can successfully capture similarities between the two epics as well. In Figure 7 the distance between our documents (i.e., individual books of the *Iliad* and *Odyssey*) is visualized in a 2-dimensional space, having applied a principal components analysis.[27] This visualization drives home two important points:

- It confirms how isolated *Iliad* 10 is among the Homeric corpus, having no close neighbors whatsoever.

- It shows that a few books of the *Iliad* lie "on the border" with the *Odyssey*: namely *Iliad* 1, 9, 23, and 24. These are precisely the books that have long been suspected of being later on the basis of linguistic features (as already observed by Monro 1891, as reported by Leaf 1900–1902:370).[28]

---

25 Among the books of the *Odyssey*, *Odyssey* 12 stands out in a fashion somewhat comparable to *Iliad* 10: the distance between it and the nearest cluster is greater than for any other book of the *Odyssey*.

26 For instance, the fact that *Iliad* 10 is clustered, within the *Iliad*, with several other books (such as *Iliad* 1, 9, 23 and 24, among others) that are known to exhibit late linguistic features (see the discussion below).

27 Principal component analysis (henceforth PCA) is a process that can be employed to reduce the dimensionality of the data. Instead of working with 100 dimensions (one for each feature), PCA will create new features that combine tiny pieces of all existing features, crafting the smallest number of features necessary to capture all variance in the data. Each further principal component explains a progressively smaller proportion of variance, for which reasons the first two principal components are usually best representative of patterns in the data. The two features represented in Fig. 7 explain 15.3% and 9.9% of the variance in the data (25% total). While this only represents a fraction of the total, this is enough to capture the sharpest distinctions.

28 Per Monro, linguistic features shared among these books are as follows: perfects in -κα from verbs in -έω; use of ἐπί with the accusative of extension over; ἐνί for μετά meaning 'among' with persons, and with abstract words; ἐκ meaning 'in consequence of'; use of the definite article (on which see Bozzone and Guardiano 2015, 2018); ἄν with the first person of the optative; ὅς τε with the infinitive; δεῖ for χρή; ἄν with the infinitive. A full treatment of each feature goes beyond the scope of this paper.
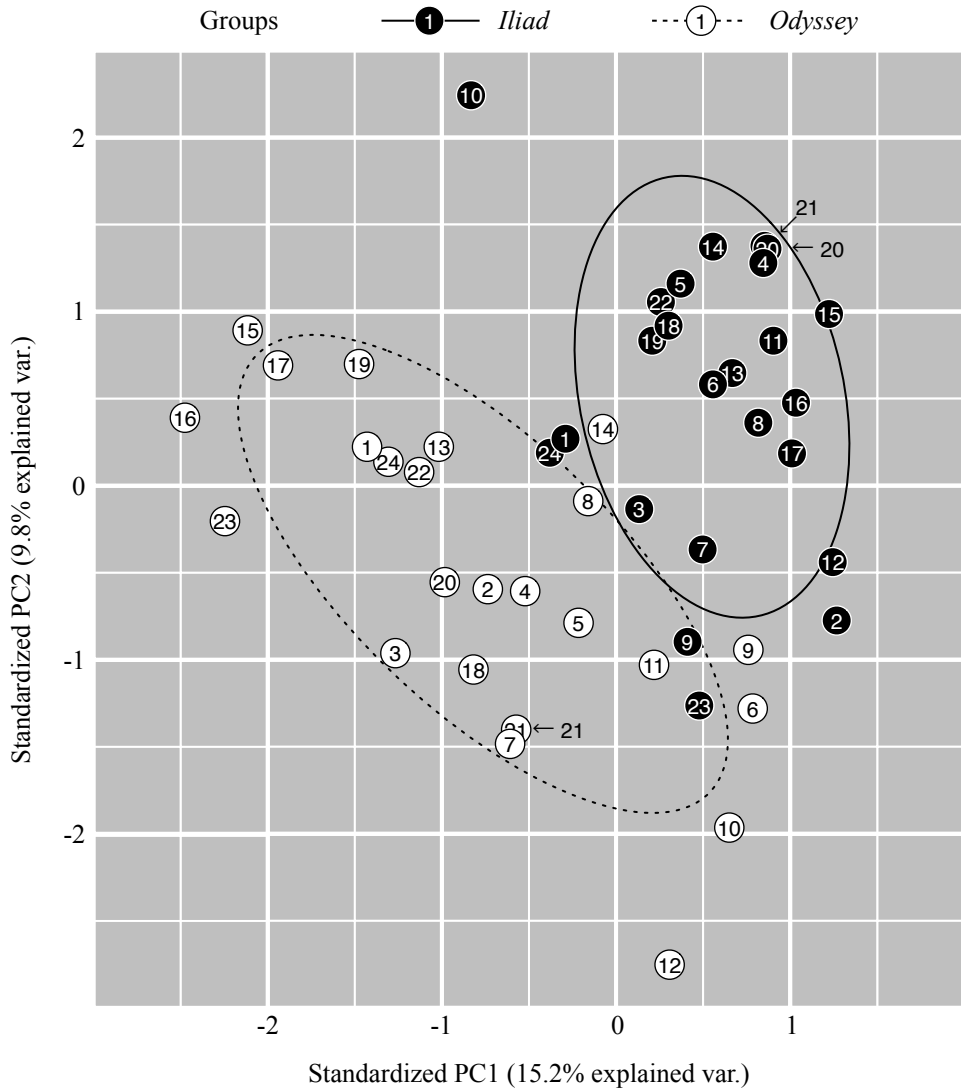
Fig. 7. Principal components analysis, top 100-word bigrams

4.2 The Iliad *by itself*

We will now consider the *Iliad* by itself and see whether our methods can help us to recover some plausible units of textualization. In Figure 8 we generated a dendrogram using the top 100 bigrams, a co-association matrix, cosine distance, and average linkage clustering, and cut it into four groups (an arbitrary number, given that the true number of clusters is unknown). The results obtained are in many

respects interesting (and, moreover, representative of several analyses using other feature sets tested for this study):

- *Iliad* 10 once again stands apart, forming its own group (Group 1)

- Group 3 represents a recurrent cluster (over several different QA analyses) consisting of well-behaved, unproblematic books. Content-wise, these are relatively less exciting books: Achilles is away from the battlefield, and a great deal of fighting takes place (most of it not particularly high-stakes, with the exception of *Iliad* 16).

- Groups 2 and 4, on the other hand, appear to cluster together some famous "troublemaker" books in the structure of the poem.

Quite remarkably, Group 2 contains some of the most famous analytical problems in the *Iliad*: these are all books that have been suspected of being secondary additions to the story of the μῆνις in terms of both content and language.[29] Several of these (*Iliad* 2, 3, 7) are almost universally regarded as adaptations of other materials, originally belonging to the first year of the war. In other cases, the language or content singles them out as late (*Iliad* 9, 12, 23).

Specifically, *Iliad* 2 contains the catalogue of the ships, which "as modern critics have almost unanimously recognized, was not composed for its present place, but was adapted to it" (Leaf 1900–1902:86). *Iliad* 3 comprises the τειχοσκοπία 'viewing from the walls', in which Priam does not recognize the main Greek chieftains (Agamemnon, Odysseus, Ajax, and Idomeneus) in the tenth year of the war.[30] Perhaps less famously, *Iliad* 7 contains arguably the greatest problem in all of the *Iliad*, namely, the construction of the wall protecting the Achaean ships in the tenth year of the war (another event that would be better suited for the earliest stages of the conflict). *Iliad* 9 contains the embassy to Achilles, also long suspected of being a later interpolation (and, as noted above, the language here is similar to *Iliad* 10, 23, and 24).[31]

---

29 Here and below, we rely on Leaf 1920 as a compact reference summarizing the results of the analytical line of inquiry concerning the structure of the *Iliad*. Another, more up-to-date (and extensive) resource on this topic is Zambarbieri 1988–1990. Zambarbieri 2002–2004 covers the *Odyssey*.

30 See Jamison 1994 for a comparative perspective on the episode and its meaning within the theme of the "counterabduction" of a bride. On the Helen myth in general, see now Edmunds 2016.

31 It is a known Homeric paradox that material alleged to be particularly ancient might be nested among the most recently textualized passages. This is the case for the alleged reflex of the PIE formula *k̂léwos n̥dʰgʷʰitóm* "imperishable fame," occurring as κλέος ἄφθιτον (*Il.* 9.413) in

Iliad 12, which recounts the battle at the wall protecting the Achaean ships (and the exploits of the Trojan allies Glaucus and Sarpedon), according to Leaf (1900–2:524) "cannot belong to any but the last strata of the *Iliad*" and is "lacking real artistic unity." Finally (as already noted) *Iliad* 23, which contains the funeral and funeral games of Patroclus, is closer in language and phraseology to *Iliad* 24, as well as the *Odyssey*.[32]
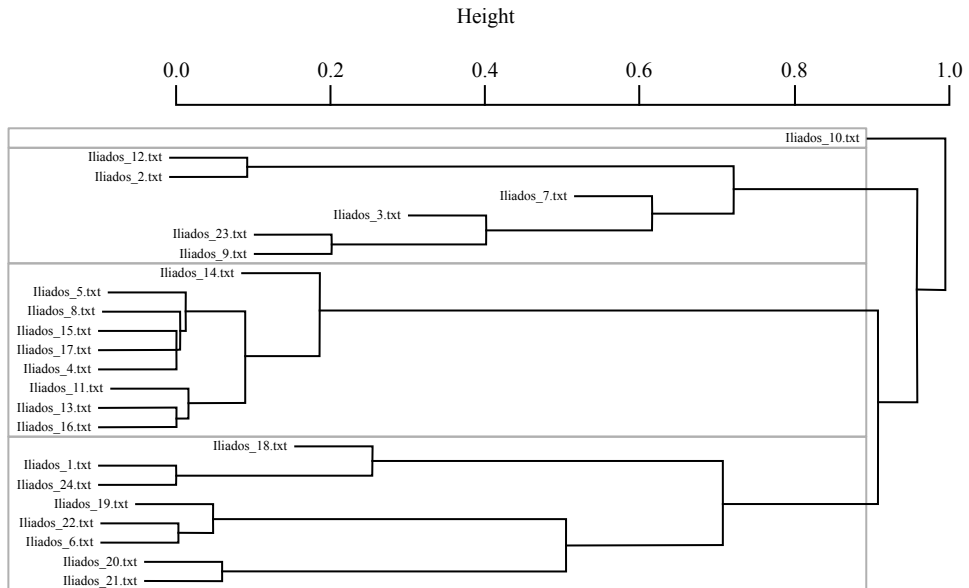

Fig. 8. Word bigrams top 100
(co-association matrix, cosine distance, average-linkage clustering)

Group 4 is also interesting in that it appears to contain some clear thematic units (despite the fact that our method relies on the frequency of non-content words). It is suggestive, for instance, that *Iliad* 1 and *Iliad* 24 would be grouped together, marking the beginning and end of the poem. This grouping could be interpreted as a trace of the process of textualization of the monumental poem. This configuration is perhaps reminiscent of the *Ṛgveda*, where the scholarly consensus

---

Achilles' famous reply to Agamemnon's offer (the closest comparandum is Vedic *śrávas … akṣitám* in *Ṛgveda* 1.9.7; see discussion in Watkins 1995:12–3, 173–8). A similar paradox exists in *Iliad* 10.260–5, which accurately describes a Mycenaean-age boar-tusk helmet, an item that would not have been in use in Greece for several centuries at the moment of textualization of the epics (Everson 2004:7–10).

32   See Bozzone 2014:81–3 for a list of *Odyssey*-like phraseological features in *Iliad* 23.

is that parts of Books 1 and 10 (the first and the last book respectively), which contain some of the youngest materials, make up a layer of textualization distinct from the Family Books (Books 2–7), which constitute the oldest nucleus of the *Ṛgveda*.[33] Specifically, one could envision a scenario in which the same poet, operating at a relatively more recent stage in the tradition (hence the more modern linguistic features in *Il.* 1 and 24) would have taken care to compose (or recompose) a beginning and an end for an *Iliad* that would have included a multitude of previous materials.[34]

Another intriguing grouping is that of Books 6 and 22, which contain the tragedy of Hector (and Andromache), as well as the most vivid rendering of their characters (these are also some of the most beloved and anthologized passages in Homer). In this node, one might tentatively want to see the (very skilled) hand of an individual poet.

The other books in this group are the preparation for Achilles' return to battle (*Iliad* 18 and 19) and his actual return (20 and 21). Perhaps unsurprisingly, *Iliad* 18, which contains the long description of the design of Achilles' shield (perhaps the most famous example of *ekphrasis* in ancient Greek literature), is the most isolated book in this group.

A SCAP analysis of the books of the *Iliad* delivers some results that are similar, though not identical, to what was just discussed. Again, *Iliad* 10 appears fairly isolated (it is the document that merged into a cluster tree at the greatest height, though not fully apart from all other documents, as in Figure 8), and we can contrast a cluster (on the right) of relatively unproblematic books (*Iliad* 5, 13, 16, 17, 12, 15, 11, and 8),[35] with a couple of clusters collecting "troublemaker" books: on the left, we can observe a cluster largely pertaining to Achilles (comprising *Iliad* 9, 1, 19, 18, 23, 22, and 24) and containing many linguistically recent books. In the middle, we can identify a subcluster containing all of the books pertaining to the first year of the war (namely, *Iliad* 2, 3, and 7).

Of course, there are some differences between the two models: some books are grouped differently in SCAP, such as *Iliad* 4, 14, 6 (here belonging to the middle cluster), and 12 (here grouped in the right-hand cluster). Some of the promising thematic units we discussed earlier (like *Iliad* 1 and 24, and *Iliad* 6 and 22) are also

---

33 See Jamison and Brereton 2014:14–8 on the textualization and transmission of the *Ṛgveda*.

34 To further support this scenario, one could also point to some clear thematic parallels between *Iliad* 1 and 24, where *Iliad* 1 begins with a father (Chryses) attempting to ransom his daughter, and *Iliad* 24 ends with a father (Priam) successfully ransoming his dead son (Hector).

35 Of these, however, we might remember that *Iliad* 12 was previously grouped with "troublemaker" books by the word bigram feature set.

not so clearly on view here. On the one hand, we should take this fact as a reminder not to become immediately attached to some appealing interpretations of the data, and not to take the results of a given QAA configuration as incontrovertible truth.[36] On the other hand, it is important to remark that some of these groupings are indeed stable across parameters of analysis, so that it is reasonable to assign them some validity.
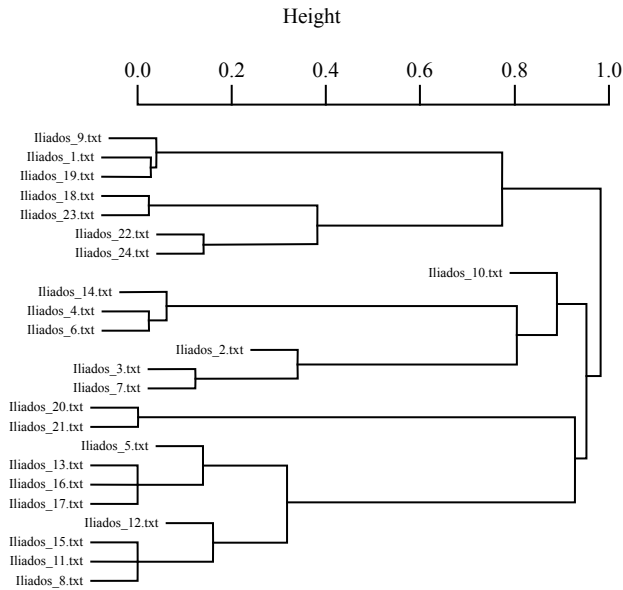
Height

| | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |

Iliados_9.txt
Iliados_1.txt
Iliados_19.txt
Iliados_18.txt
Iliados_23.txt
Iliados_22.txt
Iliados_24.txt
Iliados_10.txt
Iliados_14.txt
Iliados_4.txt
Iliados_6.txt
Iliados_2.txt
Iliados_3.txt
Iliados_7.txt
Iliados_20.txt
Iliados_21.txt
Iliados_5.txt
Iliados_13.txt
Iliados_16.txt
Iliados_17.txt
Iliados_12.txt
Iliados_15.txt
Iliados_11.txt
Iliados_8.txt

Fig. 9. SCAP 500 trigrams, *Iliad* only
(co-association matrix, cosine distance, average-linkage clustering)

5 Conclusion

The results presented here should be regarded as very much preliminary: much more work needs to be done in both refining the quantitative methods employed and aligning them more closely with existing theories of the composition of the Homeric epics. Nevertheless, the fact that some quantitative analyses based on innocuous-seeming linguistic features should yield results that replicate some key points of the current scholarly consensus (the isolation of *Iliad* 10, a clear split between *Iliad* vs. *Odyssey*), and that match known thematic groupings in the poems

---

36   It might be the case, however, that one method might prove better at "recovering the signal form the noise" than another one; only further research in this direction can provide such answers.

(the "first year of the war" books in the *Iliad*, or the ἀπόλογοι in the *Odyssey*[37]) is very encouraging. Concretely, we believe that this first pilot study supports the following conclusions.

5.1 Iliad *and* Odyssey *together: A single author unlikely*

The results of Section 3 speak against a single authorial presence for *Iliad* and *Odyssey* taken together. They also speak against a single authorial presence for the whole *Iliad* or the *Odyssey* taken in isolation (since individual books of Homer are routinely grouped together with other archaic or even Hellenistic hexametric poetry). In this light, the fact that the methods in Section 4 can find a clear division between the *Iliad* and the *Odyssey* most likely reflects a difference in the time of textualization (*Odyssey* overall later than *Iliad*), rather than two clear distinct authorial hands. Some books of the *Iliad* (exactly the ones that one might expect) stand closer to the *Odyssey*, which may also reflect a later chronology (this observation could be pushed towards speculation that the textualization of the *Odyssey* partially overlapped with the textualization of the youngest parts of the *Iliad*).

5.2 *Within the individual epics: A multiple-event model more likely*

Following the conclusions in Section 3, our results in Section 4 seem to better accord with a multiple-event model of textualization of the individual epics. The recurrent clusters that we have found and that correspond to recognized "thematic units" could reflect units of composition/textualization. These units might, in some cases, be ascribed to the work of a given individual poet (e.g., a "Hector's poet" for *Iliad* 6 and 22). Some might be linked specifically to the compilation of the *Iliad* as a single monumental poem (e.g., the cluster of *Iliad* 1 and *Iliad* 24). In some cases, they could simply reflect chronology (e.g., later additions to the *Iliad*, such as Book 10).

All of these observations are complicated by many additional considerations (such as the issue of book divisions, mentioned in n.10), to which we plan to turn in future work. Nevertheless, we hope to have shown that, when properly tuned, the techniques of QAA have the potential to help us to detect the subconscious habits of individual poets involved in the creation of the epics, and thus contribute to an untangling of the problem of Homeric authorship.

---

37   For which visit https://github.com/rpsandell/WeCIEC32.

References

Andersen, Øivind, and Dag T. T. Haug (eds.). 2012. *Relative Chronology in Early Greek Epic Poetry*. Cambridge: Cambridge University Press.

Bozzone, Chiara. 2014. Constructions: A New Approach to Formularity, Discourse, and Syntax in Homer. Ph.D. diss., University of California, Los Angeles.

Bozzone, Chiara, and Cristina Guardiano. 2015. Adnominal ὁ ἡ τό in Homer: Tracking the Spread of a Syntactic Innovation. Paper 23 March presented at the Colloquium on Ancient Greek Linguistics, Rome.

———. 2018. Adnominal ὁ ἡ τό in the Language of Homer: Syntactic Change or Stylistic Variation? Paper 19 April presented at the conference "Variation and Contact in the Ancient Indo-European Languages," Oxford and Pisa.

Burrows, John. 2002. "Delta": A Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing* 17.267–87.

Cassio, Albio Cesare. 2009. The Language of Hesiod and the *Corpus Hesiodeum*. In Montanari, Tsagalis, and Rengakos 2009, 179–201.

Cingano, Ettore. 2009. The Hesiodic Corpus. In Montanari, Tsagalis, and Rengakos 2009, 91–130.

Danek, Georg. 1988. *Studien zur Dolonie*. Vienna: Österreichische Akademie der Wissenschaften.

———. 2012. The Doloneia Revisited. In Andersen and Haug 2012, 106–21.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. Stylometry with R: A Package for Computational Text Analysis. *R Journal* 8.107–21.

Edmunds, Lowell. 2016. *Stealing Helen: The Myth of the Abducted Wife in Comparative Perspective*. Princeton: Princeton University Press.

Everson, Tim. 2004. *Warfare in Ancient Greece: Arms and Armour from the Heroes of Homer to Alexander the Great*. London: Sutton.

Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorseten Vitt. 2017. Understanding and Explaining Delta Measures for Authorship Attribution. *Digital Scholarship in the Humanities* 32.ii4–ii16.

Foley, John Miles. 2007. "Reading" Homer through Oral Tradition. *College Literature* 34.1–28.

Frantsezkou, Georgia, Efstathios Stamatatos, Stefanos Gritzalis, Carole E. Chaski, and Blake Stephen Howald. 2007. Identifying Authorship by Byte-Level N-Grams: The Source Code Author Profile (SCAP) Method. *International Journal of Digital Evidence* 6.1–18.

Heiden, Bruce. 1998. The Placement of "Book Divisions" in the *Iliad*. *Journal of Hellenic Studies* 118.68–81.

———. 2000. The Placement of "Book Divisions" in the *Odyssey*. *Classical Philology* 95.247–59.

Hunter, Richard. 2009. Hesiod's Style: Towards an Ancient Analysis. In Montanari, Tsagalis, and Rengakos 2009, 253–69.

Jamison, Stephanie W. 1994. Draupadi on the Walls of Troy: *Iliad* 3 from an Indic Perspective. *Classical Antiquity* 13.5–16.

Jamison, Stephanie W., and Joel P. Brereton (trans.). 2014. *The Rigveda: The Earliest Religious Poetry of India*. New York: Oxford University Press.

Janko, Richard. 2012. πρῶτον τε καὶ: Relative Chronology and the Literary History of the Early Greek Epos. In Andersen and Haug 2012, 20–43.

Jensen, Minna Skafte. 2011. *Writing Homer: A Study Based on Modern Fieldwork*. Copenhagen: The Royal Danish Academy of Sciences and Letters.

Juola, Patrick. 2006. Authorship Attribution. *Foundations and Trends in Information Retrieval* 1.233–334.

———. 2012. An Overview of the Traditional Authorship Attribution Subtask. *Proceedings of PAN/CLEF 2012*. Rome. https://pan.webis.de/downloads/publications/papers/juola_2012.pdf (12 October 2021).

Layton, Robert, Paul Wetters, and Richard Dazeley. 2011. Automated Unsupervised Authorship Analysis Using Evidence Accumulation Clustering. *Natural Language Engineering* 19:95–120.

Leaf, Walter. 1900–2. *The Iliad, Edited, with Apparatus Criticus, Prolegomena, Notes, and Appendices*. London: Macmillan.

Lord, Albert Bates. 1960. *The Singer of Tales*. Cambridge, MA: Harvard University Press.

———. 1991. *Epic Singers and Oral Tradition*. Ithaca: Cornell University Press.

Monro, Daniel Binning. 1891. *A Grammar of the Homeric Dialect*. Oxford: Clarendon Press.

Montanari, Franco, Christos Tsagalis, and Antonios Rengakos (eds.). 2009. *The Brill Companion to Hesiod*. Leiden: Brill.

Nagy, Gregory. 1996. *Homeric Questions*. Austin: University of Texas Press.

———. 2009. *Homer the Classic*. Washington, DC: Center for Hellenic Studies.

———. 2020: Review of Jensen 2011. https://chs.harvard.edu/curated-article/gregory-nagy-review-of-writing-homer-a-study-based-on-results-from-modern-fieldwork-by-minna-skafte-jensen/ (4 February 2022).

Nesselrath, Heinz-Günther. 2011. Geschichte der Homerforschung. In Antonios Rengakos and Bernhard Zimmermann (eds.), *Homer-Handbuch: Leben – Werk – Wirkung*, 175–98. Stuttgart: Metzler.

Parry, Milman. 1971. *The Making of Homeric Verse: The Collected Papers of Milman Parry*. Ed. Adam Parry. Oxford: Oxford University Press.

Passa, Enzo. 2016. L'epica. In Albio Cesare Cassio (ed.), *Storia delle lingue letterarie greche*, 139–96. Florence: Le Monnier.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org.

Ready, Jonathan. 2019. *Orality, Textuality, and the Homeric Epics: An Interdisciplinary Study of Oral Texts, Dictated Texts, and Wild Texts*. Oxford: Oxford University Press.

Schmidt, Drew, and Christian Heckendorf. 2021. Fast n-Gram "Tokenization," Version 3.2.0. https://cran.rproject.org/web/packages/ngram/ (16 February 2022).

Statamatos, Efstathios. 2009. A Survey of Modern Authorship Attribution Methods. *Journal of the American Society for Information Science and Technology* 60.538–56.

Stover, Justin A., and Mike Kestemont. 2016. The Authorship of the *Historia Augusta*: Two New Computational Studies. *Bulletin of the Institute of Classical Studies* 59.140–57.

Thomson, Norman D. 1969. A Computer Experiment on the Chronology of Xenophon. *Computers and the Humanities* 4.109–13.

Turner, Frank. 2011. The Homeric Question. In Ian Morris and Barry Powell (eds.), *A New Companion to Homer*, 123–45. Leiden: Brill.

Wachter, Rudolf. 2007. Greek Dialects and Epic Poety: Did Homer Have to Be an Ionian? In Miltiade Hatzopoulos (ed.), *Φωνῆς χαρακτήρ ἐθνικός: Actes du Ve congrès internationale de dialectologie greque*, 317–28. Paris: Boccard.

Watkins, Calvert. 1995. *How to Kill a Dragon: Aspects of Indo-European Poetics*. New York: Oxford University Press.

West, Martin (ed.). 1998. *Homeri Ilias*. Stuttgart: Teubner.

———. 2010. *The Making of the Iliad: Disquisition and Analytical Commentary*. Oxford: Oxford University Press.

———. 2014. *The Making of the Odyssey*. Oxford: Oxford University Press.

Zambarbieri, Mario. 1988–90. *L'Iliade com'è*. Milan: Cisalpino-Goliardica.

———. 2002–4. *L'Odissea com'è*. Milan: LED Edizioni Universitarie.