Exploring intertextuality across the Homeric poems through language models

Maria Konstantinidou[1]*, John Pavlopoulos[2], Elton Barker[3]

[1]Democritus University of Thrace, Greece
[2]Athens University of Economics and Business, Greece
[3]The Open University, U.K.

Abstract

Past research has modelled statistically the language of the Homeric poems, assessing the degree of surprisal for each verse through diverse metrics and resulting to the HoLM resource. In this study we utilise the HoLM resource to explore cross-poem affinity at the verse level, looking at Iliadic verses and passages that are less surprising to the Odyssean model than to the Iliadic one and vice-versa. Using the same tool, we investigate verses that evoke greater surprise when assessed by a local model trained solely on their source book, compared to a global model trained on the entire source poem. Investigating deeper on the distribution of such verses across the Homeric poems we employ supervised learning to further analyse quantitatively cross-poem affinity in selected books.

1 Background

The precise process by which the monumental ancient Greek epics, the Iliad and the Odyssey, came into being remains a point of speculation. That they have survived to the present day may be traced back to the Hellenistic period of ancient Greece (c. 330 BCE), when early scholars focused on issues of textualisation, primarily editing, in order to curate the canonical version of Homer. Less clear is how the poems arrived at this point. Though ever more detailed and in-depth references to them had been occurring over the previous two centuries, there is no documentation relating to the moment of their composition, primarily because they were the product of a vibrant oral society. Rather than being seen as the beginning of a Western tradition, it is more fruitful to think of the Homeric poems as coming at the end of a long tradition of in-performance improvisation, where poets recut the cloth of what they had inherited to weave new stories. The way into thinking about the oral traditionality (Foley, 1991) of these poems is through their language, and,

in particular, the repeated phrases or epithets that have long been regarded as a characteristic feature of them. Phrases such as "swift-footed Achilles" — a line that recurs throughout the Iliad, for example – are not designed to capture a moment in a specific way but rather "trigger a chain of associations" (Graziosi and Haubold, 2005, p. 53) in the minds of audiences, who have grown up with these stories and poems of this kind. The more familiar with other (earlier) uses of such phrases, the more an audience can derive meaning from their present application (Barker and Christensen, 2019). Since "oral poetry works like a language, only more so" (Foley, 2002, p. 127), there is great potential in leveraging language modelling for better understanding how the Homeric poets have been put together. Such work might not be able to resolve the so-called Homeric Question: whether, that is, one person — let's call him Homer — composed both (or one of) the Iliad and Odyssey in the form that have come down to us. Yet, it is the contention of this paper that language modelling can lift the curtain on the mechanics of oral competitive poetics, either by drawing attention to the points of connection between the poems or to other epics (such as those of Hesiod), or, on the contrary, by revealing moments of rupture from the norm. In this way, we hope to set out some ground rules for identifying, and thinking about, the practice by which individual passages generate meaning by playing on audience expectations and their very familiarity with traditional story patterns, themes, and phraseology.

Our starting point is the HoLM resource, developed to assist scholars studying linguistic heterogeneity within the Homeric poems at the level of different structural elements (verses, passages and books) (Pavlopoulos et al., 2024), where related work is also discussed. In this study, we use the cross-score metric to calculate the number of verses per book exhibiting greater linguistic affinity (i.e.,

---

reduced surprise) with the opposite poem than to their original source poem. Such verses, either individual or in clusters, suggest complexities beyond simple interpolations. They hint at potential contamination between the poems or common origins for specific passages. As a means of generating supplementary data complementing existing material, we train three text classifiers to assess the verses of nine books selected from both poems, five from the Iliad and four from the Odyssey.

1.1 The HoLM resource

The HoLM resource uses character level statistical language models to score the Iliad and the Odyssey with a variety of metrics that assess each verse's linguistic unexpectedness to the trained models (Pavlopoulos et al., 2024). The dataset comprises a 'cross-score' computed for each verse, designed to compare the degree of unexpectedness across the two poems. In this work, we also consider relations between books of the same poem. To compare unexpectedness between the individual source book (local model) of a verse and its entire source poem (global model), we use the two Perplexity (PPL) scores provided in HoLM, 'local PPL' and 'global PPL'.

1.2 The formulaic character of Homeric poetry

Homeric poetry, much like other forms of oral literature, fundamentally relies on repetition, both linguistic and thematic. These repetitions serve several crucial purposes, acting as mnemonic devices that aid the poet in structuring the material, and as triggering devices that enable an audience to derive meaning from it — all the more critical for narratives as extensive and all-encompassing as the Iliad and the Odyssey. As has been long recognised, the use of formulas — repetitive epithets, phrases, half-verses, and even entire verses — constitutes a significant feature of Homeric poetry (Parry, 1971). These formulas function as the building blocks of the poetry, ensuring a smooth and continuous poetic flow. For instance, recurring phrases like "rosy-fingered dawn" or "swift-footed Achilles" serve not only to describe characters and scenes vividly but also to fit the metrical requirements of the epic's dactylic hexameter. This technique provides the poet with ready-made segments of verse that can be adapted to various narrative contexts, thus facilitating the composition of long, complex stories in real-time performance. At the same time, these repetitive elements also enhance an audience's understanding of the thematic coherence of the story-in-performance, as well as appreciation for the story it has to tell. They help create a sense of continuity and unity, by enabling an audience to anchor different parts of the narrative and grasp key ideas, particularly when heard in and against the stories that have been sung before. The extensive use of these formulas results in a high degree of repetition within the Homeric poems, both intra- and inter-poem. Identical or near-identical verses, often repeated multiple times, are scattered across the poems. Of the 15,683 verses in our version of the Iliad, 2,019 are duplicate (approximately 13%); that is, they are repeated one or more times. In the Odyssey 1,884 out of the 12,107 verses (approximately 15.5%) are duplicates.[1] There are many more near-duplicate verses, typically hemistichs (half-verses), and a lot of shorter formulas consisting of two or three words. So well-established is the idea of formularity in oral poetry, that scholars need to argue in favour of the uniqueness and the non-formulaic nature of Homeric diction, estimating that at least one third of it is *not* affected by formulas (Finkelberg, 2020). Dealing with formulas in a computational study presents several complex challenges that necessitate a comprehensive, separate investigation. Key issues include defining what constitutes a formula (e.g., whether two words should be considered as one) and understanding how these formulas interact with the metrical structure of the verses (Bozzone, 2022). Additionally, the overall language modelling of the text must be considered in connection with repeated expressions: recent studies have established that the density of formulas in Homeric texts is not exceptional and that contemporary speech exhibits a comparable degree of formularity (Erman and Warren, 2000). In our study, duplicate and near-duplicate verses are not excluded for training. Additionally, they score, as expected, a lower PPL both with the source poem model and the other poem model, since the same formulas can be found in both works. Table 1 presents examples of duplicate verses repeated within the Iliad. Table 2 shows recurrent verses in both poems.

---

[1]HoLM uses the (Allen, 1931) edition for the Iliad and for the Odyssey the (von der Mühll, 1962) one.

Verse no	Text
16,711	μῆνιν ἀλευάμενος ἑκατηβόλου ἀπόλλωνος
1,297	ἄλλο δέ τοι ἐρέω σὺ δ᾽ ἐνὶ φρεσὶ βάλλεο σῇσι
5,444	μῆνιν ἀλευάμενος ἑκατηβόλου ἀπόλλωνος
20,19	τὸν δ᾽ ἀπαμειβόμενος προσέφη νεφεληγερέτα ζεύς
22,182	τὴν δ᾽ ἀπαμειβόμενος προσέφη νεφεληγερέτα ζεύς
24,64	τὴν δ᾽ ἀπαμειβόμενος προσέφη νεφεληγερέτα ζεύς
8,477	τὴν δ᾽ ἀπαμειβόμενος προσέφη νεφεληγερέτα ζεύς
5,764	τὴν δ᾽ ἀπαμειβόμενος προσέφη νεφεληγερέτα ζεύς
23,93	τὸν δ᾽ ἀπαμειβόμενος προσέφη πόδας ὠκὺς ἀχιλλεύς
14,311	τὴν δ᾽ ἀπαμειβόμενος προσέφη νεφεληγερέτα ζεύς

Table 1: The verses with the ten lowest PPL scores in the Iliad are all duplicates

2 Motivation and method

Our objective is to broaden the scope of research on the phenomenon of unexpectedness by moving beyond the mere identification of surprising verses or passages within a model trained on the source text. Among the verses identified as unexpected in HoLM, we perform a quantitative analysis specifically on verses/passages that appear to be linguistically more surprising to their immediate surroundings than to other, more remote parts of the Homeric poems. In short, we focus on three levels of surroundings that provide increasingly broader contexts for assessing the linguistic surprise of verses or passages:

- **Immediate surroundings**: This refers to the immediate context of a verse or passage within its own book; specifically, the verses directly preceding and following the target verse. We investigate this level by seeking consecutive or near-consecutive outlier verses.

- **Individual book > Source poem**: This level expands the scope beyond the immediate surroundings to include the entirety of the book containing the verse or passage, compared to the source poem from which it originates. It assesses the verse's surprise factor within the context of its book in relation to the entirety of its source poem.

- **Source poem > Other poem**: This evaluates the level of surprise of a verse or passage within its source poem, juxtaposed with the surprise calculated using a model trained on the entirety of the other Homeric poem (e.g., the Iliad compared to the Odyssey, or vice versa). Specifically, it examines how the unexpectedness of the verse within its own context contrasts with its unexpectedness when

assessed against the entirety of the alternative Homeric work.

This method lays the groundwork for investigating internal transposition of text within each poem, as well as Odyssean elements in the Iliad and vice versa. Further systematic study of such passages may help not only unveil patterns of interpolation itself, but also to shed light on what is consciously or instinctively perceived as 'Iliadic' or 'Odyssean', thereby ultimately unlocking insights into the agonistic, compositional basis of either poem.

3 Assessing proximity with the other poem

3.1 Positive cross-score

To identify verses and passages that may be linguistically more distant to their source poem than to the other one, we use the cross score. For a given verse, this is the difference between the PPL for that verse computed with the model trained on the source poem and the equivalent PPL computed with the model trained on the other poem. A positive cross value (PCV) for a verse means that the verse is more surprising to the source poem model than it is to the model trained on the other poem. We used PCVs to identify possible passages that may exhibit greater source poem surprise (clusters of more than two PCVs). Since a PCV is a rarity and to ensure that individual verses are not isolated from their surroundings, we also took into account the top 10% of verses with the highest negative scores (NCV). As mentioned above, few verses have a positive cross score: in the Iliad there are 511 PCVs in total and 375 if we remove the duplicate verses among them. In the Odyssey, we identified 272 PCVs (235 after duplicate verse elimination). Thus, the Iliad contains far more such verses, even after allowing for its greater length compared to the Odyssey (Fig. 1).

The greatest concentration is found in Books 24 (Fig. 3), 9 and 1 of the Iliad. The lowest concentration is found in Odyssey 7. Of the 783 PCVs, approximately one in five (173) are duplicates. Far less duplicate verses have a positive cross-score in the Odyssey (0.3%) than in the Iliad (0.87%). This could be due to the fact that in the Odyssey, unique common verses that appear across poems are fewer in number but are repeated more frequently compared to those in the Iliad(Fig. 2 and Table 2).

The low number of PCVs does not allow a reliable statistical analysis at the book level and a

Verse	Total count	Iliad count	Odyssey count
καί μιν φωνήσας ἔπεα πτερόεντα προσηύδα	30	15	15
τὸν δ᾽ ἀπαμειβόμενος προσέφη πολύμητις ὀδυσσεύς	30	5	25
ἦμος δ᾽ ἠριγένεια φάνη ῥοδοδάκτυλος ἠώς	22	2	20
αὐτὰρ ἐπεὶ πόσιος καὶ ἐδητύος ἐξ ἔρον ἔντο	21	7	14
ἀλλ᾽ ἄγε μοι τόδε εἰπὲ καὶ ἀτρεκέως κατάλεξον	17	4	13
διογενὲς λαερτιάδη πολυμήχαν᾽ ὀδυσσεῦ	17	7	10
ὀχθήσας δ᾽ ἄρα εἶπε πρὸς ὃν μεγαλήτορα θυμόν	11	7	4
τὸν δ᾽ ἠμείβετ᾽ ἔπειτα γερήνιος ἱππότα νέστωρ	11	8	3
ἀτρεΐδη κύδιστε ἄναξ ἀνδρῶν ἀγάμεμνον	10	8	2
τὴν δ᾽ ἀπαμειβόμενος προσέφη νεφεληγερέτα ζεύς	10	7	3

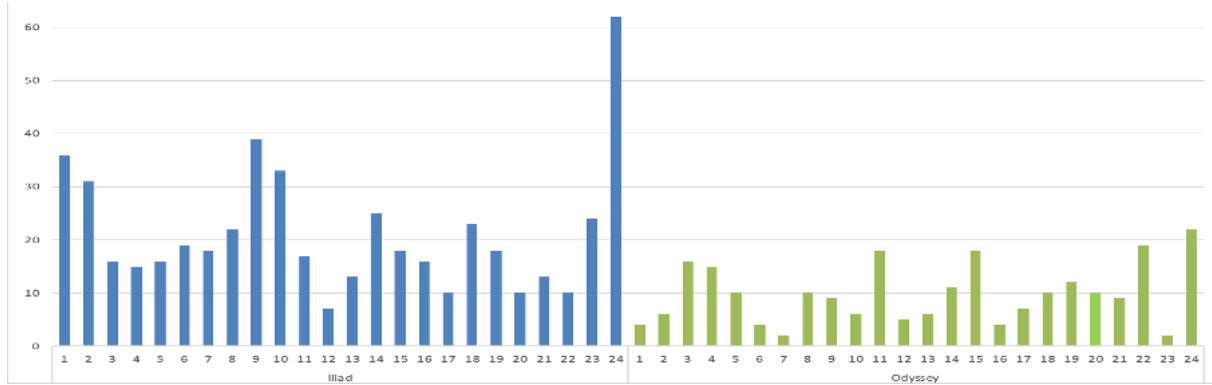Table 2: The common verses in the two poems with the most occurances.

Figure 1: Number of verses with positive cross scores per book in the Iliad and the Odyssey
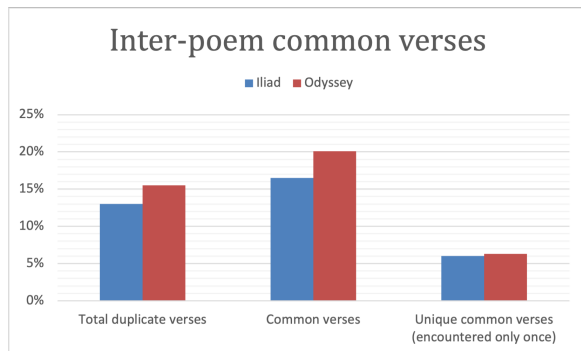

Figure 2: Verses in each poem also found in the other (percentage of total poem verses)

more decisive tool should be used for this purpose. Nevertheless, the cross-score metric can be useful to identify potential passages of interest.

4 Intra-poem unexpectedness: global versus local PPL

Local PPL is computed by training a statistical language model on the whole of the source book, excluding only the textual part that is being scored. As global PPL we consider the PPL score computed by a model trained on the source poem. In the Iliad, of the 1568 verses within the top 10 percentile of global PPL score, only 510 also rank among the top 10 percentile of Local PPL score. This means that 66% of the top surprising verses to the Global Iliad model are not surprising to their local Book model (and vice versa). If we examine the top 20 percentile, then 1536 out of the 3233 (48%) are equally surprising both globally and locally. In the Odyssey, it is a similar ratio, with 1123 verses universally surprising out of the 2421 globally surprising ones. The books with the greatest number of high local PPL verses are: Iliad 12, 22 and 24, and 1, 8 and 11 of the Odyssey (Fig. 4).

5 Zooming in

As is appropriate for works with such a lengthy and involved compositional history, the macroscopic book-level analysis ultimately aims at identifying distinctive narrative segments with higher concentration of PCVs, indicating a closer affinity with the other poem than their source poem. Using the books that stood out in the statistical analysis of PCVs, we focus on Iliad 1, 2, 9, 10 and 24. Book 10 is probably the most discussed book in terms of its authenticity; it is still commonly regarded as interpolated (or at least extended parts of it)
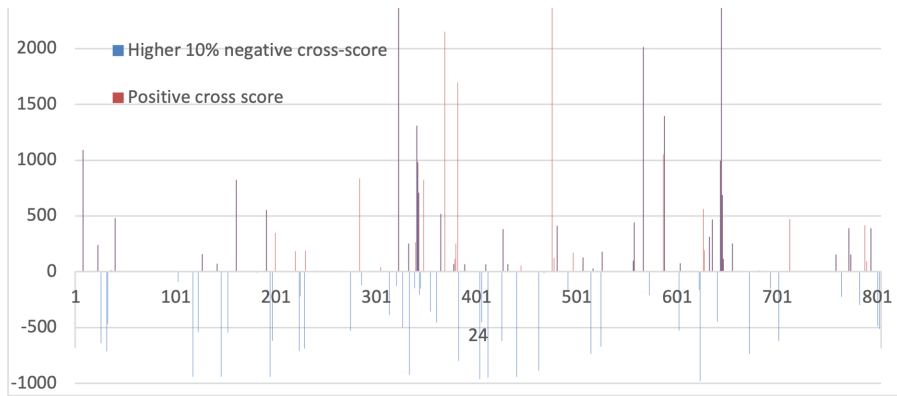
(a) HoLM SLM
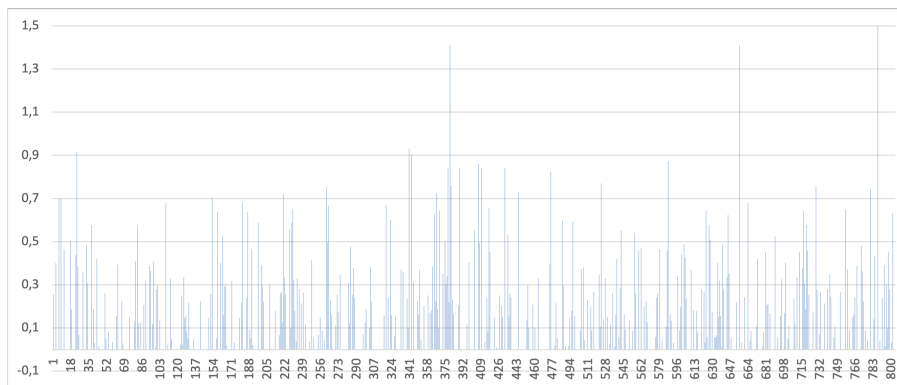
(b) RNNLM

Figure 3: Iliad book 24: (a) PCV and near positive cross-score verses computed with the HoLM SLM models and (b) PCV computed with the RNN model

(Danek, 2012). Book 2, which heralds 'the great gathering of armies', has also been discussed extensively in the literature (see for instance (Karanika, 2020)), again due to its atypical content, since it includes extensive lists, not least of which is the famous catalogue of ships. However, it is Book 24 that stands out from our book-level analysis of the HoLM resource: it exhibits the highest number of PCVs (Fig. 1) as well as the highest rate of locally surprising verses (Fig. 4). From the Odyssey, we selected books 11, 15, 22 and 24, the ones with the highest number of PCVs. Book 11 demonstrates in addition the highest number of high local PPL in the poem. Together with Book 24, they also present a more coherent picture of surprising passages with high concentration of PCVs (groups of verses clustered in close proximity).

5.1 Machine Learning for verse classification

To capture greater depth and range of language dependencies, we trained supervised learning algorithms to classify verses between the two poems (i.e., source v. other poem). This results in each

unseen verse being marked as 'source-surprising' or not. We opted for three traditional machine learning algorithms for this experiment,[2] K-nearest neighbours (KNN), logistic regression (LR), and random forests (RF). All the algorithms operated on top of term-frequency inverse-document-frequency (TFIDF; documents are verses in our case) features, using character n-grams (i.e., sequences of two to five characters), maximum document frequency of 0.5 and minimum document frequency of 5 (i.e., we ignore n-grams in more than half and less than 5 verses).

The classifiers were trained on the whole of both poems, excluding books 1, 2, 9, 10 and 24 of the Iliad and books 11, 15, 22, and 24 of the Odyssey. We kept 20% of the verses, randomly selected across both poems, for evaluation purposes. This left us with 12,103 verses from Iliad and 9,861 verses from Odyssey for training. As is shown in Table 3, LR was the best in classifying the poem a verse belongs in, followed by RF and KNN. All

---

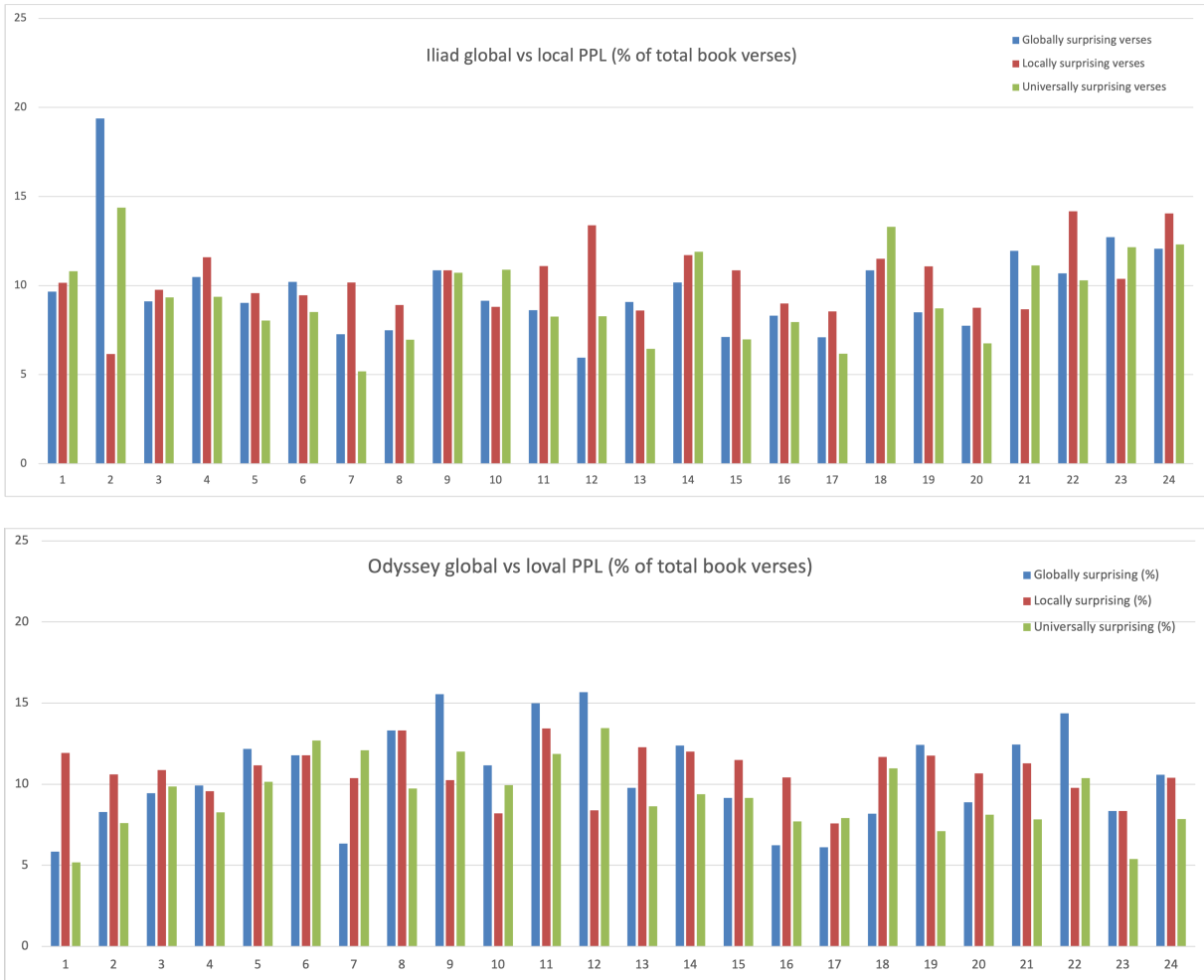[2]We used the scikit-learn library for the implementations.

264

Figure 4: Global vs local PPL per book in the two poems

	ODYSSEY			ILIAD		
	P	R	F1	P	R	F1
LR	**0.75**	**0.68**	**0.72**	**0.76**	**0.82**	**0.79**
KNN	0.70	0.66	0.68	0.73	0.77	0.75
RF	0.73	0.65	0.69	0.74	0.81	0.77
RAND	0.45	0.50	0.47	0.55	0.50	0.52

Table 3: Precision, Recall and F1 per algorithm per poem. In bold the best per column.

three algorithms, however performed considerably better than a random baseline (RAND), classifying the verse randomly.

We also used these three classifiers to yield predictions per verse from the left out books. The distribution of source-surprising verses across the 9 books is shown in Fig. 6, but we observe that there is a positive correlation between the classifications of the three models (Fig. 5).

5.2 Classification vs PCV

Comparing the attribution of verses to the other poem in the four books, the classifiers largely support the SLM PCVs. In the Iliad, 79% of the PCVs are also flagged by the classifiers as Odyssean; in the Odyssey, 71% of PCVs are flagged as Iliadic. At the same time, the three classifiers substantially increase the quantity of source-surprising verses, revealing a clearer view. Among then, they mark as source-surprising another 1,622 verses in the Iliad and 1,192 in the Odyssey.

6 Discussion

Compared to the PCVs, a large number of groups of source-surprising verses (up to four excluding duplicates) can be readily observable in all of the individual books evaluated by the ML classifiers. Further merging groups located closely together requires closer examination and carefully selected criteria. It is, however, feasible to discern poten-
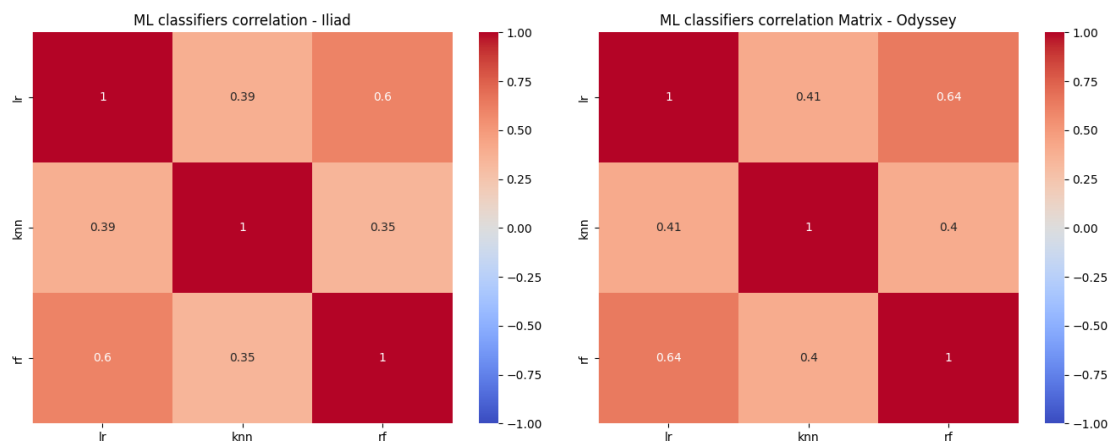
Figure 5: Pearson correlation between the machine learning classifiers on verses from held-out books of Iliad (the heatmap on the left) and Odyssey (on the right).
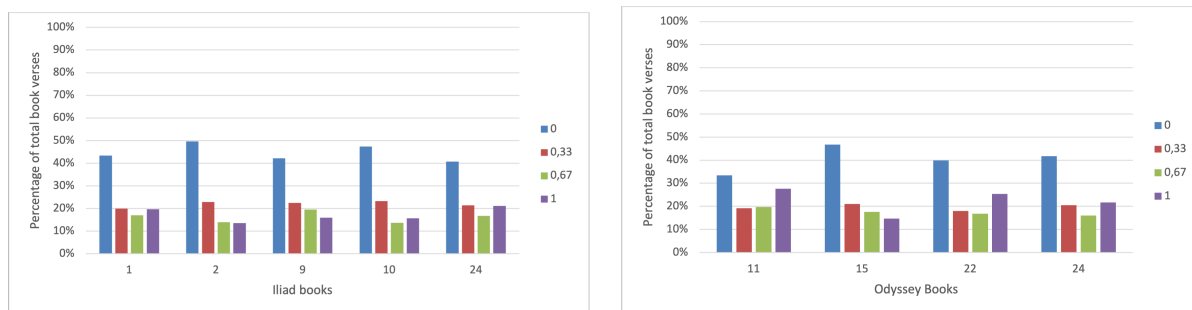

Figure 6: ML rate agreement in identifying source-surprising verses per book (0: no model classifies the verse as source-surprising; 1: all three models classify the verse as source-surprising). Number of verses shown as percentage of the book's total verses.

tial patterns among some of the passages picked out. These include lists (such as the catalogue of ships in Iliad 2 and the list of women in Odyssey 11), but also a number of similes, as well as narrations referring to the past. From a literary perspective, the books that emerge as related to the 'other' poem reveal close correspondences. These moments of contact could simply be down to the protagonist of either poem and their prominence in the other poem, namely Odysseus in Iliad 2, 9 and 10, and Achilles in Odyssey 11 (noting that the wrath of Achilles is the headline of the Iliad, and the return home of Odysseus the subject of the Odyssey). But, as well as being insufficient to explain all the cases (particularly in the Odyssey), the presence of the protagonist arguably better indicates heightened moments of thematic cross-over between the epics. As well as being prominent in these books of the Iliad, Odysseus also acts in an 'Odyssean' manner, most notably in Iliad 10, the book which some critics still doubt or consider as

a late 'add on', precisely because of its seemingly unIliadic story of night adventure, ambush and deceit (led, of course, by Odysseus) (Barker, 2009). In Iliad 9 Odysseus is prominent as the leader of the embassy to Achilles, where his rhetorical skills are on display (and seen through by Achilles). In Iliad 2, Odysseus again takes control of the narrative, after Agamemnon's disastrous 'testing' of the troops: it is Odysseus to whom Athena goes (as she so often does in the Odyssey) and who notably holds back the Achaeans as they rush to the ships to go home, an event that, the Homeric narrator remarks, would have been 'beyond fate' (Barker, 2009).. The rivalry between these alternative epic traditions is taken up in Odyssey 11, where an ambushed Achilles is left behind in the Underworld bemoaning his early death and anxious for news of his son, even as Odysseus continues on his journey home to reunite with his l(Edwards, 1985). Odyssey 22 is the moment when Odysseus's banqueting halls become an Iliadic battleground,

as Odysseus takes on and slaughters all the suitors who have been eating his son out of house and home. Odyssey 24 opens with another scene of (un)Iliadic heroes in the underworld — Achilles and Agamemnon praising each other no less — and culminates in another battle, when Odysseus, accompanied by both his father and son, takes on and kills the families of the suitors (Barker, 2009; Barker and Christensen, 2019). It is also striking that the beginning and ending of the Iliad is marked out as resonating strongly with the other tradition, as if self-consciously aware of its place in the tradition. Analysing individual passages is simultaneously more straightforward and open to speculation, especially in defining their boundaries. Nevertheless, a number clearly stand out when considering both PCVs and the ML models classifications. Such an example of source-surprising verses is in Iliad book 10 (263-279) where the arming of Odysseus is described in a distinct section of the book. In the Odyssey, the models mark as source-surprising the catalogue of women in book 11 (specifically verses 255-272 and 299-330). This is also a section mentioned in literature as a possible interpolation and further discussed in the same context in (Pavlopoulos and Konstantinidou, 2023).

7 Conclusions

Our two methods for assessing the level of inter-poem surprise largely converge in identifying specific books and passages as notably surprising within their respective poems. An initial expert analysis of the flagged passages reveals potential patterns recognized by the models; notably, Books 1, 9, 10, and 24 of the Iliad and Books 11, 15, 22, and 24 of the Odyssey contain the highest concentration of such verses. Within these books, shorter passages appear to contribute more significantly to these findings.

Further analysis of shorter verse clusters with positive or almost positive cross score seem promising in revealing both linguistic and thematic criteria associated with either poem. It may also reveal lexical features that weigh more in each poem's language modelling.

Future research may focus on these aspects, including catalogs, direct speech, gender-related topics and discourse, as well as proper names and content words. Additionally, it should aim to examine the Homeric poems within their closer historical context and model them alongside other ancient authors and genres, such as Hesiod and lyric poetry.

References

T.W. Allen. 1931. *Homeri Ilias*, volume 2 and 3. Clarendon Press, Oxford.

Elton T. E. Barker. 2009. *Entering the Agon: Dissent and Authority in Homer, Historiography, and Tragedy*. Oxford University Press.

Elton T. E. Barker and Joel P. Christensen. 2019. *Homer's Thebes: Epic Rivalries and the Appropriation of Mythical Pasts*. Center for Hellenic Studies.

Chiara Bozzone. 2022. Technologies of orality: Formularity, meter, and kunstsprache in homer. In *Rethinking Orality II: The Mechanisms of the Oral Communication System in the Case of the Archaic Epos*, pages 51–82, Berlin, Boston. De Gruyter.

Georg Danek. 2012. The Doloneia revisited. In Øvind Andersen and Dag T. T. Haug, editors, *Relative Chronology in Early Greek Epic Poetry*, pages 106–121. Cambridge University Press, Cambridge.

Anthony T. Edwards. 1985. *Achilles in the Odyssey : ideologies of heroism in the Homeric epic*. Verlag Anton Hain.

Britt Erman and Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text & Talk*, 20(1):29–62.

Margalit Finkelberg. 2020. 6. oral theory and the limits of formulaic diction. In *Homer and Early Greek Epic, Collected Essays*, pages 53–65, Berlin, Boston. De Gruyter.

John Miles Foley. 1991. *Immanent art : from structure to meaning in traditional oral epic*. Bloomington (Ind.) : Indiana University Press.

John Miles Foley. 2002. *How to Read an Oral Poem*. University of Illinois Press.

Barbara Graziosi and Johannes Haubold. 2005. *Homer: The Resonance of Epic*. Duckworth. EPrint Processing Status: DRO Team to ask author to provide full text (pre-publication version).

Andromache Karanika. 2020. Catalogues. In *The Cambridge Guide to Homer*, page 108–109. Cambridge University Press.

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

John Pavlopoulos and Maria Konstantinidou. 2023. Computational authorship analysis of the Homeric poems. *International Journal of Digital Humanities*, 5:45–64.

John Pavlopoulos, Ryan Sandell, Maria Konstantinidou, and Chiara Bozzone. 2024. HoLM: Analysing Linguistic Unexpectedness in Homeric Poetry. In *LREC-COLING*.

P. von der Mühll. 1962. *Homeri Odyssea*. Helbing & Lichtenhahn, Basel.