



ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ  
ΣΤΙΣ ΨΗΦΙΑΚΕΣ ΜΕΘΟΔΟΥΣ ΓΙΑ ΤΙΣ ΑΝΘΡΩΠΙΣΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ

**Γλωσσική μοντελοποίηση για την απόδοση  
συγγραφικής πατρότητας σε Ομηρικά κείμενα**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΦΑΣΟΗ ΜΑΡΙΑ**

ΑΘΗΝΑ, ΝΟΕΜΒΡΙΟΣ 2020



ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ  
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ  
ΣΤΙΣ ΨΗΦΙΑΚΕΣ ΜΕΘΟΔΟΥΣ ΓΙΑ ΤΙΣ ΑΝΘΡΩΠΙΣΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ

**Γλωσσική μοντελοποίηση για την απόδοση  
συγγραφικής πατρότητας σε Ομηρικά κείμενα**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΦΑΣΟΗ ΜΑΡΙΑ**

Επίβλεψη: Ιωάννης Παυλόπουλος, Λέκτορας (fixed-term) στο Πανεπιστήμιο της Στοκχόλμης, Σουηδία – Διδάσκων του ΠΜΣ Ψηφιακές μέθοδοι για τις Ανθρωπιστικές Επιστήμες του Οικονομικού Πανεπιστημίου Αθηνών

Συν-επίβλεψη: Μαρία Κωνσταντινίδου, Επίκουρη καθηγήτρια στο Τμήμα Ελληνικής Φιλολογίας του Δημοκριτείου Πανεπιστημίου Θράκης

Κριτές: Ιωάννης Παυλόπουλος, Λέκτορας (fixed-term) στο Πανεπιστήμιο της Στοκχόλμης, Σουηδία – Διδάσκων του ΠΜΣ Ψηφιακές μέθοδοι για τις Ανθρωπιστικές Επιστήμες του Οικονομικού Πανεπιστημίου Αθηνών

Μαρία Κωνσταντινίδου, Επίκουρη καθηγήτρια στο Τμήμα Ελληνικής Φιλολογίας του Δημοκριτείου Πανεπιστημίου Θράκης

Ίων Ανδρουτσόπουλος, Αναπληρωτής καθηγητής στο Τμήμα Πληροφορικής του Οικονομικού Πανεπιστημίου Αθηνών

ΑΘΗΝΑ, ΝΟΕΜΒΡΙΟΣ 2020



## Περίληψη

Στην παρούσα διπλωματική εργασία εφαρμόστηκαν γλωσσικά μοντέλα, ώστε να μελετηθεί η απόδοση συγγραφικής πατρότητας ομηρικών κειμένων, που κατά την κρατούσα άποψη αποδίδονται στον ποιητή Όμηρο. Για την επίτευξη του σκοπού αυτού εφαρμόστηκαν στατιστικά και νευρωνικά γλωσσικά μοντέλα, τα οποία εκπαιδεύτηκαν στα ομηρικά έπη (Ιλιάδα και Οδύσσεια) και βαθμολογήθηκαν τόσο στα ίδια τα έπη όσο και στους καλούμενους Ομηρικούς ύμνους. Η βαθμολόγηση πραγματοποιήθηκε αφενός μεν μετρώντας με αυτόματο τρόπο τη γλωσσική συγγένεια ενός δείγματος κειμένου με το σύνολο του εκάστοτε γλωσσικού μοντέλου, αφετέρου δε ταξινομώντας αποσπάσματα ομηρικών κειμένων.

Με αυτόν τον τρόπο, εντοπίστηκαν αποσπάσματα ομηρικών κειμένων να εμφανίζουν περισσότερη συγγένεια από άλλα αποσπάσματα σε γλωσσικό επίπεδο με το σύνολο των δεδομένων εκπαίδευσης. Πιο συγκεκριμένα, η ραψωδία *Ο* της *Ιλιάδας* με το σύνολο του έπους της *Ιλιάδας* και η ραψωδία *β* της *Οδύσσειας* με το συνολικό έπος της *Οδύσσειας*. Έπειτα, παρατηρήσαμε ότι η ραψωδία *δ* της *Οδύσσειας* εμφανίζει περισσότερη γλωσσική ομοιότητα με το έπος της *Ιλιάδας* και αντίστοιχα η ραψωδία *Τ* της *Ιλιάδας* με το έπος της *Οδύσσειας*. Ακολούθως, μελετήσαμε τη γλωσσική συγγένεια των Ομηρικών ύμνων *Εἰς Ἀπόλλωνα*, *Εἰς Ἀφροδίτην*, *Εἰς Δημήτραν* και *Εἰς Ἑρμῆν* με τα ομηρικά έπη και καταλήξαμε στο συμπέρασμα ότι ο ύμνος *Εἰς Ἀφροδίτην* εμφανίζει τη μεγαλύτερη γλωσσική συγγένεια με τα ομηρικά έπη. Τέλος, τα εν λόγω γλωσσικά μοντέλα αντιπαραβλήθηκαν με κρίσεις ανθρώπων-επισημειωτών, ώστε να ερευνηθεί η ικανότητά τους να κατηγοριοποιούν ορθά δοθέντα αποσπάσματα αρχαίων κειμένων στο αντίστοιχο έργο υπαγωγής τους. Συμπεράναμε ότι τα νευρωνικά γλωσσικά μοντέλα μακράς – βραχείας μνήμης δύνανται να ταξινομούν με μεγαλύτερη επιτυχία ομηρικά κείμενα στο αντίστοιχο έργο υπαγωγής τους απ' ό,τι τα στατιστικά γλωσσικά μοντέλα και η ανθρώπινη κρίση.

**Λέξεις κλειδιά:** Ομηρικά έπη, Ομηρικοί ύμνοι, Επεξεργασία Φυσικής Γλώσσας, Απόδοση συγγραφικής πατρότητας, στατιστικά γλωσσικά μοντέλα, νευρωνικά γλωσσικά μοντέλα μακράς - βραχείας μνήμης



# Abstract

## Language Modeling for Authorship attribution in Homeric texts

In the present thesis, language models were applied in order to study the authorship attribution of Homeric texts, which according to the prevailing view are attributed to the poet Homer. To achieve this goal, statistical and neural language models were applied, which were trained on the Homeric epics (Iliad and Odyssey) and were rated both in the epics themselves and in the so-called Homeric hymns. The grading was performed on the one hand by automatically measuring the linguistic affinity of a sample of text with the whole linguistic model, on the other hand by classifying excerpts from Homeric texts.

In this way, excerpts from ancient texts were found to show more affinity than other excerpts at the linguistic level with the set of training data. More specifically, the rhapsody “*O*” of Iliad with the whole epic of the Iliad and the rhapsody “*β*” of Odyssey with the total epic of Odyssey. Next, we observed that the rhapsody “*δ*” of Odyssey shows more linguistic resemblance to the epic of Iliad and respectively the rhapsody “*T*” of Iliad to the epic Odyssey. Then, we studied the linguistic affinity of the Homeric hymns “*To Apollo*”, “*To Aphrodite*”, “*To Demeter*” and “*To Hermes*” with the Homeric epics and concluded that the hymn “*To Aphrodite*” shows the greatest linguistic affinity with the Homeric epics. Finally, these linguistic models were contrasted with human-annotators interpretation in order to investigate their ability to classify correctly given excerpts of ancient texts into their respective subordinate work. We conclude that neural language models long short-term memory can classify Homeric texts more successfully in their respective subordinate work than statistical language models and human interpretation.

**Keywords:** Homeric epics, Homeric hymns, Natural Language Processing, Authorship attribution, Statistical Language Models (SLM), Neural language models Long Short-Term Memory (LSTM)

*Στην οικογένειά μου*





## Ευχαριστίες

Με την περάτωση της παρούσας διπλωματικής εργασίας, θα ήθελα πρωτίστως να ευχαριστήσω θερμά τον επιβλέποντά μου, κύριο Ιωάννη Παυλόπουλο, για την καθοδήγησή του, την υπομονή που έδειξε καθ' όλη τη διάρκεια της διπλωματικής εργασίας και την ευκαιρία που μου έδωσε να συνδυάσω τις φιλολογικές με τις υπολογιστικές γνώσεις στο έργο της *απόδοσης συγγραφικής πατρότητας* στον επικό ποιητή Όμηρο. Έπειτα, θα ήθελα να ευχαριστήσω θερμά τη συν-επιβλέπουσά μου, κυρία Μαρία Κωνσταντινίδου, για τη διαρκή συμβολή της σε όλη την πορεία της διπλωματικής εργασίας και για την ευχερή πρόσβαση στο σύνολο των δεδομένων. Ακόμη, θα ήθελα να ευχαριστήσω θερμά τον κύριο Ίωνα Ανδρουτσόπουλο για την αξιολόγηση της διπλωματικής εργασίας. Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για όλη την υποστήριξή τους.



## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

<b>1.Εισαγωγή.....</b>	<b>6</b>
<b>2.Βιβλιογραφική επισκόπηση .....</b>	<b>8</b>
<b>3.Δεδομένα.....</b>	<b>12</b>
3.1 Περιγραφή δεδομένων .....	12
3.2 Προ-επεξεργασία δεδομένων (Data preprocessing) .....	13
<b>4.Μέθοδοι γλωσσικής μοντελοποίησης και μετρήσεις αξιολόγησης .....</b>	<b>18</b>
4.1 Στατιστικό Γλωσσικό Μοντέλο (Statistical Language Model – SLM) .....	19
4.2 Νευρωνικό Γλωσσικό Μοντέλο Μακράς – Βραχείας Μνήμης (Long Short-Term Memory – LSTM) .....	21
4.3 Ανθρώπινη αξιολόγηση .....	22
4.4 Μετρήσεις αξιολόγησης .....	23
4.4.1 Περιπλοκότητα (Perplexity) .....	24
4.4.2 F-measure (ή F1-score) .....	25
<b>5.Πειραματικά αποτελέσματα.....</b>	<b>27</b>
5.1.1 Απόδοση συγγραφικής πατρότητας μέσω Στατιστικών Γλωσσικών Μοντέλων (SLM) .....	27
5.1.2 Ταξινόμηση ομηρικών κειμένων μέσω Στατιστικών Γλωσσικών Μοντέλων (SLM) και Νευρωνικών Γλωσσικών Μοντέλων Μακράς - Βραχείας Μνήμης (LSTM).....	39
<b>6.Συζήτηση και μελλοντικές κατευθύνσεις .....</b>	<b>44</b>
<b>Βιβλιογραφία .....</b>	<b>48</b>
<b>Παράρτημα.....</b>	<b>51</b>

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1. Σημεία στίξης και εκδοτικά σύμβολα που αφαιρέθηκαν από τα κωδικοποιημένα δεδομένα.....	14
Πίνακας 2. Χαρακτήρες Ιλιάδας, Οδύσσειας και Ομηρικών ύμνων σε lowercase.....	17
Πίνακας 3. Κατάταξη πειραμάτων που εφαρμόστηκαν στην παρούσα μελέτη.....	18
Πίνακας 4. Confusion matrix για ένα πρόβλημα ταξινόμησης δύο κλάσεων.....	25
Πίνακας 5. Διαστήματα εμπιστοσύνης και μέσοι όροι των ραψωδιών της Ιλιάδας με τα IliadSLM's.....	30
Πίνακας 6. Διαστήματα εμπιστοσύνης και μέσοι όροι των ραψωδιών της Οδύσσειας με τα OdysseySLM's.....	32
Πίνακας 7. Διαστήματα εμπιστοσύνης και μέσοι όροι των ραψωδιών της Οδύσσειας με το IliadSLM.....	34
Πίνακας 8. Διαστήματα εμπιστοσύνης και μέσοι όροι των ραψωδιών της Ιλιάδας με το OdysseySLM.....	35
Πίνακας 9. Μήκος σε χαρακτήρες στους Ομηρικών ύμνων.....	35
Πίνακας 10. Διαστήματα εμπιστοσύνης και μέσοι όροι των Ομηρικών ύμνων με το IliadSLM.....	36
Πίνακας 11. Διαστήματα εμπιστοσύνης και μέσοι όροι των Ομηρικών ύμνων με το OdysseySLM.....	37
Πίνακας 12. F1-score για τα αποσπάσματα της Ιλιάδας και της Οδύσσειας του ερωτηματολογίου.....	41

## ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1. Τυπολογία υφομετρικών χαρακτηριστικών.....	9
Σχήμα 2. Αρχιτεκτονική LSTM.....	22
Σχήμα 3. Μέθοδος Bootstrapping για τα βαθμολογημένα δεδομένα του πειράματος «Απόδοση συγγραφικής πατρότητας μέσω Στατιστικών Γλωσσικών Μοντέλων (SLM)».....	29
Σχήμα 4. Χιαστί βαθμολόγηση στην Ιλιάδα και την Οδύσσεια.....	33

## ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διάγραμμα 1. Μήκος σε χαρακτήρες στις ραψωδίες της Ιλιάδας.....	15
Διάγραμμα 2. Μήκος σε μοναδικούς χαρακτήρες στις ραψωδίες της Ιλιάδας.....	15
Διάγραμμα 3. Μήκος σε λέξεις στις ραψωδίες της Ιλιάδας.....	15
Διάγραμμα 4. Μήκος σε μοναδικές λέξεις στις ραψωδίες της Ιλιάδας.....	15
Διάγραμμα 5. Μήκος σε χαρακτήρες στις ραψωδίες της Οδύσσειας.....	15
Διάγραμμα 6. Μήκος σε μοναδικούς χαρακτήρες στις ραψωδίες της Οδύσσειας.....	15
Διάγραμμα 7. Μήκος σε λέξεις στις ραψωδίες της Οδύσσειας.....	16
Διάγραμμα 8. Μήκος σε μοναδικές λέξεις στις ραψωδίες της Οδύσσειας.....	16
Διάγραμμα 9. Μήκος σε χαρακτήρες στους Ομηρικούς ύμνους.....	16
Διάγραμμα 10. Μήκος σε μοναδικούς χαρακτήρες στους Ομηρικούς ύμνους.....	16
Διάγραμμα 11. Μήκος σε λέξεις στους Ομηρικούς ύμνους.....	16
Διάγραμμα 12. Μήκος σε μοναδικές λέξεις στους Ομηρικούς ύμνους.....	16
Διάγραμμα 13. IliadSLM's ανά ραψωδία Ιλιάδας με μοντέλα εκπαιδευμένα στις υπολειπόμενες 23 ραψωδίες της Ιλιάδας – κυκλική εκπαίδευση.....	30
Διάγραμμα 14. OdysseySLM's ανά ραψωδία Οδύσσειας με μοντέλα εκπαιδευμένα στις υπολειπόμενες 23 ραψωδίες της Οδύσσειας – κυκλική εκπαίδευση.....	31
Διάγραμμα 15. IliadSLM ανά ραψωδία Οδύσσειας με μοντέλο εκπαιδευμένο σε όλες τις ραψωδίες της Ιλιάδας (9.600 χαρακτήρες από κάθε ραψωδία).....	33
Διάγραμμα 16. OdysseySLM ανά ραψωδία Ιλιάδας με μοντέλο εκπαιδευμένο σε όλες τις ραψωδίες της Οδύσσειας (9.600 χαρακτήρες από κάθε ραψωδία).....	34
Διάγραμμα 17. IliadSLM ανά Ομηρικό ύμνο με μοντέλο εκπαιδευμένο σε όλες τις ραψωδίες της Ιλιάδας (9.600 χαρακτήρες από κάθε ραψωδία).....	36
Διάγραμμα 18. OdysseySLM ανά Ομηρικό ύμνο με μοντέλο εκπαιδευμένο σε όλες τις ραψωδίες της Οδύσσειας (9.600 χαρακτήρες από κάθε ραψωδία).....	37
Διάγραμμα 19. Θέση του Ομηρικού ύμνου Είς Άφροδίτην συγκριτικά με τα Ομηρικά έπη.....	38
Διάγραμμα 20. F1-score για τα αποσπάσματα της Ιλιάδας και Οδύσσειας του ερωτηματολογίου.....	41
Διάγραμμα 21. Συνολική επίδοση F1-score επισημειωτών, SLM και LSTM για τα Ιλιαδικά και Οδυσειακά αποσπάσματα του ερωτηματολογίου.....	41
Διάγραμμα 22. Συνολική επίδοση F1-score μέσου επισημειωτή, καλύτερου επισημειωτή καθώς και των γλωσσικών συστημάτων, SLM και LSTM, για τα Ιλιαδικά και Οδυσειακά αποσπάσματα του ερωτηματολογίου.....	42

Διάγραμμα 23. Precision επισημειωτών για τα Ιλιαδικά και Οδυσσειακά αποσπάσματα του ερωτηματολογίου.....	43
Διάγραμμα 24. Recall επισημειωτών για τα Ιλιαδικά και Οδυσσειακά αποσπάσματα του ερωτηματολογίου.....	43



## ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1. Παράδειγμα κωδικοποίησης TEI.....	14
Εικόνα 2. Παράδειγμα της κωδικοποίησης one-hot embeddings.....	17
Εικόνα 3. Εμφάνιση των λέξεων «Ιλιάδα» και «Οδύσσεια» σε Αγγλικά κείμενα.....	20
Εικόνα 4. Ενδεικτικό παράδειγμα αποσπάσματος από τον Ομηρικό ύμνο «Εἰς Δημήτραν».....	23

# 1. Εισαγωγή

Ο ομηριστής Albin Lesky πριν από περίπου 40 χρόνια δήλωσε αναφορικά με τη σύγχρονη ομηρική έρευνα ότι “Η ενασχόληση με το ομηρικό ζήτημα από την εποχή του Friedrich August Wolf (1759 – 1824) μπορεί να χαρακτηριστεί ως το πιο αμφιλεγόμενο κεφάλαιο της φιλολογικής έρευνας” [11]. Ο ποιητής Όμηρος και η σύνθεση των ομηρικών επών αποτέλεσαν αντικείμενο βαθέως προβληματισμού από την αρχαιότητα,<sup>1</sup> εντούτοις το ομηρικό ζήτημα ανέκυψε τον 19<sup>ο</sup> αιώνα. Το εν λόγω ζήτημα, κατά την κρατούσα άποψη, αφορά στην αμφισβήτηση, αφενός της ύπαρξης του ποιητή Ομήρου και της πατρότητας των επών, αφετέρου εάν η σύνθεση των επών πραγματοποιήθηκε από έναν ή περισσότερους συνθέτες [25]. Κατά τον 20<sup>ο</sup> αιώνα, που ακολούθησε της διατύπωσης του ομηρικού ζητήματος, δημοσιεύτηκαν πολλά σπουδαία έργα για τον Όμηρο και για τα θέματα που συζητούνταν στους φιλολογικούς κύκλους του 19<sup>ου</sup> αιώνα. Παρά την πληθώρα των ακαδημαϊκών μελετών και τις νέες μεταφράσεις των *ομηρικών επών*, το ομηρικό ζήτημα δυστυχώς δεν επιλύθηκε μέχρι σήμερα. Στην αρχαιότητα πολλά έργα αποδίδονται στον Όμηρο, συμπεριλαμβανομένων των *Ομηρικών ύμνων*, της *Βατραχομομαχίας*, του *Μαργίτου* και της *Θηβαΐδος*. Ωστόσο, μεταξύ όλων των έργων που αποδίδονται στον Όμηρο, μόνο δύο μπορούν να θεωρηθούν αληθινά έργα του και ως εκ τούτου να χαρακτηριστούν ως τα πρώτα ποιητικά δημιουργήματα της Ευρώπης, που συντέθηκαν με τη συνδρομή της γραφής, ήτοι η *Ιλιάδα* και η *Οδύσσεια*, αν και κάποιοι ερευνητές αποδίδουν μόνο την *Ιλιάδα* στον Όμηρο [25]. Κατόπιν του προαναφερθέντος ιστορικού των προσπαθειών για την *απόδοση της συγγραφικής πατρότητας* πρωτίστως της *Ιλιάδας* και της *Οδύσσειας* και δευτερευόντως των καλουμένων ως *Ομηρικών ύμνων* στον Όμηρο, δημιουργήθηκε η πρόσκληση της έρευνας και με το πρίσμα της συνδρομής των υπολογιστικών μεθόδων.

---

<sup>1</sup> J. E. Sandys, *A History of Classical Scholarship* (Νέα Υόρκη, 1967· 1<sup>η</sup> έκδοση 1903 – 1908), τόμ. 3· U. von Wilamowitz-Moellendorf, *History of Classical Scholarship*, μτφρ. A. Harris (Λονδίνο, 1982· 1<sup>η</sup> γερμανική έκδοση 1921)· Davison (1962b)· Turner (1981) 135-186.

Τα ερωτήματα που θα μας απασχολήσουν στην παρούσα εργασία είναι τα εξής:

1. Υπάρχουν ραψωδίες στην *Ιλιάδα* και αντίστοιχα την *Οδύσσεια* που εμφανίζουν μεγαλύτερη γλωσσική συγγένεια με το σύνολο του αντίστοιχου έπους;
2. Υπάρχουν ραψωδίες στην *Ιλιάδα* και την *Οδύσσεια* που παρεκκλίνουν από το γλωσσικό ύφος των *ομηρικών επών*;
3. Πόσο ομοιάζουν γλωσσικά οι *Ομηρικοί ύμνοι*: *Εἰς Ἀπόλλωνα*, *Εἰς Ἀφροδίτην*, *Εἰς Δημήτραν* και *Εἰς Ἑρμῆν* στα *ομηρικά έπη*;
4. Μπορούν τεχνητά γλωσσικά μοντέλα να κατηγοριοποιήσουν με μεγαλύτερη επιτυχία αποσπάσματα της *Ιλιάδας* και της *Οδύσσειας* στο αντίστοιχο έπος απ' ότι η ανθρώπινη κρίση;

## 2. Βιβλιογραφική επισκόπηση

Η Επεξεργασία της Φυσικής Γλώσσας, ευρέως γνωστή με τον όρο *NLP* (*Natural Language Processing*), είναι η τεχνολογία που χρησιμοποιείται από τις μηχανές για την κατανόηση, την ανάλυση καθώς και την ερμηνεία των φυσικών γλωσσών του ανθρώπου [16]. Αξιοποιείται σε πλήθος εφαρμογών ανάμεσα στις οποίες βρίσκεται και η *απόδοση συγγραφικής πατρότητας* (*Authorship attribution*). Η αυτοματοποιημένη *απόδοση συγγραφικής πατρότητας* είναι το ζητούμενο αναγνώρισης του συγγραφέα ενός ανώνυμου κειμένου ή κειμένου του οποίου η πατρότητα αμφισβητείται [12].

Ανατρέχοντας στα ιστορικά δεδομένα εντοπίζουμε τις πρώτες προσπάθειες *απόδοσης συγγραφικής πατρότητας* τον 18<sup>ο</sup> αι. στην Ευρώπη στα έργα του Άγγλου ποιητή και θεατρικού συγγραφέα, Ουίλλιαμ Σαίξπηρ [28]. Συνεχίζοντας στον 19<sup>ο</sup> αι., το ζητούμενο αυτό επανέρχεται σε ένα από τα σπουδαιότερα προβλήματα στον τομέα των Κλασικών Σπουδών, δηλαδή στο θέμα της χρονολόγησης των πλατωνικών διαλόγων [28]. Τον 20<sup>ο</sup> αι. διάσημο παράδειγμα *απόδοσης συγγραφικής πατρότητας* καθίσταται η πατρότητα των Ομοσπονδιακών κειμένων (*The Federalistic Papers*) των ΗΠΑ,<sup>2</sup> εκ των οποίων 12 κείμενα από αυτά πιθανολογείται πώς γράφτηκαν από τους συγγραφείς Alexander Hamilton και James Madison [7]. Το ενδιαφέρον για το πεδίο αυτό παραμένει και σήμερα, τον 21<sup>ο</sup> αι., αμείωτο και γίνεται ολοένα πιο αυξανόμενο και στην επιστήμη των υπολογιστών.

Η *απόδοση συγγραφικής πατρότητας* βρίσκει πρόσφορο έδαφος σε μία πληθώρα εφαρμογών, ανάμεσα στις οποίες είναι η διαδικασία ανίχνευσης λογοκλοπής [10], η εγκληματολογική έρευνα — στην περίπτωση αναγνώρισης του γραφέα ενός ανώνυμου εγγράφου [2] — και η διαδικασία εντοπισμού ηλεκτρονικού «ψαρέματος» μέσω ηλεκτρονικών μηνυμάτων, γνωστού ως «*phishing*» [5].

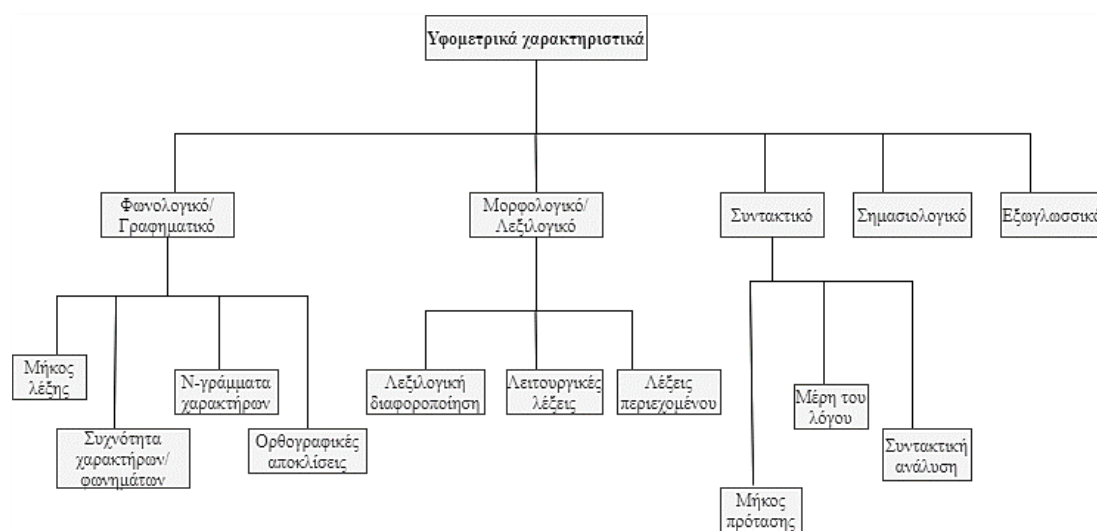
Τα γλωσσικά μοντέλα (*Language models-LMs*) αποτελούν ακρογωνιαίο λίθο σε ποικίλους τομείς του *NLP*, όπως φερειπείν στην αυτόματη μετάφραση [21], [13]. Εντούτοις, λίγες είναι οι μελέτες που έχουν εφαρμόσει γλωσσικά μοντέλα για το καθήκον της *απόδοσης συγγραφικής πατρότητας*, ανάμεσα στις οποίες βρίσκεται μία μελέτη που εξετάζει την απόδοση γλωσσικών μοντέλων βάσει ποικίλων

---

<sup>2</sup> Πρόκειται για μία συλλογή 85 άρθρων και δοκιμίων που εκδόθηκε στις 27 Οκτωβρίου 1787 και γράφτηκε από τον Alexander Hamilton, James Madison και John Jay αναφορικά με το σύνταγμα των Ηνωμένων Πολιτειών.

χαρακτηριστικών, άμεσα εξαρτωμένων από τη γλώσσα των εκάστοτε δεδομένων [19]. Κάθε φορά η επιλογή των κατάλληλων χαρακτηριστικών είναι ιδιαίτερα σημαντική και έτσι συνήθως αποκλείονται μη πληροφοριακά χαρακτηριστικά [22], αλλά ωστόσο πιο σπάνια χαρακτηριστικά ενδεχομένως να έχουν σημαντική επίδραση στην τελική απόδοση του γλωσσικού μοντέλου.

Ο προσδιορισμός της *συγγραφικής πατρότητας* ενός κειμένου προκύπτει από την ποσοτική έρευνα ορισμένων στοιχείων, που ονομάζονται «*υφομετρικά χαρακτηριστικά*». Τα στοιχεία αυτά, βάσει του γλωσσικού επιπέδου υπαγωγής τους, διακρίνονται στις εξής τέσσερις κατηγορίες: φωνολογικά/γραφηματικά, λεξιλογικά/μορφολογικά, συντακτικά και σημασιολογικά (βλ. Σχήμα 1) [23].



Σχήμα 1. Τυπολογία υφομετρικών χαρακτηριστικών

Ένα ευρύ φάσμα χαρακτηριστικών έχει εφαρμοστεί στο έργο προσδιορισμού της *απόδοσης συγγραφικής πατρότητας* και πολλές μελέτες κατέληξαν στο συμπέρασμα ότι η χρήση γλωσσικών μοντέλων *n-γραμμάτων χαρακτήρων* (*character n-grams*) — οι οποίοι χαρακτήρες εντάσσονται στα φωνολογικά/γραφηματικά «*υφομετρικά χαρακτηριστικά*» — αποβαίνει σε αρκετές περιπτώσεις αποδοτική. Συγκεκριμένα, στη μελέτη [19] χρησιμοποιήθηκαν μέθοδοι εκτίμησης της *απόδοσης συγγραφικής πατρότητας* από τη *Στατιστική Γλωσσική Μοντελοποίηση* (*Statistical Language Modeling – SLM*). Η προσέγγιση *n-γραμμάτων χαρακτήρων* πέτυχε απόδοση τεχνολογίας αιχμής (*state-of-the-art*) σε τρία σύνολα δεδομένων που αφορούσαν τρεις γλώσσες: Ελληνικά, Αγγλικά και Κινεζικά. Πιο συγκεκριμένα, στη μελέτη [19] επετεύχθη βελτίωση ακρίβειας (*accuracy*) περίπου 18% συγκριτικά με τα καλύτερα δημοσιευμένα αποτελέσματα για το σύνολο δεδομένων της ελληνικής γλώσσας. Ένας

παράγοντας που επηρεάζει την απόδοση ενός γλωσσικού μοντέλου είναι η τεχνική ομαλοποίησης (*Smoothing technique*),<sup>3</sup> η οποία στις περισσότερες περιπτώσεις δεν έχει σημαντική επίδραση στην απόδοση συγγραφικής πατρότητας. Είναι δυνατό να χρησιμοποιηθεί οποιαδήποτε τεχνική ομαλοποίησης σε αυτό το πρόβλημα και να αποκτηθεί συγκρίσιμη απόδοση, επειδή οι βαθμολογίες της μέτρησης αξιολόγησης *Perplexity*,<sup>4</sup> που παράγονται, είναι σχεδόν πάντα ίδιες [19].

Τα αποτελέσματα από τα πειράματα της προαναφερθείσας μελέτης [19] είναι ικανοποιητικά, αλλά παρόλα αυτά τα *SLM* εξετάζουν περιορισμένο μήκος συμφραζομένων (*context*) [17]. Αυτή η παράμετρος αποτρέπει το γλωσσικό μοντέλο να λάβει πληροφορίες από μεγαλύτερες ακολουθίες (*sequences*) συμφραζομένων, οι οποίες ενδέχεται να είναι ωφέλιμες για την απόδοση συγγραφικής πατρότητας. Πιο πρόσφατες έρευνες στη γλωσσική μοντελοποίηση πέτυχαν αποτελέσματα τεχνολογίας αιχμής (*state-of-the-art*) μοντέλων [17] μέσω Αναδρομικών Νευρωνικών Δικτύων (*Recurrent Neural Network-RNN*) και Μοντέλων Μακράς - Βραχείας Μνήμης (*Long Short-Term Memory-LSTM*). Τα αναφερόμενα Νευρωνικά Γλωσσικά Μοντέλα (*Neural Language Models*), εν αντιθέσει των Στατιστικών Γλωσσικών Μοντέλων, κρίνονται επωφελή για την επεξεργασία διαδοχικών δεδομένων (*sequential data*), με αιτιολογία ότι οι επαναλαμβανόμενες αλληλεξαρτώμενες συνδέσεις τους επιτρέπουν να επεξεργάζονται πληροφορίες από μεγαλύτερες ακολουθίες.

Σε ένα γλωσσικό μοντέλο κρίνεται σημαντικό να αποσαφηνίσουμε και στη συνέχεια να ακολουθήσουμε, αν το μοντέλο θα προσεγγίσει τα δεδομένα σε επίπεδο χαρακτήρα ή σε επίπεδο λέξης. Στη συνέχεια, χάριν συντομίας, θα αναφερόμαστε στα γλωσσικά μοντέλα χαρακτήρων με τη συντομογραφία *CLLM* (*Character-Level Language Model*) και στα γλωσσικά μοντέλα λέξεων *WLLM* (*Word-Level Language Model*). Τα περισσότερα συστήματα απόδοσης συγγραφικής πατρότητας εκτελούν κατά βάση την ανάλυσή τους σε επίπεδο λέξεων [19]. Ωστόσο, σε αρκετές γλώσσες, όπως τις Ασιατικές (π.χ. Ιαπωνικά, Κινεζικά κ.ά.), τα όρια των λέξεων δεν είναι σαφή και επομένως η αναγνώρισή τους καθίσταται δυσχερής στο κείμενο [19]. Τα στατιστικά *CLLM* βρίσκουν πεδίο εφαρμογής σε κάθε γλώσσα, ακόμα και σε μη γλωσσικές ακολουθίες όπως της μουσικής [3] και του DNA [15]. Ταυτοχρόνως χρησιμοποιούνται

---

<sup>3</sup> Βλ. σελ. 13-17 στο <http://web.stanford.edu/~jurafsky/slp3/3.pdf> [ημερομηνία πρόσβασης: 29 Νοεμβρίου 2020].

<sup>4</sup> Για τη μέτρηση αξιολόγησης *Perplexity* βλ. Υποενότητα 4.4.1 της παρούσας εργασίας.

ευρέως σε εφαρμογές εξόρυξης δεδομένων [24]. Τα στατιστικά και νευρωνικά γλωσσικά μοντέλα χαρακτήρων έχουν εφαρμοστεί σε έρευνες [19], [20] και έχουν σημειώσει καλύτερες αποδόσεις από τα αντίστοιχα γλωσσικά μοντέλα λέξεων.

Ανατρέχοντας σε μελέτες προσδιορισμού απόδοσης συγγραφικής πατρότητας οφείλουμε να αναφερθούμε στην περίοδο της Κλασικής αρχαιότητας. Χαρακτηριστική περίπτωση αποτελεί η επαλήθευση συγγραφέα μέσω *Μοντέλων Διανυσματικού Χώρου*<sup>5</sup> (*Vector Space Models-VSM*) στο έργο «Απομνημονεύματα περί του Γαλατικού πολέμου» (*Commentarii de Bello Gallico*) του Ιουλίου Καίσαρα (100-44 π.Χ.). Ενώ το έργο αποδίδεται παραδοσιακά στον Καίσαρα, ωστόσο εικάζεται ότι συνέβαλε σημαντικά στη συγγραφή του σώματος κειμένου (*corpus*) ο Αύλος Χίρτιος (*Aulus Hirtius*), ένας από τους πιο αξιόπιστους στρατηγούς του [9]. Άλλη αξιοσημείωτη περίπτωση αποτελεί η μελέτη απόδοσης συγγραφικής πατρότητας στην τραγωδία «Ρήσος» (453 π.Χ.) του Ευριπίδη με *Μηχανές Διανυσμάτων Υποστήριξης* (*Support Vector Machines-SVM*)<sup>6</sup> και αλγορίθμου εντοπισμού λογοκλοπής (*Plagiarism*) [14].

Στο πλαίσιο της παρούσας έρευνας εξετάσαμε την περίπτωση του Ομήρου, ποιητή κορυφαίου βεληνεκού της Κλασικής αρχαιότητας. Ειδικότερα, μελετήσαμε το ποσοστό κατά το οποίο, *Στατιστικά* και *Νευρωνικά* γλωσσικά μοντέλα χαρακτήρων, που έχουν εκπαιδευτεί σε Ομηρικά δεδομένα, είναι σε θέση να ανιχνεύσουν, εάν αρχαία κείμενα, που κατά την κρατούσα άποψη αποδίδονται στον ποιητή Όμηρο, φέρουν την πατρότητα του τελευταίου. Προκειμένου να επιτύχουμε τον εν λόγω σκοπό, χρησιμοποιήσαμε *SLM* και *LSTM*, αξιολογώντας με τη σειρά μας τις δύο κατηγορίες γλωσσικών μοντέλων χρησιμοποιώντας τις μετρήσεις αξιολόγησης *Perplexity* και *F-measure* (ή *F1-score*).

---

<sup>5</sup> Πρόκειται για μία τεχνική ανάκτησης δεδομένων, που προτάθηκε από τον Gerard Salton στις αρχές του 1960. Το μοντέλο βασίζεται στο μετασχηματισμό των κειμένων σε διανύσματα με συντεταγμένες πραγματικούς αριθμούς χρησιμοποιώντας την έννοια του διανυσματικού χώρου από τη γραμμική άλγεβρα ([http://ipl.cs.aueb.gr/info\\_ret/vsm/vsm\\_6.htm](http://ipl.cs.aueb.gr/info_ret/vsm/vsm_6.htm)) [ημερομηνία πρόσβασης: 29 Νοεμβρίου 2020].

<sup>6</sup> Οι *Μηχανές Διανυσμάτων Υποστήριξης* (*Support Vector Machines-SVM*) αποτελούν μία σύγχρονη αποτελεσματική προσέγγιση της επίλυσης ζητημάτων κατηγοριοποίησης (Διαμαντάρας και Μπότσης, 2019, σ.186).

### 3. Δεδομένα

Στο κεφάλαιο αυτό αναλύουμε και περιγράφουμε τα δεδομένα της παρούσας έρευνας. Αναλυτικότερα, παρουσιάζονται γενικές πληροφορίες για τα δεδομένα καθώς και το στάδιο της *προ-επεξεργασίας των δεδομένων (Data preprocessing)* για την εκπαίδευση των *στατιστικών και νευρωνικών γλωσσικών μοντέλων* που θα περιγράψουμε στη συνέχεια.

#### 3.1 Περιγραφή δεδομένων

Το σύνολο των δεδομένων (dataset) που χρησιμοποιήθηκε για το πείραμα ήταν τα *ομηρικά έπη*, η *Ιλιάδα* (15.693 στίχοι) και η *Οδύσσεια* (12.110 στίχοι), καθώς και 4 από τους 33 καλούμενους *Ομηρικούς ύμνους*: *Εἰς Απόλλωνα* (546 στίχοι), *Εἰς Ἀφροδίτην* (293 στίχοι), *Εἰς Δημήτραν* (495 στίχοι) και *Εἰς Ἑρμῆν* (580 στίχοι). Η σχετικά άνιση έκταση των δύο *ομηρικών επών* είναι ανάλογη του διαφορετικού θεματικού τους περιβάλλοντος [27]. Η *Ιλιάδα* χαρακτηρίζεται κατά βάση ως πολεμικό ποίημα, όπως μας προδιαθέτει ο Όμηρος στο προοίμιό του έργου με τη λέξη «μῆνις» (= οργή, θυμός)<sup>7</sup> του Αχιλλέα και η *Οδύσσεια* ως μεταπολεμικό ποίημα που περιγράφει τις περιπέτειες του Οδυσσέα προκειμένου να επιστρέψει στην πατρίδα του χωρίς να αποκλείει τα πολεμικά επεισόδια, όπως για παράδειγμα τη «Μνηστηροφονία» (= φόνος των μνηστήρων). Οι *Ομηρικοί ύμνοι* αποτελούν μία συλλογή από 33 αφηγηματικά ποιήματα αφιερωμένα σε θεούς με έκταση από 3 έως 580 στίχους. Ορισμένοι συγγραφείς αποδίδουν τους *Ομηρικούς ύμνους* στον Όμηρο, ωστόσο οι Αλεξανδρινοί φιλόλογοι φαίνεται να είχαν αφαιρέσει αυτή τη συλλογή από το συνολικό έργο του ποιητή, πράγμα το οποίο είχε σημαντικές συνέπειες στο κύρος των ύμνων ως κλασικά αναγνώσματα [18]. Οι ύμνοι που επιλέξαμε να μελετήσουμε στην παρούσα εργασία υπάγονται στην κατηγορία των «μεγάλων ύμνων», μαζί με τον ύμνο *Εἰς Βάκχον*, που δε συμπεριλαμβάνεται στα δεδομένα.

Τα εν λόγω δεδομένα, *ομηρικά έπη* και οι 4 επιλεγθέντες *Ομηρικοί ύμνοι*, αντλήθηκαν από το *Thesaurus Linguae Graecae (TLG)*,<sup>8</sup> το οποίο αποτελεί μία βάση δεδομένων πλήρους κειμένου, που περιλαμβάνει όλα τα σωζόμενα κείμενα της αρχαίας

<sup>7</sup> Μετάφραση της λέξης «μῆνις»: Liddell, H. G., Scott, R., Jones, H. S., McKenzie, R., & Thesaurus Linguae Graecae Project. (2011). *The online Liddell-Scott-Jones Greek-English lexicon*. Irvine, CA: University of California, Irvine.

<sup>8</sup> <http://stephanus.tlg.uci.edu/> [ημερομηνία πρόσβασης: 29 Νοεμβρίου 2020].



ελληνικής γραμματείας από τον 8<sup>ο</sup> αι. π.Χ. μέχρι το 600 μ.Χ. Επιπλέον, περιέχει ιστοριογραφικά και λεξικογραφικά κείμενα, όπως και κείμενα σχολιαστών της περιόδου 600 μ.Χ. έως το 1453 μ.Χ. Αναφέρονται σε ακόλουθες παραπομπές οι εκδόσεις που χρησιμοποιεί το *TLG* για την *Ιλιάδα*,<sup>9</sup> την *Οδύσσεια*<sup>10</sup> καθώς και τους *Ομηρικούς ύμνους*<sup>11</sup>. Τα κείμενα ήταν κωδικοποιημένα σε *Text Encoding Initiative (TEI)*,<sup>12</sup> το οποίο αποτελεί μία κοινοπραξία, που αναπτύσσει και διατηρεί συλλογικά ένα πρότυπο για την αναπαράσταση κειμένων σε ψηφιακή μορφή. Κύριο έργο του καθίσταται ένα σύνολο κατευθυντήριων γραμμών, που καθορίζουν μεθόδους κωδικοποίησης για μηχανικά αναγνώσιμα κείμενα, κυρίως στις ανθρωπιστικές επιστήμες, τις κοινωνικές επιστήμες και τη γλωσσολογία.

### 3.2 Προ-επεξεργασία δεδομένων (Data preprocessing)

Στον κλάδο της *Επεξεργασίας Φυσικής Γλώσσας* ένα από τα συχνότερα ζητήματα που προκύπτουν είναι ότι τα δεδομένα που τίθενται προς ανάλυση πάσχουν από πολλά και διαφορετικά προβλήματα. Για να αντιμετωπιστούν όσο το δυνατόν αποτελεσματικότερα τα προβλήματα αυτά, το σύνολο δεδομένων (dataset) πρέπει να μετατραπεί σε μία μορφή που να ταιριάζει όσο το δυνατόν καλύτερα στη γλώσσα που αντιλαμβάνονται τα πληροφοριακά συστήματα. Σε αυτό το σημείο έρχεται η *προ-επεξεργασία των δεδομένων (Data preprocessing)* για να διαχειριστεί σε ένα μεγάλο βαθμό τα προβλήματα που ανακύπτουν. Η *προ-επεξεργασία των δεδομένων* αφορά τεχνικές «καθαρισμού» των δεδομένων (*data cleaning*) με στόχο τη μετατροπή του μη επεξεργασμένου κειμένου (*raw text*) σε επεξεργασμένο κείμενο. Ενδεικτικά αναφέρουμε μερικές από τις τεχνικές αυτές: μετατροπή κειμένου σε πεζά γράμματα (*lowercasing*), αφαίρεση σημείων στίξης (*remove punctuation*), διαχωρισμός των λέξεων, ώστε οι λέξεις να αντιμετωπίζονται ως *tokens (tokenize)*, αφαίρεση τονισμού (*detone*) κ.ά. Στη δική μας περίπτωση, τα δεδομένα που χρησιμοποιήσαμε χρειάστηκε αρχικά να καθαριστούν από τη σήμανση που έφερε η κωδικοποίηση *TEI* (βλ. Εικόνα 1).

---

<sup>9</sup> M.L. West, *Homeri Ilias [Bibliotheca Scriptorum Graecorum et Romanorum Teubneriana]*. Stuttgart-Leipzig: Teubner.

<sup>10</sup> P. von der Mühl, *Homeri Odyssea*, Basel: Helbing & Lichtenhahn, 1962: 1-456.

<sup>11</sup> T.W. Allen, W.R. Halliday, and E.E. Sikes, *The Homeric hymns*, 2nd edn., Oxford: Clarendon Press, 1936.

<sup>12</sup> <https://tei-c.org/> [ημερομηνία πρόσβασης: 29 Νοεμβρίου 2020].

```

<div type="Book" n="1">
<l n="t"><label type="head">ΙΛΙΑΔΟΣ Α</label> </l>
<l n="1">Μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος </l>
<l n="2"><seg rend="Marginalia">&gt;</seg> οὐλομένην, ἣ μυρὶ' Ἀχαιοῖς ἄλγε' ἔθηκε, </l>
<l n="3"><seg rend="Marginalia">&gt;</seg> πολλὰς δ' ἰφθίμους ψυχὰς Ἀΐδι προΐαψεν </l>
<l n="4">ἠρώων, αὐτοὺς δὲ ἐλώρια τεῦχε κύνεσσιν </l>
<l n="5"><seg rend="Marginalia">&lt;</seg> οἴωνοῖσι τε πᾶσι, Διὸς δ' ἐτελείετο βουλή, </l>
<l n="6">ἔξ οὔ δὴ τὰ πρῶτα διαστήτην ἐρίσαντε </l>
<l n="7"><seg rend="Marginalia">&lt;</seg> Ἀτρεΐδης τε ἄναξ ἀνδρῶν καὶ δῖος Ἀχιλλεύς. </l>

```

Εικόνα 1. Παράδειγμα κωδικοποίησης TEI

Ύστερα απ' αυτή τη διαδικασία, τα δεδομένα της έρευνάς μας ήταν πλέον το σώμα κειμένου, δηλαδή τα *ομηρικά έπη* καθώς και οι τέσσερις *Ομηρικοί ύμνοι*, όπως αυτά διασώζονται στην πρωτότυπη μορφή τους. Ακολουθώς, αφαιρέθηκαν όλα τα σημεία στίξης καθώς και τα εκδοτικά σύμβολα που δεν προσέφεραν ιδιαίτερη σημασιολογική πληροφορία (βλ. Πίνακας 1) στο κείμενο και ακόμα το κείμενο στο σύνολό του μετατράπηκε σε πεζά γράμματα (*lowercasing*).

'	,	.	`	''	※	•	ᾀ	;	‘	†	ras
[	]	>	<	⋞	–	⌋	€	·	,	–	corr

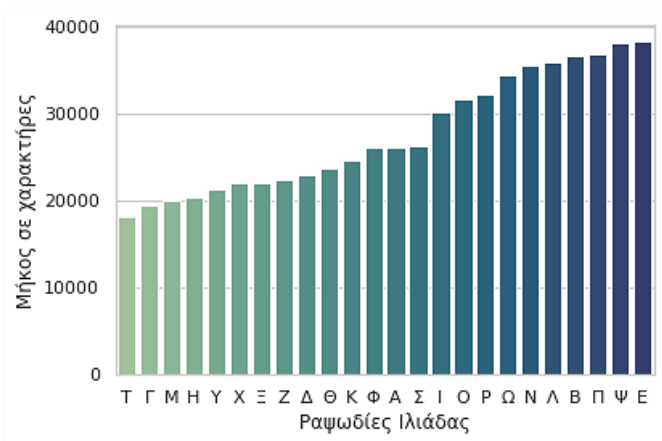
Πίνακας 1. Σημεία στίξης και εκδοτικά σύμβολα που αφαιρέθηκαν από τα κωδικοποιημένα δεδομένα

Αποφασίσαμε να μην αφαιρέσουμε τον τονισμό καθώς και τα πνεύματα από τις λέξεις, για δύο λόγους που θα αναλύσουμε ευθύς αμέσως. Ο πρώτος λόγος είναι ότι οι τόνοι και τα πνεύματα θεωρούνται μέρος της ορθογραφίας μιας λέξης, έτσι ένας διαφορετικός τόνος είναι πιθανό να αλλάξει σημασιολογικά μία λέξη (π.χ. *θολός* και *θόλος*). Αυτό συμβαίνει, ωστόσο, σε πολύ μικρό ποσοστό λέξεων και τις περισσότερες φορές καταλαβαίνουμε ποια λέξη διαβάζουμε, ακόμη και στην περίπτωση ορθογραφικού λάθους ή χωρίς την παρουσία τονισμού. Η καταγραφή των διακριτικών αυτών σημείων ξεκίνησε από τους Αλεξανδρινούς φιλόλογους και δεν καθιερώθηκαν καθολικά, παρά μόνο στη Βυζαντινή εποχή. Την περίοδο καταγραφής των *ομηρικών επών* και των *Ομηρικών ύμνων*, ο τονισμός δε γραφόταν. Η πλειοψηφία, όμως, των μαρτύρων (χειρογράφων) που χρησιμοποιήθηκαν για την έκδοση του *TLG* περιείχε τόσο τους τόνους όσο και τα πνεύματα. Ο δεύτερος λόγος που επιλέξαμε να κρατήσουμε τα διακριτικά αυτά σημεία είναι διότι ο τονισμός αποτελεί αναπόσπαστο μέρος των μετρικών κανόνων και επειδή πρόκειται για έμμετρα ποιήματα, υπάρχουν σοβαρές πιθανότητες να έχουν κάποιο ρόλο στην αλληλουχία των χαρακτήρων.

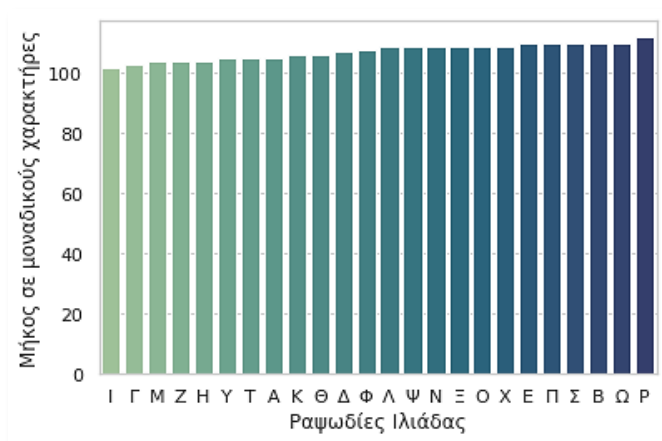
Στα διαγράμματα που ακολουθούν μπορούμε να δούμε τα μήκη σε συνολικούς και μοναδικούς χαρακτήρες καθώς και σε λέξεις των ραψωδιών της *Ιλιάδας*, της

Οδύσσειας και των Ομηρικών ύμνων: Εἰς Απόλλωνα, Εἰς Ἀφροδίτην, Εἰς Δημήτραν και Εἰς Ἑρμῆν, ὕστερα ἀπὸ τὴν προ-επεξεργασία τῶν δεδομένων που εφαρμόσαμε.

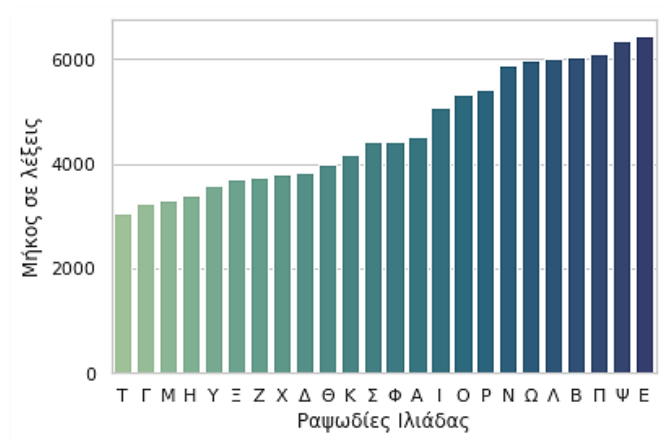
## Ιλιάδα



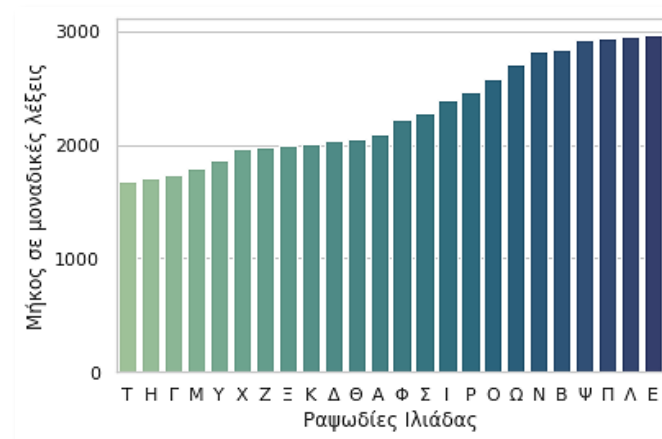
Διάγραμμα 1. Μήκος σε χαρακτήρες στις ραψωδίες της Ιλιάδας



Διάγραμμα 2. Μήκος σε μοναδικούς χαρακτήρες στις ραψωδίες της Ιλιάδας

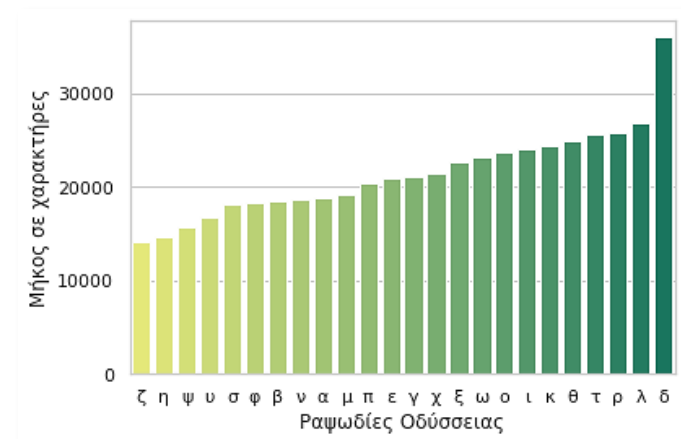


Διάγραμμα 3. Μήκος σε λέξεις στις ραψωδίες της Ιλιάδας

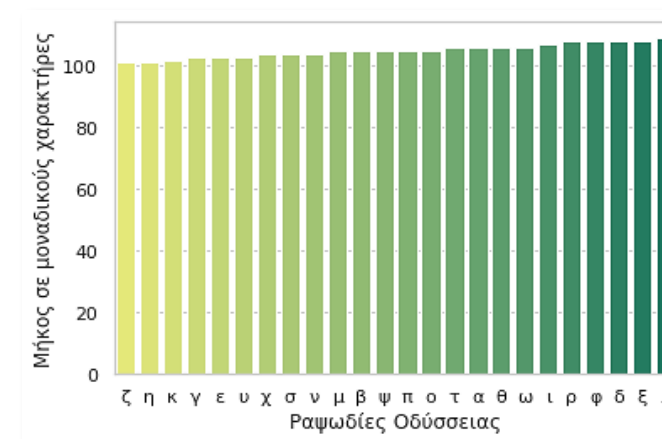


Διάγραμμα 4. Μήκος σε μοναδικές λέξεις στις ραψωδίες της Ιλιάδας

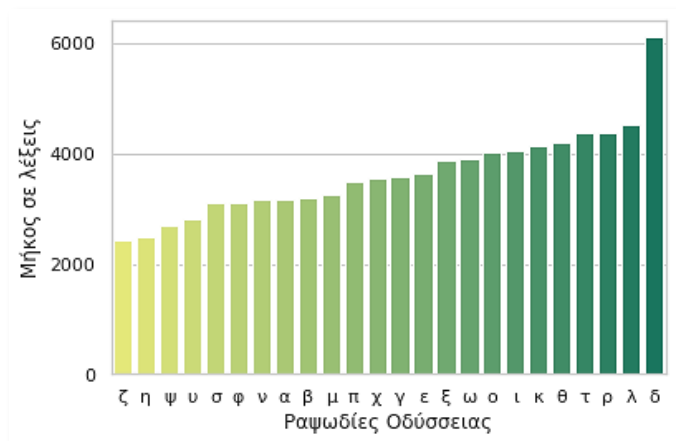
## Οδύσσεια



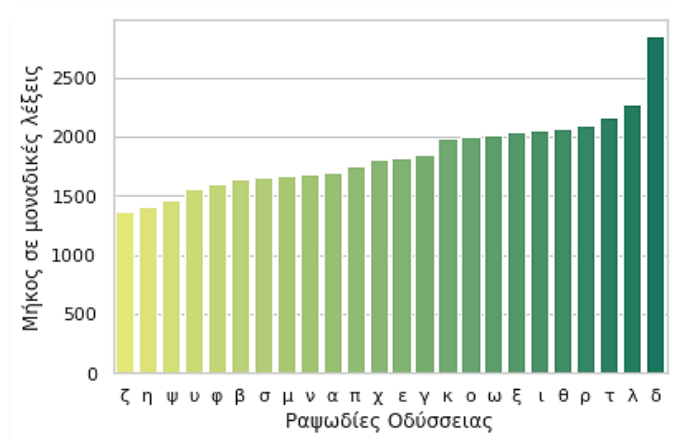
Διάγραμμα 5. Μήκος σε χαρακτήρες στις ραψωδίες της Οδύσσειας



Διάγραμμα 6. Μήκος σε μοναδικούς χαρακτήρες στις ραψωδίες της Οδύσσειας

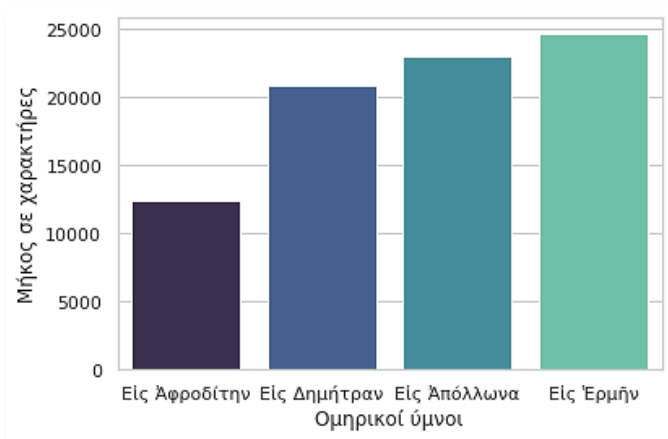


Διάγραμμα 7. Μήκος σε λέξεις στις ραψωδίες της Οδύσσειας

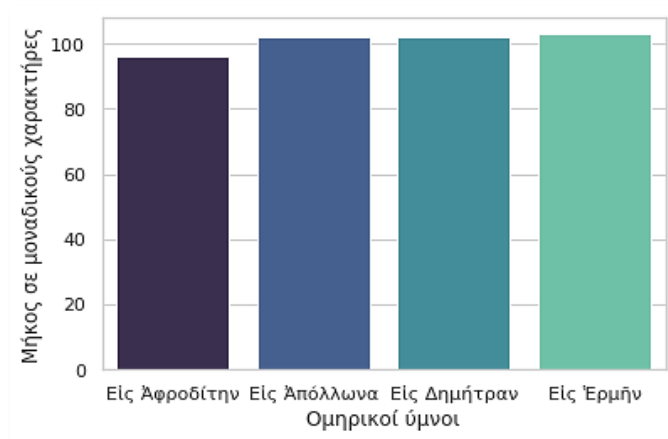


Διάγραμμα 8. Μήκος σε μοναδικές λέξεις στις ραψωδίες της Οδύσσειας

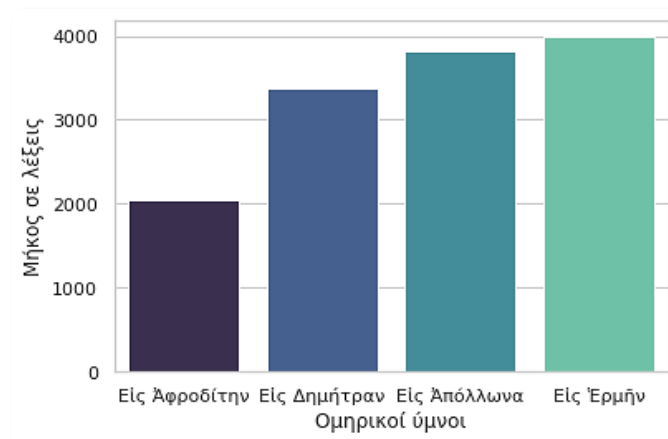
## Ομηρικοί ύμνοι



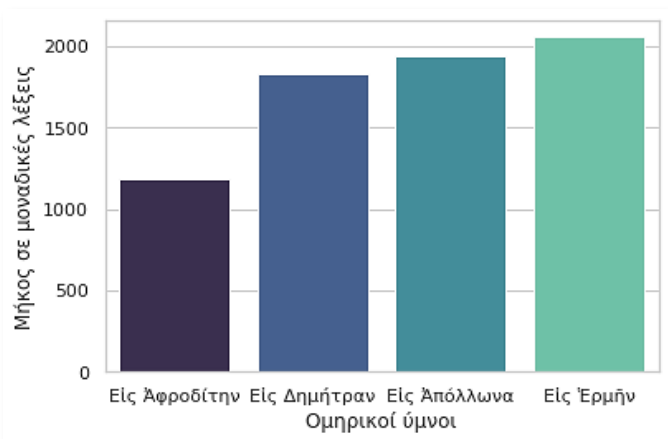
Διάγραμμα 9. Μήκος σε χαρακτήρες στους Ομηρικούς ύμνους



Διάγραμμα 10. Μήκος σε μοναδικούς χαρακτήρες στους Ομηρικούς ύμνους



Διάγραμμα 11. Μήκος σε λέξεις στους Ομηρικούς ύμνους



Διάγραμμα 12. Μήκος σε μοναδικές λέξεις στους Ομηρικούς ύμνους

Η προ-επεξεργασία δεδομένων που ακολουθήσαμε διαμόρφωσε το λεξιλόγιο των χαρακτήρων από τα κείμενα της *Ιλιάδας*, της *Οδύσσειας* και των *Ομηρικών ύμνων* στους 121 χαρακτήρες (120 γράμματα και 1 χαρακτήρας το κενό ή space). Στον Πίνακα 2. που ακολουθεί μπορούμε να δούμε αναλυτικά τους χαρακτήρες.

ı	ά	έ	ή	ί	ϋ	α	β	γ	δ	ε	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο	π	ρ	ς	σ	τ	υ	
φ	χ	ψ	ω	ϊ	ϋ	ό	ύ	ώ	ὰ	ά	ᾶ	ἄ	ᾷ	ἄ	ᾶ	ἐ	έ	ἔ	ἕ	ἔ	ἕ	ἦ	ή	ἦ	ἦ	ἦ	
ἦ	ἦ	ἦ	ι	ί	ἰ	ἱ	ἱ	ἱ	ἱ	ἱ	ὀ	ό	ὀ	ὀ	ὀ	ὐ	ύ	ὕ	ὕ	ὕ	ὕ	ὠ	ώ	ὠ	ὠ	ὠ	
ὠ	ὠ	ὠ	ὠ	ὰ	ἐ	ἦ	ι	ὀ	ὐ	ὠ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ὠ	ὠ	ὠ	ὠ	ᾱ	ᾱ	ᾱ	ἦ	ἦ	ἦ
ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	

Πίνακας 2. Χαρακτήρες *Ιλιάδας*, *Οδύσσειας* και *Ομηρικών ύμνων* σε lowercase

Μέχρι αυτό το σημείο τα βήματα που ακολουθήθηκαν για την προ-επεξεργασία δεδομένων ήταν κοινά και για το *SLM* και για το *LSTM*. Για το *LSTM* γίνεται χρήση της τεχνικής κωδικοποίησης *one-hot encoding*.<sup>13</sup>

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
L	→	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U	→	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
K	→	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	→	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	→	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

Εικόνα 2. Παράδειγμα της κωδικοποίησης *one-hot encoding*

Πηγή: <https://www.wandb.com/tutorial/text-classification-using-cnns>

[ημερομηνία πρόσβασης: 29 Νοεμβρίου 2020]

<sup>13</sup> Περισσότερες πληροφορίες για την κωδικοποίηση *one-hot encoding*

βλ. <https://www.wandb.com/tutorial/text-classification-using-cnns> [ημερομηνία πρόσβασης: 29 Νοεμβρίου 2020].

## 4. Μέθοδοι γλωσσικής μοντελοποίησης και μετρήσεις αξιολόγησης

Στο κεφάλαιο αυτό θα αναφερθούμε στην ερευνητική μεθοδολογία που ακολουθήθηκε. Πιο συγκεκριμένα, αρχικά θα περιγράψουμε τα πειράματα που εφαρμόσαμε μέσω των στατιστικών και νευρωνικών γλωσσικών μοντέλων, έπειτα θα παρουσιάσουμε τα δύο είδη γλωσσικών μοντέλων (*SLM* και *LSTM*), που υλοποιήθηκαν για τα εν λόγω πειράματα και στη συνέχεια θα αναφερθούμε στις μετρήσεις αξιολόγησης των πειραμάτων.

Προκειμένου να εξετάσουμε το ζητούμενο *απόδοσης συγγραφικής πατρότητας* στα δεδομένα της *Ιλιάδας*, της *Οδύσσειας* και των *Ομηρικών ύμνων* προβήκαμε στα πειράματα, που θα αναφέρουμε. Το πρώτο πείραμα που πραγματοποιήθηκε αφορούσε *SLM* και το δεύτερο πείραμα τη σύγκριση *SLM*, *LSTM* και *ανθρώπων-επισημειωτών* (*human-annotators*). Οι κρίσεις των *ανθρώπων-επισημειωτών* συλλέχθηκαν μέσω ερωτηματολογίου (βλ. Υποενότητα 4.3). Στον Πίνακα 3. που ακολουθεί μπορούμε να δούμε ονομαστικά τα πειράματα που υλοποιήθηκαν στην παρούσα μελέτη. Εν συνεχεία, στο Κεφάλαιο 5, θα ακολουθήσει η αναλυτική περιγραφή τους.

Πειράματα	1. Απόδοση συγγραφικής πατρότητας μέσω Στατιστικών Γλωσσικών Μοντέλων (SLM)	2. Ταξινόμηση ομηρικών κειμένων μέσω Στατιστικών Γλωσσικών Μοντέλων (SLM) και Νευρωνικών Γλωσσικών Μοντέλων Μακράς - Βραχείας μνήμης (LSTM)
Κατηγορίες γλωσσικών μοντέλων	<ol style="list-style-type: none"><li>Κυκλικά μοντέλα <i>SLM</i></li><li>Χιαστί μοντέλα <i>SLM</i></li></ol>	<ol style="list-style-type: none"><li>Ταξινόμηση βάσει <i>SLM</i></li><li>Ταξινόμηση βάσει <i>LSTM</i></li><li>Σύγκριση της κατηγορίας 1 &amp; 2 με τα αποσπάσματα του ερωτηματολογίου, που δόθηκαν στους ανθρώπους-επισημειωτές</li></ol>
Μετρήσεις αξιολόγησης	Perplexity	F-measure (ή F1-score)

Πίνακας 3. Κατάταξη πειραμάτων που εφαρμόστηκαν στην παρούσα μελέτη

## 4.1 Στατιστικό Γλωσσικό Μοντέλο (Statistical Language Model – SLM)

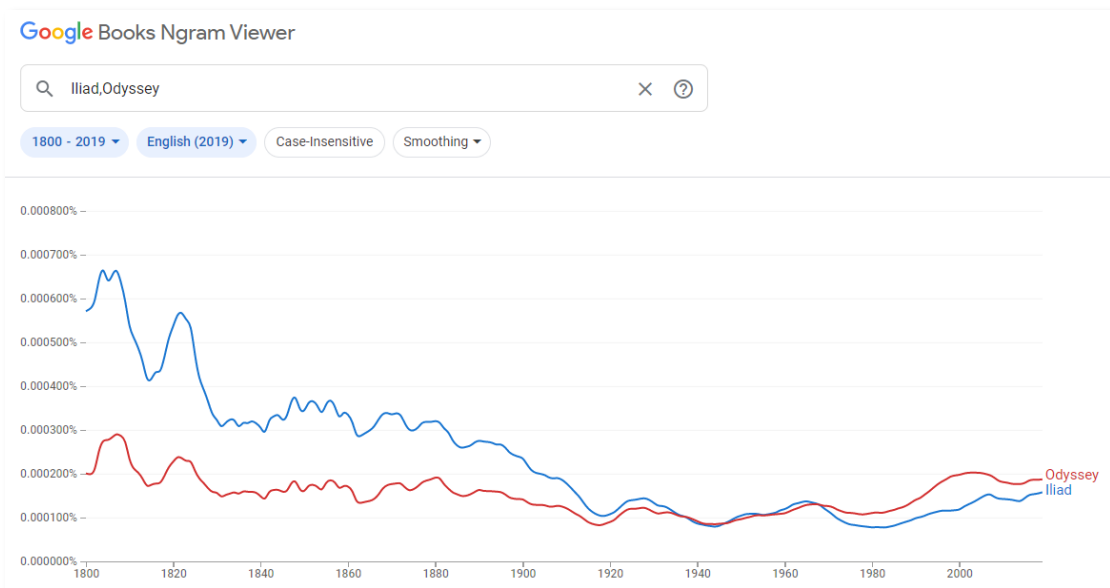
Το πείραμα για την απόδοση συγγραφικής πατρότητας επιτεύχθηκε μέσω Στατιστικών Γλωσσικών Μοντέλων (*Statistical Language Models – SLM*) επιπέδου χαρακτήρα (*Character-Level Language Models - CLLM*), ενώ το πείραμα για την ταξινόμηση ομηρικών κειμένων με *SLM* και *LSTM*. Τα στατιστικά γλωσσικά μοντέλα περιγράφονται ως μοντέλα που κατανέμουν πιθανότητες σε ακολουθίες λέξεων, όταν γνωρίζουμε τα συμφραζόμενα [8]. Μέρος της απαιτούμενης γνώσης για την πρόβλεψη των επόμενων χαρακτήρων σε ένα κείμενο εντοπίζεται μέσω αυτής της κατηγορίας μοντέλων. Για παράδειγμα, ένα γλωσσικό μοντέλο αυτής της κατηγορίας πρέπει να εκτιμήσει την πιθανότητα να εμφανιστεί ο χαρακτήρας “e” στο παράδειγμα συμφραζομένων “The color of lavender is purpl\_\_”:

$$P(\text{“e”} \mid \text{“The color of lavender is purpl\_”}) = ;$$

Τα *n-γράμματα χαρακτήρων* (*character n-grams*) συνιστούν μέχρι στιγμής ένα από τα πιο αποτελεσματικά «υφομετρικά χαρακτηριστικά» για τον εντοπισμό ενός συγγραφέα ακόμα και στην περίπτωση μικρού δείγματος κειμένου [28]. Μέσω των *n-γραμμάτων*<sup>14</sup> υπολογίζεται η πιθανότητα ένας χαρακτήρας να εμφανίζεται στο περιβάλλον *n-1*. Τα *n-grams* αποτελούν ακολουθίες από *tokens* (= συνολικός αριθμός λέξεων/χαρακτήρων κτλ.). Στο Εικόνα 3. μπορούμε να δούμε το εργαλείο *Google n-grams*, μέσω του οποίου μπορούμε να εντοπίσουμε τη συχνότητα χρήσης κάποιων λέξεων σε ένα μεγάλο φάσμα κειμένων των *Google Books*.

---

<sup>14</sup> Βλ. περισσότερες πληροφορίες για τα *N-grams* στο:  
Jurafsky, D. (2000). *Speech & language processing*. Pearson Education India.



Εικόνα 3. Εμφάνιση των λέξεων «Ιλιάδα» και «Οδύσσεια» σε Αγγλικά κείμενα

Το  $n$  στον όρο  $n$ -gram υποδηλώνει τον αριθμό των στοιχείων που εξετάζονται κάθε φορά. Έτσι μπορούμε να συναντήσουμε μονό-γράμματα (*unigrams*) με  $n=1$ , δι-γράμματα (*bigrams*) με  $n=2$ , τρι-γράμματα (*trigrams*) με  $n=3$  κ.ο.κ. Προκειμένου να γίνει περισσότερο εύληπτη η επιμέρους ανάλυση ενός κειμένου σε  $n$ -γράμματα *χαρακτήρων*, παραθέτουμε ενδεικτικά την ανάλυση της ακόλουθης πρότασης από το προοίμιο της *Ιλιάδας* σε τρι-γράμματα *χαρακτήρων*:

- Πρόταση από τη ραψωδία Α της *Ιλιάδας*:

«Μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος οὐλομένην»

- Ανάλυση σε τρι-γράμματα *χαρακτήρων*:

[μῆν], [ῆνι], [νιν], [ιν\_], [ν\_ᾶ], [\_ᾶε], [ᾶει], [ειδ], [ιδε],  
[δε\_], [ε\_θ], [\_θε], [θεᾶ], [εᾶ\_], [ᾶ\_π], [\_πη], [πηλ], [ηλη],  
[ληϊ], [ηϊά], [ϊάδ], [άδε], [δεω], [εω\_], [ω\_ᾶ], [\_ᾶχ], [ᾶχι],  
[χιλ], [ιλῆ], [λῆο], [ῆος], [ος\_], [ς\_ο], [\_ού], [ούλ], [ύλο],  
[λομ], [ομέ], [μέν], [ένη], [νην]

Το *SLM* χρησιμοποιεί τους προηγούμενους  $n-1$  χαρακτήρες σε μία ακολουθία για να προβλέψει τον επόμενο χαρακτήρα. Η ιδανική τιμή  $n$  για τα δεδομένα της παρούσας έρευνας επιλέχθηκε να είναι  $n=3$ . Ο κατάλληλος αριθμός *χαρακτήρων* ( $n$ ) αποφασίστηκε δοκιμάζοντας *δι-γράμματα*, *τρι-γράμματα* και *τέτρα-γράμματα* *χαρακτήρων*. Παρατηρήσαμε ότι τα *τρι-γράμματα* *χαρακτήρων* πετύχαιναν την ιδανική



μέτρηση *Perplexity*<sup>15</sup> μεταξύ των υπολοίπων ( $n=2$  και  $n=4$ ) μετρήσεων. Ωστόσο, η παραπάνω μέθοδος ορισμένες φορές έχει πρακτικά προβλήματα, διότι απαιτείται ένας μεγάλος αριθμός προτάσεων εκπαίδευσης για να περιέχονται όλοι οι δυνατοί συνδυασμοί λέξεων στο περικείμενο.

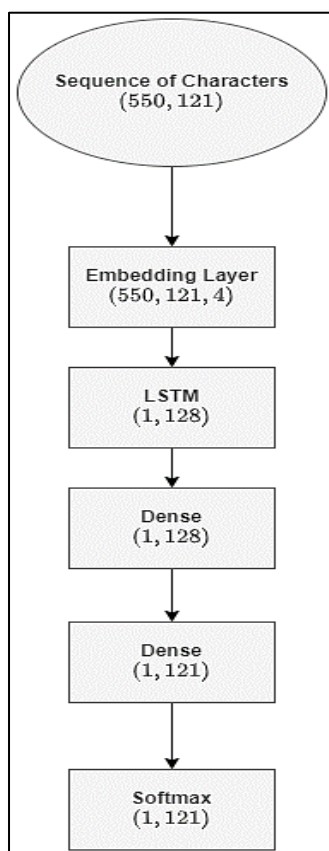
## 4.2 Νευρωνικό Γλωσσικό Μοντέλο Μακράς – Βραχείας Μνήμης (Long Short-Term Memory – LSTM)

Το πείραμα για την ταξινόμηση ομηρικών κειμένων μέσω στατιστικών γλωσσικών μοντέλων και νευρωνικών γλωσσικών μοντέλων μακράς – βραχείας μνήμης υλοποιήθηκε χρησιμοποιώντας, μεταξύ άλλων, και Νευρωνικά Γλωσσικά Μοντέλα *Long Short-Term Memory – LSTM*. Η εν λόγω κατηγορία γλωσσικών μοντέλων προέρχεται από τη γλωσσική μοντελοποίηση μέσω Νευρωνικών Δικτύων (*Neural Networks – NN*). Στις αρχές της δεκαετίας 2000 προτάθηκε ένα από τα πρώτα Νευρωνικά μοντέλα φυσικής γλώσσας [1]. Αυτό το δίκτυο ονομάστηκε *Neural Network Language Model (NNLM)*, αποτελείτο από 3 στρώματα και επρόκειτο για ένα απλό δίκτυο πρόσθιας τροφοδότησης (*feed – forward*) [26]. Λίγα χρόνια αργότερα από την παρουσίαση του *NNLM* προτάθηκε ένα Αναδρομικό Νευρωνικό Μοντέλο (*Recurrent Neural Network Language Model – RNNLM*) για τη γλωσσική μοντελοποίηση, το οποίο βελτίωσε σε μεγάλο βαθμό την απόδοση του *NNLM* [17]. Στα Αναδρομικά Νευρωνικά Μοντέλα (*Recurrent Neural Networks – RNN*) ανήκει η δεύτερη κατηγορία μοντέλου που χρησιμοποιήσαμε, το *LSTM*. Τα *LSTM* μοντέλα χρησιμοποιούνται σε εφαρμογές *NLP*, μία από τις οποίες είναι και η απόδοση συγγραφικής πατρότητας [20]. Επιλέξαμε να εξετάσουμε συγκριτικά την απόδοση του *SLM* με αυτή του *LSTM* στο δεύτερο πείραμα που εφαρμόσαμε, το οποίο αφορούσε στην ταξινόμηση ομηρικών κειμένων. Τα *LSTM* έχουν σημειώσει σημαντικά καλύτερες επιδόσεις σε σύγκριση με τα *SLM* και ένας παράγοντας γι' αυτό είναι η δυνατότητά τους να παράξουν μοντέλα από μεγαλύτερα περιβάλλοντα συμφραζομένων. Τα Νευρωνικά Γλωσσικά Μοντέλα λαμβάνουν υπόψη την ολότητα των προηγούμενων λέξεων και χαρακτήρων σε ένα κείμενο [17], σε αντίθεση με τα *SLM* που περιορίζονται σε ένα μικρό μήκος περιβάλλοντος για την εκτίμηση της πιθανότητας. Ακόμη, ένα από τα ουσιώδη πλεονεκτήματα του μοντέλου *LSTM* είναι η μεγάλη διάρκεια μνήμης που διαθέτει.

---

<sup>15</sup> Για τη μέτρηση *Perplexity* θα μιλήσουμε αναλυτικά στην Υποενότητα 4.4.1.

Η αρχιτεκτονική του  $LSTM^{16}$  γλωσσικού μοντέλου που υλοποιήθηκε στην παρούσα εργασία αποτυπώνεται στο Σχήμα 2. που ακολουθεί.



Σχήμα 2. Αρχιτεκτονική LSTM

### 4.3 Ανθρώπινη αξιολόγηση

Για την υλοποίηση του δεύτερου πειράματος που αφορούσε στη σύγκριση των δύο γλωσσικών μοντέλων ( $SLM$  και  $LSTM$ ) και ύστερα στη σύγκριση αυτών των μοντέλων με τις κρίσεις των ανθρώπων-επισημειωτών (*human-annotators*), δημιουργήσαμε ένα ερωτηματολόγιο (βλ. Παράρτημα), το οποίο απευθύνθηκε σε φοιτητές και αποφοίτους φιλολογικών τμημάτων. Ο αριθμός των ανθρώπων-επισημειωτών, που συμμετείχαν στο ερωτηματολόγιο ήταν 23. Συγκεκριμένα, συμμετείχαν 5 προπτυχιακοί φοιτητές Φιλολογίας, 3 απόφοιτοι Φιλολογίας, 10 μεταπτυχιακοί φοιτητές με προπτυχιακές σπουδές Φιλολογίας, 5 καθηγητές Φιλολόγοι σε Β/θμια εκπαίδευση. Σκοπός του ερωτηματολογίου ήταν να διερευνηθεί η ικανότητά

<sup>16</sup> Βλ. περισσότερες πληροφορίες για το  $LSTM$  στο: Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

των ανθρώπων-επισημειωτών να εντοπίζουν, χωρίς καμία αναζήτηση πηγών, όπως για παράδειγμα σε βιβλία ή στο διαδίκτυο, ορθά τα αποσπάσματα που αποδίδονται στον επικό ποιητή Όμηρο. Οι κρίσεις των τελευταίων αντιπαραβλήθηκαν με τις κρίσεις των *SLM* και *LSTM* μοντέλων για τα *Ιλιαδικά* και τα *Οδυσσειακά* αποσπάσματα που περιλαμβάνονταν στο ερωτηματολόγιο.

Τα αποσπάσματα του ερωτηματολογίου κυμαίνονταν από 580 έως 610 χαρακτήρες. Πιο συγκεκριμένα, 6 χωρία προέρχονταν από την *Ιλιάδα*, 6 από την *Οδύσσεια*, 2 από τον *Ομηρικό ύμνο «Εἰς Δημήτραν»*, 2 από τον *Ομηρικό ύμνο «Εἰς Ἀπόλλωνα»*, 2 από τον *Ομηρικό ύμνο «Εἰς Ἑρμῆν»* και 2 από τον *Ομηρικό ύμνο «Εἰς Ἀφροδίτην»*.

⋮

1. Ὡς φάτο, τὴν δ' Ὑπεριονίδης ἡμείβετο μῦθος· Πρῆς ἠὲ κόμου θυγάτηρ Δῆμιτερ ἄνασσα εἰδήσεις· διὲ γὰρ μέγα σ' ἄζομαι ἢ δ' ἑλεῖραι ἀχνυμένην περὶ παιδί τανυσφύρω· οὐδέ τις ἄλλος αἴτιος ἀθανάτων εἰ μὴ νεφεληγερέτα Ζεὺς, ὃς μιν ἔδωκε· Αἰδὴ θαλερὴν κεκλιθῆσθαι ἄκοιτιν αὐτοκασιγνήτω· ὃ δ' ὑπὸ ζόφῳ ἠερόεντα ἀρπάξας ἵπποισιν ἄγεν μεγάλα ἰάχουσιν· ἀλλὰ, θεά, κατάπαυε μέγαν γόνον· οὐδέ τί σε χρήμας αὐτῶς ἅπλητον ἔχειν χόλον· οὐ τοι ἀεικὴς γαμβρὸς ἐν ἀθανάτοισι Πολυσημάντωρ Αἰδωνεύς αὐτοκασιγνήτος· καὶ ὁμόσπορος· ἀμφὶ δὲ τιμὴν ἔλλαχεν ὥς τὰ πρῶτα διάτριχα δασμὸς ἐτύχθη τοῖς μεταναιετᾶσι τῶν Ἑλλαχε κοῖρανός ἐστιν. \*

☐ ΝΑΙ (ΙΛΙΑΔΑ)

☐ ΝΑΙ (ΟΔΥΣΣΕΙΑ)

☐ ΝΑΙ

☐ ΟΧΙ

☐ ἌΛΛΟ

Εικόνα 4. Ενδεικτικό παράδειγμα αποσπάσματος από τον Ομηρικό ύμνο «Εἰς Δημήτραν»

Τα 8 αποσπάσματα των *Ομηρικών ύμνων* δεν αξιολογήθηκαν με τα γλωσσικά μοντέλα διότι η χρησιμότητα αυτών ήταν να υπάρχει μία ποικιλία στη θεματική των αποσπασμάτων, πέραν της *Ιλιάδας* και της *Οδύσσειας*, που ο μέσος *επισημειωτής* φιλόλογος είναι περισσότερο εξοικειωμένος. Γι' αυτόν ακριβώς τον λόγο, επιλέξαμε και αποσπάσματα που φέρουν κοντινή στη γλώσσα των *Ομηρικών επών* διάλεκτο.

#### 4.4 Μετρήσεις αξιολόγησης

Για την αξιολόγηση του *SLM* και του *LSTM* χρησιμοποιήσαμε δύο μετρήσεις αξιολόγησης: η πρώτη μέτρηση ήταν η *περιπλοκότητα* (*Perplexity*) και η δεύτερη μέτρηση το *F-measure* (ή *F1-score*), η οποία χρησιμοποιήθηκε και για την απόδοση των ανθρώπων-επισημειωτών στα αποσπάσματα του ερωτηματολογίου.

#### 4.4.1 Περιπλοκότητα (Perplexity)

Η πιο συνηθισμένη μέτρηση αξιολόγησης ενός γλωσσικού μοντέλου είναι η *περιπλοκότητα* (*Perplexity*), η οποία ορίζεται σύμφωνα με αυτόν τον τύπο:

$$PP(p) := 2^{H(p)} = 2^{-\sum_x p(x) \log_2 p(x)}$$

*Εξίσωση 1. Τύπος περιπλοκότητας (Perplexity)*

όπου  $H(p)$  είναι η *εντροπία* [6] του μοντέλου.

Ένα αποδοτικό και ως εκ τούτου καλό γλωσσικό μοντέλο σημειώνει συνήθως μικρή *εντροπία*  $H$  και *περιπλοκότητα* που συνήθως προσεγγίζει τη μονάδα. Όσο πιο κακή είναι η απόδοση του γλωσσικού μοντέλου τόσο μεγαλύτερη συνηθίζεται να είναι η *εντροπία* και πολύ περισσότερο η *περιπλοκότητα* που σχετίζεται εκθετικά με την *εντροπία* [26]. Αφού υπολογίσαμε τη μέτρηση *Perplexity*, χρησιμοποιήσαμε μία στατιστική μέθοδο, που χρησιμοποιείται για τον προσδιορισμό του εύρους των πιθανών τιμών της εκτιμώμενης παραμέτρου αναφοράς, τα *Διαστήματα Εμπιστοσύνης* (*Confidence Intervals* ή *CI*). Τα *Διαστήματα Εμπιστοσύνης* χρησιμοποιούνται για την ασφαλέστερη εκτίμηση μίας παραμέτρου ενός πληθυσμού βάσει ενός τυχαίου δείγματος αυτού του πληθυσμού<sup>17</sup>. Τα *Διαστήματα Εμπιστοσύνης* υπολογίστηκαν για κάθε σειρά 20 τυχαίων αποσπασμάτων ανά κείμενο που βαθμολογήσαμε (βλ. αναλυτικά στο Κεφάλαιο 5). Αναλυτικότερα, υπολογίσαμε το μέσο όρο (*mean*) *Perplexity* κάθε σειράς 20 τυχαίων αποσπασμάτων καθώς και το τυπικό σφάλμα του μέσου όρου (*standard error of mean* ή *sem*), ορίσαμε επίπεδο εμπιστοσύνης (*confidence*) 95% και εν τέλει υπολογίσαμε το κάτω όριο (*lower*) και το άνω όριο (*upper*) των διαστημάτων εμπιστοσύνης.

Μέσω ενός παραδείγματος που θα περιγράψουμε, θα εξηγήσουμε τον λόγο για τον οποίο χρησιμοποιήσαμε τα *Διαστήματα Εμπιστοσύνης* στην παρούσα εργασία. Ας υποθέσουμε ότι θέλουμε να εξετάσουμε, αν η ραψωδία  $A$  ομοιάζει γλωσσικά στο σύνολο του έπους της *Ιλιάδας* ή όχι. Αρχικά, κάνουμε μία μηδενική υπόθεση (*null hypothesis*) ότι η ραψωδία  $A$  της *Ιλιάδας* δεν ανήκει στην *Ιλιάδα*, έχοντας σκοπό να αποδείξουμε το αντίθετο, δηλαδή ότι η ραψωδία  $A$  ανήκει στην *Ιλιάδα*. Το γλωσσικό μοντέλο της *Ιλιάδας* εκπαιδεύεται με τις υπόλοιπες 23 ραψωδίες, πλην της ραψωδίας

---

<sup>17</sup> Βλ. περισσότερες πληροφορίες για τα *Διαστήματα Εμπιστοσύνης* (*Confidence Intervals*) στο Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American statistical Association*, 82(397), 171-185.

*A*. Για να εξετάσουμε την εναπομείνασα ραψωδία *A* λαμβάνουμε τυχαία αποσπάσματα από τη ραψωδία αυτή, ώστε να βαθμολογήσουμε το γλωσσικό μοντέλο με τη μέτρηση *Perplexity*. Αφού λάβουμε τις τιμές *Perplexity* από τη βαθμολόγηση των τυχαίων αποσπασμάτων, υπολογίζουμε τον μέσο όρο γι' αυτές τις τιμές καθώς και το τυπικό σφάλμα του μέσου όρου, ορίζουμε επίπεδο εμπιστοσύνης 95% και υπολογίζουμε τα Διαστήματα Εμπιστοσύνης. Εν τέλει, απορρίπτουμε την αρχική μας υπόθεση (η ραψωδία *A* δεν ανήκει στην *Ιλιάδα*) ή δεχόμαστε το αντίστροφό της (η ραψωδία *A* ανήκει στην *Ιλιάδα*) σύμφωνα με το όριο εμπιστοσύνης που έχουμε ορίσει.

#### 4.4.2 F-measure (ή F1-score)

Η μέτρηση αξιολόγησης *F-measure* (ή *F1-score*) αποδεικνύεται χρήσιμη σε πολλές εφαρμογές ταξινόμησης στιγμιότυπων, όπου συχνά ερχόμαστε αντιμέτωποι με ένα πρόβλημα ταξινόμησης δύο κλάσεων<sup>18</sup> (π.χ 1<sup>η</sup> κλάση: 0 (=Οδύσσεια) και 2<sup>η</sup> κλάση: 1(=Ιλιάδα). Υπάρχουν τέσσερις περιπτώσεις στιγμιότυπων για την ταξινόμηση δύο κλάσεων, που περιγράφονται με τον πίνακα *Confusion matrix* (βλ. Πίνακας 4). Έτσι, σε ένα πρόβλημα δυαδικής ταξινόμησης, η επίδοση του ταξινομητή (*classifier*) όσον αφορά την ταξινόμηση των στιγμιότυπων ανάμεσα στις δύο κλάσεις, αξιολογείται βάσει κάποιων μετρήσεων αξιολόγησης, ανάμεσα στις οποίες βρίσκεται και το *F-measure* (ή *F1-score*).

	Αρνητικά: Ταξινομούνται Στην Κλάση 0	Θετικά: Ταξινομούνται Στην Κλάση 1
Ανήκουν στην Κλάση 0	True Negatives (TN)	False Positives (FP)
Ανήκουν στην Κλάση 1	False Negatives (FN)	True Positives (TP)

Πίνακας 4. *Confusion matrix* για ένα πρόβλημα ταξινόμησης δύο κλάσεων

Πηγή: Διαμαντάρας και Μπότσης, 2019, σ.36

Προκειμένου να ορίσουμε το *F-measure* (ή *F1-score*) θα αναφερθούμε σε δύο διαφορετικές μετρήσεις αξιολόγησης. Η πρώτη μέτρηση ονομάζεται *Ευστοχία* (*Precision*) και η δεύτερη *Ανάκληση* (*Recall*). Η *Ευστοχία* (*Precision*) δείχνει το ποσοστό από τα στιγμιότυπα που κατηγοριοποιούνται στην Κλάση 1 και ανήκουν πράγματι στην Κλάση 1 φερειπείν, ενώ η *Ανάκληση* (*Recall*) δείχνει το ποσοστό των στιγμιότυπων που ανήκουν στην Κλάση 1 και κατηγοριοποιούνται στην Κλάση 1 [26].

<sup>18</sup> Με τον όρο *κλάση* εννοείται μία συλλογή από πράγματα με κοινά χαρακτηριστικά και ιδιότητες.

Οι τιμές των δύο προαναφερθεισών μετρήσεων αξιολόγησης κυμαίνονται από 0 έως 1. Για να περιγραφεί εξ' ολοκλήρου η επίδοση ενός ταξινομητή, οι μετρήσεις αυτές συνδυάζονται με τη μέτρηση *F-measure* (ή *F1-score*), η οποία ορίζεται κατ' αυτόν τον τρόπο:

$$F - measure = \frac{Precision * Recall}{(Precision + Recall) / 2}$$

*Εξίσωση 2. Τύπος F-measure (ή F1-score)*

Η μέτρηση αξιολόγησης *F-measure* (ή *F1-score*) επίσης κυμαίνεται από 0 έως 1 και επιτυγχάνει τη μέγιστη τιμή (1) στην περίπτωση που ισχύει  $Precision = Recall = 1$ , το οποίο συμβαίνει στη περίπτωση που ταξινομηθούν όλα τα στιγμιότυπα ορθά [26].

## 5. Πειραματικά αποτελέσματα

Σε αυτό το κεφάλαιο θα περιγράψουμε τα πειράματα που υλοποιήσαμε και έπειτα θα παρουσιάσουμε την αξιολόγηση των πειραμάτων, βάσει της μεθοδολογίας που περιγράψαμε στο Κεφάλαιο 4.

### 5.1 Αναλυτική περιγραφή πειραμάτων

Στην παρούσα ενότητα θα περιγράψουμε αναλυτικά τα πειράματα που εφαρμόσαμε για την *απόδοση συγγραφικής πατρότητας στα ομηρικά έπη* και τους *Ομηρικούς ύμνους*. Στις υποενότητες 5.1.1 και 5.1.2, που ακολουθούν, θα δούμε τις δύο κατηγορίες πειραμάτων.

#### 5.1.1 Απόδοση συγγραφικής πατρότητας μέσω Στατιστικών Γλωσσικών Μοντέλων (SLM)

Το πρώτο πείραμα που εφαρμόσαμε στην παρούσα έρευνα αφορούσε στα *Στατιστικά Γλωσσικά Μοντέλα (SLM)*. Τα γλωσσικά μοντέλα για να εκπαιδευτούν χρειάζονται δεδομένα, τα οποία ονομάζονται δεδομένα εκπαίδευσης. Για να καθορίσουμε τα δεδομένα εκπαίδευσης, μετρήσαμε το μήκος χαρακτήρων κάθε ραψωδίας της *Ιλιάδας* και κάθε ραψωδίας της *Οδύσσειας*, έχοντας πρώτα εφαρμόσει την απαραίτητη επεξεργασία (βλ. Υποενότητα 3.2) στην *Ιλιάδα* και την *Οδύσσεια*. Επειδή οι ραψωδίες δεν είχαν ίσο μήκος χαρακτήρων (βλ. Υποενότητα 3.2), αποφασίσαμε να περικόψουμε χαρακτήρες από κάθε ραψωδία, ώστε όλες οι ραψωδίες, τόσο της *Ιλιάδας* όσο και της *Οδύσσειας*, να είναι ισομήκεις μεταξύ τους. Κατ' αυτόν τον τρόπο επιλέξαμε τα *δεδομένα εκπαίδευσης* και από την *Ιλιάδα* και από την *Οδύσσεια* να είναι οι πρώτοι 9.600 χαρακτήρες από κάθε ραψωδία, ώστε να αφαιρεθούν χαρακτήρες ακόμη και από τις μικρότερες σε μήκος χαρακτήρων ραψωδίες.

Υλοποιήθηκαν γλωσσικά μοντέλα *SLM*<sup>19</sup> για δύο κατηγορίες *SLM* (βλ. Πίνακας 3) με  $n=3$ . Ξεκινώντας με την πρώτη κατηγορία, δηλαδή με τα κυκλικά μοντέλα *SLM*, δημιουργήθηκαν 24 γλωσσικά μοντέλα επιπέδου χαρακτήρα για την *Ιλιάδα* (*IliadSLM's*) και 24 γλωσσικά μοντέλα αντίστοιχα για την *Οδύσσεια* (*OdysseySLM's*). Τα εν λόγω μοντέλα ακολούθησαν *κυκλικό τρόπο εκπαίδευσης (Cross validation)*

---

<sup>19</sup> Για το *SLM* που χρησιμοποιήθηκε βλ.

<https://github.com/ipavlopoulos/lm/blob/master/markov/models.py> [ημερομηνία πρόσβασης: 29 Νοεμβρίου 2020].

σύμφωνα με την εξής λογική: κάθε ένα από τα 24 γλωσσικά μοντέλα εκπαιδεύτηκε σε διαφορετικό υποσύνολο των ραψωδιών πλήθους 23 ραψωδιών και κάθε φορά βαθμολογήθηκε με την εναπομείνασα 24<sup>η</sup> ραψωδία του αντίστοιχου έπους (*Ιλιάδας* / *Οδύσσειας*).

---

**Αλγόριθμος 1: Εφαρμογή SLM σε *Ιλιάδα* και *Οδύσσεια***

---

*Ο αλγόριθμος μπορεί να γενικευτεί για οποιοδήποτε γλωσσικό μοντέλο, είτε στατιστικό είτε νευρωνικό.*

**Initialize** *Iliad\_models* as a list of 24 n-gram models where  $n = 3$

**Initialize** *Odyssey\_models* as a list of 24 n-gram models where  $n = 3$

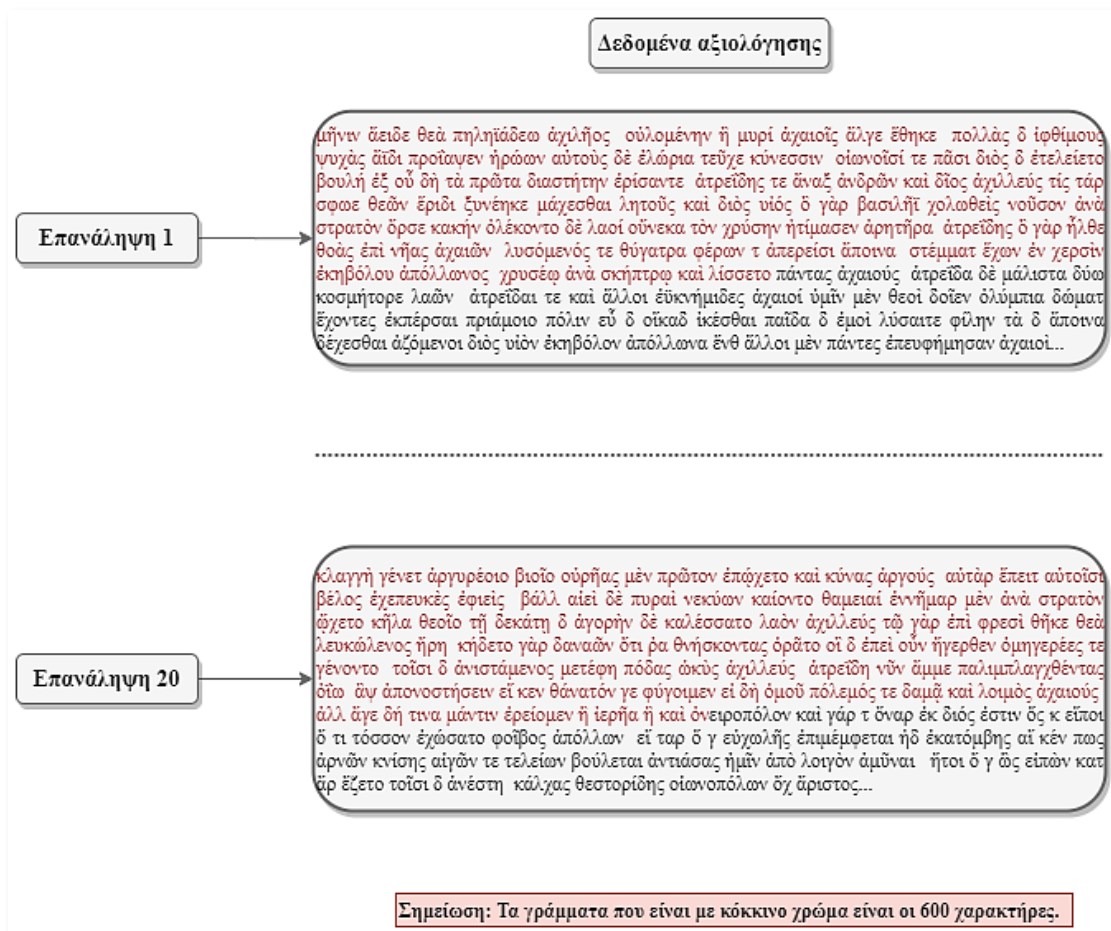
**Initialize** *Iliad\_train* as a list of all rhapsodies

**Initialize** *Odyssey\_train* as a list of all rhapsodies

1. **repeat** from  $m=0$  to  $m= 24$ ,  $step=1$
  2.     **repeat** from  $r=0$  to  $r= 24$ ,  $step=1$
  3.         **if**  $m$  is equal to  $r$ :
  4.             **continue**
  5.         **Train** *Iliad\_models*[ $m$ ] using samples from *Iliad\_train*[ $r$ ]
  6.         **Train** *Odyssey\_models*[ $m$ ] using samples from *Odyssey\_train*[ $r$ ]
- 

Για τη βαθμολόγηση των 24 *Ιλιαδικών μοντέλων* (*IliadSLM's*) και των 24 *Οδυσειακών μοντέλων* (*OdysseySLM's*) χρησιμοποιήθηκε η μέθοδος *Bootstrapping*. Πρόκειται για μία στατιστική τεχνική αναδειγματολειψίας (*Bootstrap statistical resampling technique*), που αναπαράγει ποικίλα σύνολα εκπαίδευσης, τα οποία χρησιμοποιούνται για την εκπαίδευση των μελών που συνιστούν τη συνδυαστική μέθοδο [4]. Έτσι, επιλέχθηκαν, μέσω της μεθόδου *Bootstrapping*, 20 τυχαία αποσπάσματα των 600 χαρακτήρων από το συνολικό μήκος χαρακτήρων κάθε ραψωδίας (βλ. Σχήμα 3).

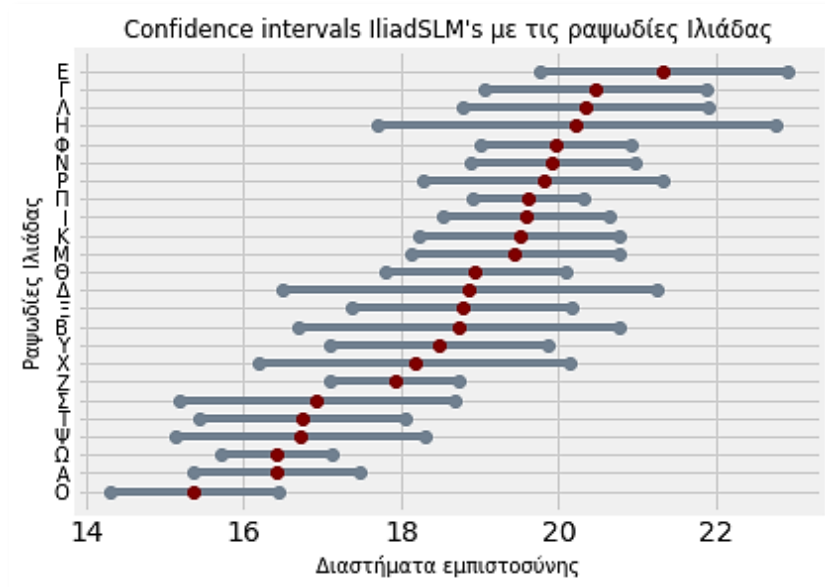




Σχήμα 3. Μέθοδος Bootstrapping για τα βαθμολογημένα δεδομένα του πειράματος «Απόδοση συγγραφικής πατρότητας μέσω Στατιστικών Γλωσσικών Μοντέλων (SLM)»

Το πρώτο βήμα ήταν να δημιουργηθεί ένας αλγόριθμος για την απόσπαση των 600 χαρακτήρων κάθε φορά και ύστερα αυτή η διαδικασία να επαναληφθεί 20 φορές, ώστε τελικά να έχουμε συνολικά 20 τυχαία αποσπάσματα από κάθε ραψωδία.

Ύστερα, τα αποσπάσματα αυτά βαθμολογήθηκαν με τη μέτρηση αξιολόγησης *Perplexity*. Στο διάγραμμα που ακολουθεί βλέπουμε τα Διαστήματα Εμπιστοσύνης από τα ποσοστά της μέτρησης *Perplexity* καθώς και τους μέσους όρους για τα Ιλιάδικα στατιστικά γλωσσικά μοντέλα (*IliadSLM's*) με βαθμολογημένα δεδομένα 20 τυχαία αποσπάσματα των 600 χαρακτήρων από την εναπομείνουσα κάθε φορά ραψωδία της *Ιλιάδας*. Η κόκκινη βούλα (•) σε κάθε γραμμή αποτυπώνει το μέσο όρο της μέτρησης *Perplexity* για τα βαθμολογημένα αποσπάσματα της εκάστοτε ραψωδίας.



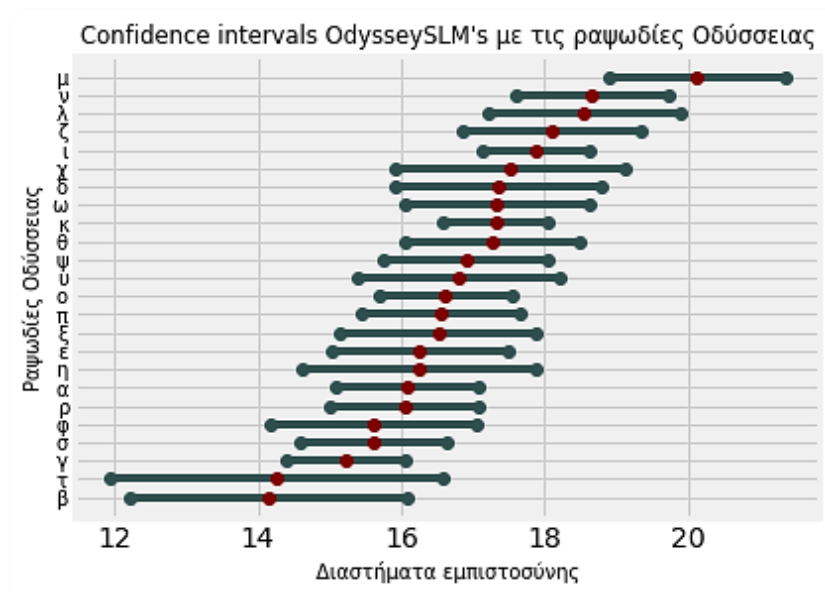
Διάγραμμα 13. *IliadSLM's* ανά ραψωδία *Ιλιάδας* με μοντέλα εκπαιδευμένα στις υπολειπόμενες 23 ραψωδίες της *Ιλιάδας* – κυκλική εκπαίδευση

Στο πιο πάνω διάγραμμα διαπιστώνουμε ότι η ραψωδία **Ο** της *Ιλιάδας*, που βαθμολογήθηκε με το *IliadSLM* ραψωδιών *Α-Ξ* και *Π-Ω* παρουσιάζει τη μικρότερη μέτρηση *Perplexity* με *Διαστήματα Εμπιστοσύνης* (14.30, 16.43) και μέσο όρο 15.37. Από την άλλη πλευρά, η ραψωδία που φαίνεται να παρεκκλίνει από το γλωσσικό ύφος της *Ιλιάδας* είναι η **Ε** με *Διαστήματα Εμπιστοσύνης* (19.76, 22.90) και μέσο όρο 21.32. Αναλυτικότερα, παραθέτουμε τα *Διαστήματα Εμπιστοσύνης* και τους μέσους όρους για όλες τις *Ιλιαδικές* ραψωδίες.

Ραψωδίες <i>Ιλιάδας</i>		CI lower	CI upper	Mean
0	Ο	14.299770	16.434077	15.366923
1	Α	15.354530	17.460495	16.407512
2	Ω	15.722314	17.110117	16.416216
3	Ψ	15.130209	18.310778	16.720493
4	Τ	15.435039	18.057280	16.746159
5	Σ	15.179634	18.679969	16.929801
6	Ζ	17.105616	18.724502	17.915059
7	Χ	16.183580	20.146054	18.164817
8	Υ	17.096202	19.871685	18.483943
9	Β	16.697362	20.771276	18.734319
10	Ξ	17.378657	20.166488	18.772573
11	Δ	16.495957	21.240275	18.868116
12	Θ	17.796020	20.082407	18.939213
13	Μ	18.133347	20.758543	19.445945
14	Κ	18.238413	20.756650	19.497532
15	Ι	18.517178	20.643811	19.580494
16	Π	18.912785	20.308790	19.610787
17	Ρ	18.279913	21.325358	19.802635
18	Ν	18.881109	20.967373	19.924241
19	Φ	18.998307	20.926327	19.962317
20	Η	17.691803	22.744251	20.218027
21	Λ	18.768618	21.905221	20.336920
22	Γ	19.052347	21.875799	20.464073
23	Ε	19.764017	22.895483	21.329750

Πίνακας 5. *Διαστήματα εμπιστοσύνης* και μέσοι όροι των ραψωδιών της *Ιλιάδας* με τα *IliadSLM's*

Στο ακόλουθο διάγραμμα μπορούμε να δούμε τα *Διαστήματα Εμπιστοσύνης* και τον *μέσο όρο* από τα ποσοστά της μέτρησης *Perplexity* για τα *Οδυσσειακά* στατιστικά γλωσσικά μοντέλα (*OdysseySLM's*) με βαθμολογημένα δεδομένα 20 τυχαία αποσπάσματα των 600 χαρακτήρων από την εναπομείναισα κάθε φορά ραψωδία της *Οδύσσειας*.



Διάγραμμα 14. *OdysseySLM's* ανά ραψωδία *Οδύσσειας* με μοντέλα εκπαιδευμένα στις υπολειπόμενες 23 ραψωδίες της *Οδύσσειας* – κυκλική εκπαίδευση

Σε αυτήν την κατηγορία, όπως μπορούμε να παρατηρήσουμε, η ραψωδία με τη μικρότερη μέτρηση *Perplexity* είναι η **β** της *Οδύσσειας* με *Διαστήματα Εμπιστοσύνης* (12.22, 16.08) και *μέσο όρο* 14.15. Η ραψωδία αυτή βαθμολογήθηκε με το *OdysseySLM* ραψωδιών *α* και *γ - ω*. Ενώ, η ραψωδία **μ** φαίνεται να παρεκκλίνει από το γλωσσικό ύφος της *Οδύσσειας* με *Διαστήματα Εμπιστοσύνης* (18.90, 21.36) και *μέσο όρο* 20.12. Ακολουθεί, ο πίνακας με τα *Διαστήματα Εμπιστοσύνης* και τους *μέσους όρους* όλων των βαθμολογημένων *Οδυσσειακών ραψωδιών*.

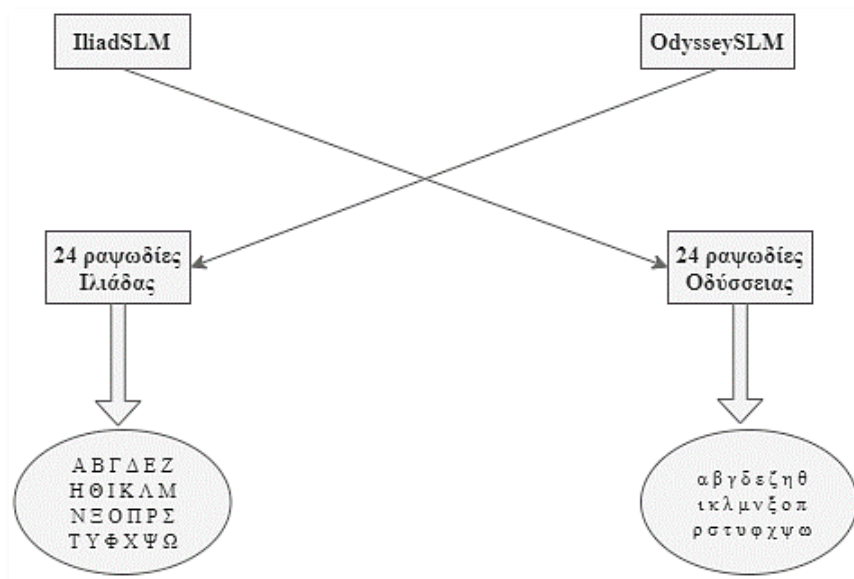
	Ραψωδίες Οδύσσειας	CI lower	CI upper	Mean
0	β	12.223103	16.076212	14.149658
1	τ	11.920747	16.583902	14.252324
2	γ	14.400627	16.041253	15.220940
3	σ	14.586299	16.635139	15.610719
4	φ	14.164126	17.057503	15.610815
5	ρ	15.011027	17.076340	16.043684
6	α	15.097956	17.070622	16.084289
7	η	14.627129	17.870444	16.248787
8	ε	15.018339	17.494983	16.256661
9	ξ	15.150343	17.885246	16.517795
10	π	15.446622	17.662985	16.554803
11	ο	15.681693	17.546580	16.614136
12	υ	15.398444	18.219246	16.808845
13	ψ	15.745613	18.054932	16.900272
14	θ	16.065231	18.490800	17.278015
15	κ	16.579618	18.048192	17.313905
16	ω	16.042658	18.628995	17.335827
17	δ	15.909186	18.780997	17.345092
18	χ	15.903742	19.116192	17.509967
19	ζ	17.124346	18.623099	17.873722
20	ς	16.863804	19.353861	18.108832
21	λ	17.210101	19.881259	18.545680
22	ν	17.585948	19.719002	18.652475
23	μ	18.891708	21.363748	20.127728

Πίνακας 6. Διαστήματα εμπιστοσύνης και μέσοι όροι των ραψωδιών της Οδύσσειας με τα *OdysseySLM*'s

Για τη δεύτερη κατηγορία γλωσσικών μοντέλων *SLM*, δηλαδή τα χιαστί μοντέλα *SLM* (βλ. Πίνακας 3) υλοποιήθηκαν δύο γλωσσικά μοντέλα επιπέδου χαρακτήρα, ένα για την *Ιλιάδα* (*IliadSLM*) και ένα για την *Οδύσσεια* (*OdysseySLM*). Το *Ιλιαδικό μοντέλο* (*IliadSLM*) εκπαιδεύτηκε σε όλες τις ραψωδίες της *Ιλιάδας*, ακολουθώντας και σε αυτήν την περίπτωση την απόσπαση 9.600 χαρακτήρων από την αρχή κάθε ραψωδίας της *Ιλιάδας*. Αντίστοιχα, το *Οδυσειακό μοντέλο* (*OdysseySLM*) εκπαιδεύτηκε σε όλες τις ραψωδίες της *Οδύσσειας*, δηλαδή με 9.600 χαρακτήρες από την αρχή κάθε ραψωδίας της *Οδύσσειας*. Συνολικά, το κάθε ένα γλωσσικό μοντέλο (*IliadSLM* και *OdysseySLM*) εκπαιδεύτηκε σε 230.400 χαρακτήρες. Τόσο το *IliadSLM* όσο και το *OdysseySLM* βαθμολογήθηκαν χιαστί. Πιο συγκεκριμένα, το *IliadSLM* βαθμολογήθηκε με κάθε ραψωδία της *Οδύσσειας*<sup>20</sup> και το *OdysseySLM* με κάθε ραψωδία της *Ιλιάδας*<sup>21</sup> (βλ. Σχήμα 4).

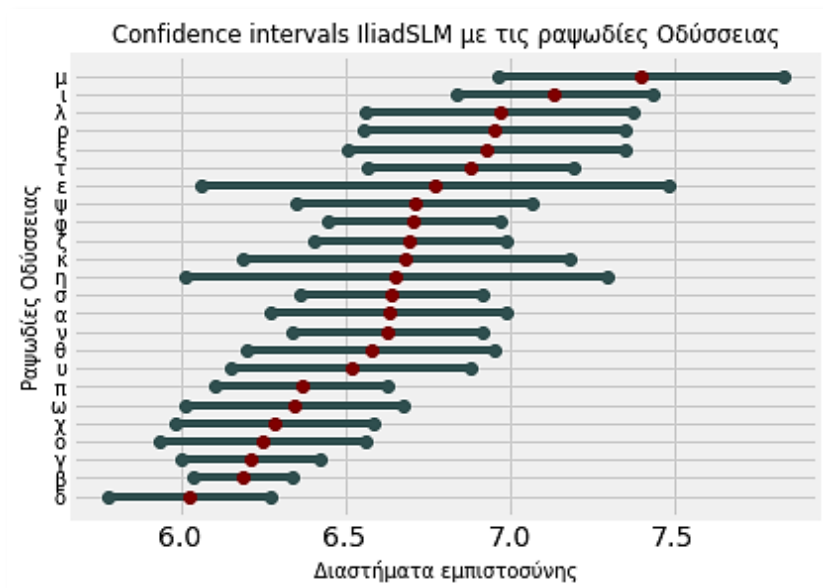
<sup>20</sup> Οι ραψωδίες της *Οδύσσειας* συμβολίζονται με πεζά γράμματα.

<sup>21</sup> Οι ραψωδίες της *Ιλιάδας* συμβολίζονται με κεφαλαία γράμματα.



Σχήμα 4. Χιαστί βαθμολόγηση στην Ιλιάδα και την Οδύσσεια

Τα βαθμολογημένα δεδομένα από κάθε ραψωδία της *Ιλιάδας* και της *Οδύσσειας* ήταν και σε αυτήν την περίπτωση 20 τυχαία αποσπάσματα των 600 χαρακτήρων. Στο διάγραμμα που ακολουθεί παρουσιάζονται τα *Διαστήματα Εμπιστοσύνης* από τα ποσοστά της μέτρησης *Perplexity* καθώς και οι μέσοι όροι για το *IliadSLM* με βαθμολογημένα δεδομένα 20 τυχαία αποσπάσματα των 600 χαρακτήρων από κάθε ραψωδία της *Οδύσσειας*.



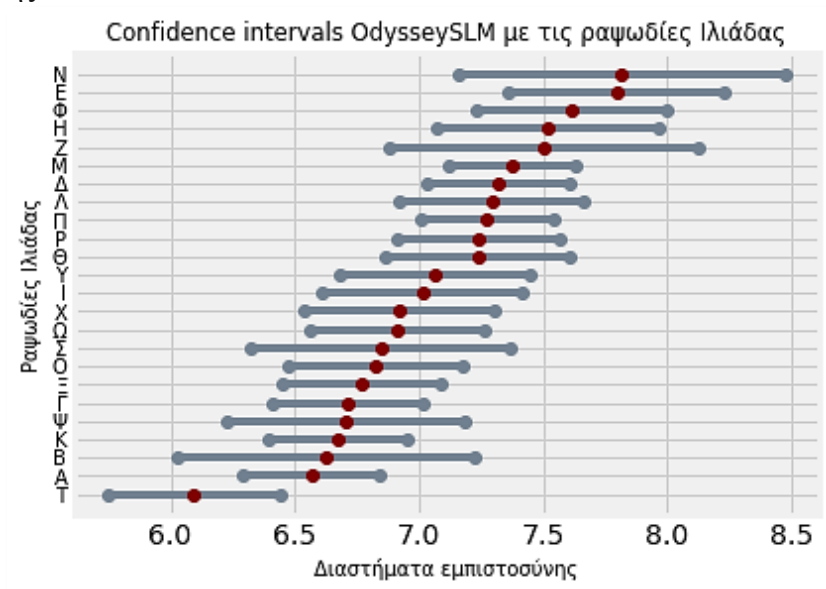
Διάγραμμα 15. *IliadSLM* ανά ραψωδία Οδύσσειας με μοντέλο εκπαιδευμένο σε όλες τις ραψωδίες της *Ιλιάδας* (9.600 χαρακτήρες από κάθε ραψωδία)

Στην κατηγορία αυτή, όπως γίνεται φανερό από το παραπάνω διάγραμμα, η ραψωδία **δ** της *Οδύσσειας* παρουσιάζει τη μικρότερη μέτρηση *Perplexity* με *Διαστήματα Εμπιστοσύνης* (5.77, 6.27) και μέσο όρο 6.02. Από την άλλη πλευρά, η ραψωδία **μ** της *Οδύσσειας* φαίνεται να παρεκκλίνει από το γλωσσικό ύφος της *Ιλιάδας* με *Διαστήματα Εμπιστοσύνης* (6.96, 7.83) και μέσο όρο 7.40. Παρακάτω, μπορούμε να δούμε τα *Διαστήματα Εμπιστοσύνης* και τους μέσους όρους όλων των ραψωδιών της *Οδύσσειας* με το *IliadSLM*.

Ραψωδίες Οδύσσειας		CI lower	CI upper	Mean
0	δ	5.777117	6.272138	6.024628
1	β	6.041329	6.337696	6.189513
2	γ	6.002012	6.424557	6.213285
3	ο	5.938457	6.562239	6.250348
4	χ	5.986607	6.587243	6.286925
5	ω	6.013532	6.677419	6.345476
6	π	6.107404	6.629920	6.368662
7	υ	6.155834	6.880704	6.518269
8	θ	6.202380	6.956665	6.579522
9	ν	6.341672	6.917933	6.629802
10	α	6.276804	6.992174	6.634489
11	σ	6.361833	6.915834	6.638834
12	η	6.012204	7.297665	6.654935
13	κ	6.190197	7.181659	6.685928
14	ζ	6.404421	6.991741	6.698081
15	φ	6.446198	6.973588	6.709893
16	ψ	6.352499	7.069827	6.711163
17	ε	6.062398	7.484997	6.773698
18	τ	6.571970	7.197212	6.884591
19	ξ	6.510114	7.351516	6.930815
20	ρ	6.557264	7.352656	6.954960
21	λ	6.561200	7.379025	6.970112
22	ι	6.841000	7.434599	7.137800
23	μ	6.963743	7.835524	7.399633

Πίνακας 7. Διαστήματα εμπιστοσύνης και μέσοι όροι των ραψωδιών της *Οδύσσειας* με το *IliadSLM*

Στο επόμενο διάγραμμα παρουσιάζονται τα *Διαστήματα Εμπιστοσύνης* και οι μέσοι όροι από τα ποσοστά της μέτρησης *Perplexity* για το *OdysseySLM* με βαθμολογημένα δεδομένα 20 τυχαία αποσπάσματα των 600 χαρακτήρων από κάθε ραψωδία της *Ιλιάδας*.



Διάγραμμα 16. *OdysseySLM* ανά ραψωδία *Ιλιάδας* με μοντέλο εκπαιδευμένο σε όλες τις ραψωδίες της *Οδύσσειας* (9.600 χαρακτήρες από κάθε ραψωδία)



Όπως γίνεται φανερό από το πιο πάνω διάγραμμα, η ραψωδία **T** της *Ιλιάδας* παρουσιάζει τη μικρότερη μέτρηση *Perplexity* με *Διαστήματα Εμπιστοσύνης* (**5.74, 6.44**) και μέσο όρο **6.09**. Ενώ, η ραψωδία που φαίνεται να παρεκκλίνει γλωσσικά από το ύφος της *Οδύσσειας* είναι η **N** της *Ιλιάδας* με *Διαστήματα Εμπιστοσύνης* (**7.15, 8.47**) και μέσο όρο **7.81**. Πιο κάτω μπορούμε να δούμε τα *Διαστήματα Εμπιστοσύνης* καθώς και τους μέσους όρους όλων των ραψωδιών της *Ιλιάδας* με το *OdysseySLM*.

	Ραψωδίες Ιλιάδας	CI lower	CI upper	Mean
0	T	5.747851	6.441432	6.094641
1	A	6.294364	6.841072	6.567718
2	B	6.026335	7.226988	6.626661
3	K	6.391262	6.952845	6.672054
4	Ψ	6.228821	7.180338	6.704579
5	Γ	6.412954	7.015177	6.714066
6	Ξ	6.448063	7.084885	6.766474
7	O	6.477618	7.179081	6.828350
8	Σ	6.323805	7.370781	6.847293
9	Ω	6.560278	7.262823	6.911551
10	X	6.534634	7.306831	6.920732
11	I	6.609295	7.418676	7.013985
12	Y	6.681905	7.448927	7.065416
13	Θ	6.865119	7.609655	7.237387
14	P	6.916161	7.565103	7.240632
15	Π	7.008989	7.540253	7.274621
16	Λ	6.923290	7.664805	7.294048
17	Δ	7.030633	7.603761	7.317197
18	M	7.121680	7.631020	7.376350
19	Z	6.879814	8.120959	7.500386
20	H	7.069681	7.965399	7.517540
21	Φ	7.230663	7.996327	7.613495
22	E	7.361692	8.229047	7.795369
23	N	7.157534	8.472324	7.814929

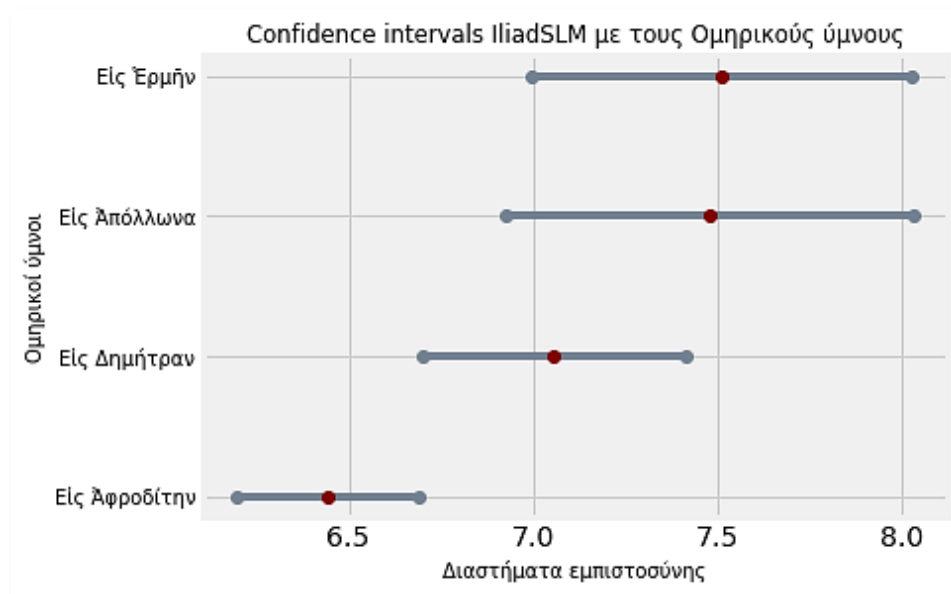
Πίνακας 8. Διαστήματα εμπιστοσύνης και μέσοι όροι των ραψωδιών της *Ιλιάδας* με το *OdysseySLM*

Ακολουθώς η δεύτερη κατηγορία *SLM* ή χιαστί μοντέλο *SLM*, δηλαδή το *IliadSLM* και το *OdysseySLM*, βαθμολογήθηκε και με τους τέσσερις *Ομηρικούς ύμνους*: «Εἰς Ἀπόλλωνα», «Εἰς Ἀφροδίτην», «Εἰς Δημήτραν» και «Εἰς Ἑρμῆν». Στον πίνακα που ακολουθεί μπορούμε να δούμε τα μήκη των *Ομηρικών ύμνων* σε χαρακτήρες.

Ομηρικός ύμνος	Μήκος ύμνου σε χαρακτήρες
Εἰς Ἀπόλλωνα	23036
Εἰς Ἀφροδίτην	12371
Εἰς Δημήτραν	20826
Εἰς Ἑρμῆν	24637

Πίνακας 9. Μήκος σε χαρακτήρες στους *Ομηρικών ύμνων*

Το *IliadSLM* και το *OdysseySLM* βαθμολογήθηκαν με 20 τυχαία αποσπάσματα των 600 χαρακτήρων από κάθε έναν *Ομηρικό ύμνο*. Στα διαγράμματα και τους πίνακες που ακολουθούν αποτυπώνονται τα αποτελέσματα του *IliadSLM* και του *OdysseySLM* με τους *Ομηρικούς ύμνους*. Συγκεκριμένα, παρουσιάζονται, αρχικά, τα αποτελέσματα του *IliadSLM* με τους *Ομηρικούς ύμνους* και ύστερα του *OdysseySLM* με τους *Ομηρικούς ύμνους*. Κάθε γραμμή είναι τα *Διαστήματα Εμπιστοσύνης* από τα ποσοστά της μέτρησης *Perplexity* κάθε *Ομηρικού ύμνου* και η κόκκινη βούλα (•) αποτυπώνει το μέσο όρο της μέτρησης *Perplexity*.



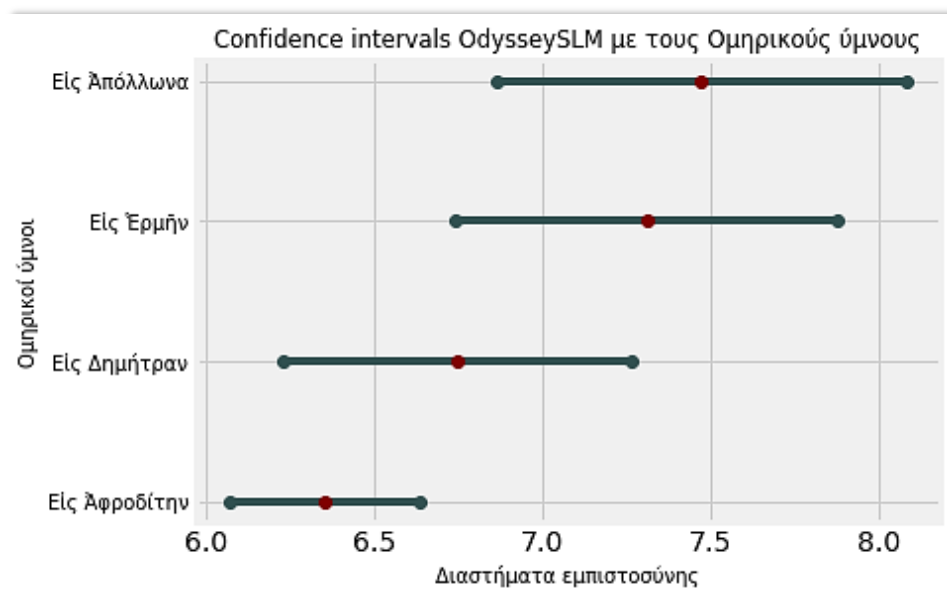
Διάγραμμα 17. *IliadSLM* ανά *Ομηρικό ύμνο* με μοντέλο εκπαιδευμένο σε όλες τις ραψωδίες της *Ιλιάδας* (9.600 χαρακτήρες από κάθε ραψωδία)

Στο πιο πάνω διάγραμμα φαίνεται ο *Ομηρικός ύμνος* «*Εἰς Ἀφροδίτην*» να παρουσιάζει τη μικρότερη μέτρηση *Perplexity* με *Διαστήματα Εμπιστοσύνης* (6.19, 6.69) και μέσο όρο 6.44. Ενώ, ο *Ομηρικός ύμνος* «*Εἰς Ἑρμῆν*» φαίνεται να παρεκκλίνει από το γλωσσικό ύφος της *Ιλιάδας* με *Διαστήματα Εμπιστοσύνης* (6.99, 8.02) και μέσο όρο 7.51. Στον πίνακα που ακολουθεί μπορούμε να δούμε αναλυτικά τα *Διαστήματα Εμπιστοσύνης* και τους μέσους όρους για τους τέσσερις *Ομηρικούς ύμνους* με το *IliadSLM*.

Ομηρικοί ύμνοι	CI lower	CI upper	Mean
0 Εἰς Ἀφροδίτην	6.196133	6.690965	6.443549
1 Εἰς Δημήτραν	6.697965	7.415069	7.056517
2 Εἰς Ἀπόλλωνα	6.927678	8.028041	7.477859
3 Εἰς Ἑρμῆν	6.993205	8.027390	7.510297

Πίνακας 10. *Διαστήματα εμπιστοσύνης και μέσοι όροι των Ομηρικών ύμνων με το IliadSLM*





Διάγραμμα 18. *OdysseySLM* ανά Ομηρικό ὕμνο με μοντέλο εκπαιδευμένο σε όλες τις ραψωδίες της Οδύσσειας (9.600 χαρακτήρες από κάθε ραψωδία)

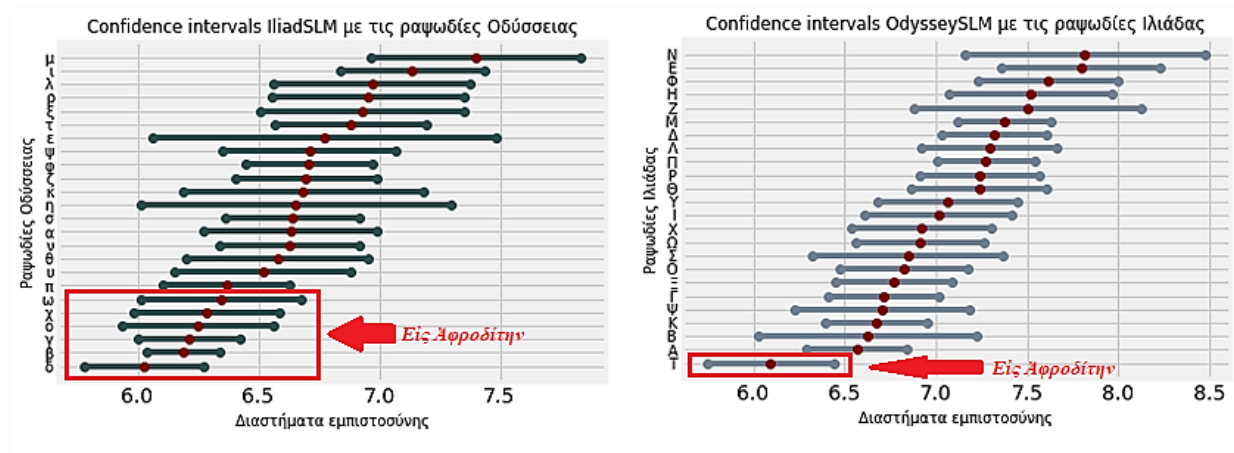
Ομοίως και στην περίπτωση του *OdysseySLM*, όπως βλέπουμε στο πιο πάνω διάγραμμα, ο Ομηρικός ὕμνος «*Εἰς Ἀφροδίτην*» παρουσιάζει τη μικρότερη μέτρηση *Perplexity* με Διαστήματα Εμπιστοσύνης (6.06, 6.63) και μέσο ὄρο 6.35. Από την άλλη πλευρά, ο Ομηρικός ὕμνος «*Εἰς Ἀπόλλωνα*» φαίνεται να παρεκκλίνει από το γλωσσικό ὕφος της Οδύσσειας με Διαστήματα Εμπιστοσύνης (6.86, 8.07) και μέσο ὄρο 7.46. Στον πίνακα που ακολουθεί αποτυπώνονται τα Διαστήματα Εμπιστοσύνης καθώς και οι μέσοι ὄροι για τους τέσσερις Ομηρικούς ὕμνους με το *OdysseySLM*.

Ομηρικοί ὕμνοι	CI lower	CI upper	Mean
0 Εἰς Ἀφροδίτην	6.067164	6.632857	6.350011
1 Εἰς Δημήτῤαν	6.224485	7.261550	6.743017
2 Εἰς Ἑρμῆν	6.740469	7.874017	7.307243
3 Εἰς Ἀπόλλωνα	6.861130	8.078110	7.469620

Πίνακας 11. Διαστήματα εμπιστοσύνης και μέσοι ὄροι των Ομηρικών ὕμνων με το *OdysseySLM*

Εἰς ὅτι αφορά την αξιολόγηση των Ομηρικών ὕμνων με το *IliadSLM* και το *OdysseySLM* γίνεται εύκολα αντιληπτό, από τα πιο πάνω διαγράμματα, πώς ο Ομηρικός ὕμνος «*Εἰς Ἀφροδίτην*» φαίνεται να συγγενεύει περισσότερο γλωσσικά με τα δύο ομηρικά ἔπη απ' ὅτι οι υπόλοιποι τρεῖς ὕμνοι. Ακόμα, παρατηρούμε πώς ο μέσος ὄρος *Perplexity* του ὕμνου βρίσκεται αρκετά κοντά με εκείνους των ραψωδιών δ, β, γ, ο, χ και ω της Οδύσσειας, όταν βαθμολογούνται με το *IliadSLM* και με τη ραψωδία Τ της *Ιλιάδας*, όταν βαθμολογείται με το *OdysseySLM* (βλ. Διάγραμμα 19). Στο σημείο αυτό

αξίζει να αναφερθεί ότι ο ύμνος «*Εἰς Ἀφροδίτην*» θεωρείται συχνά από τους μελετητές ο πιο Ομηρικός από όλους τους άλλους ύμνους, διότι βρίσκεται κοντά στα *ομηρικά έπη* από πλευράς ποιητικής γλώσσας, ύφους και θεματικής [18].



Διάγραμμα 19. Θέση του Ομηρικού ύμνου *Εἰς Ἀφροδίτην* συγκριτικά με τα *Ομηρικά έπη*

Ολοκληρώνοντας την παρουσίαση των αποτελεσμάτων του πειράματος για την απόδοση συγγραφικής πατρότητας μέσω *SLM*, πρέπει να αναφερθεί ότι, αν επαναλάβουμε το πείραμα, λόγω του ότι κάθε φορά βαθμολογούνται τυχαία δείγματα κειμένου με τη μέθοδο *Bootstrapping*, η κατάταξη των ραψωδιών τροποποιείται ελαφρώς. Μία ερμηνεία θα μπορούσε να είναι ότι η εκάστοτε αλληλουχία των 600 χαρακτήρων που λαμβάνεται κάθε φορά φαίνεται να επηρεάζει τη γλωσσική ομοιότητα των αποσπασμάτων της *X* ραψωδίας με το σύνολο των δεδομένων εκπαίδευσης του *Y* γλωσσικού μοντέλου. Ωστόσο, η κατάταξη των *Ομηρικών ύμνων* φαίνεται να παραμένει σταθερή, όσες φορές και αν επαναληφθεί το πείραμα. Η ερμηνεία που μπορεί να δοθεί σε αυτό είναι ότι για τη βαθμολόγηση των *Ομηρικών ύμνων* χρησιμοποιήθηκαν τα ίδια χιαστί γλωσσικά μοντέλα *SLM*, ενώ για τις ραψωδίες της *Ιλιάδας* και της *Οδύσσειας* — αναφορικά με το κυκλικό μοντέλο *SLM* — χρησιμοποιήθηκαν γλωσσικά μοντέλα που εκπαιδεύτηκαν σε διαφορετικά δεδομένα εκπαίδευσης. Συνεπώς, φαίνεται ότι η μέτρηση *Perplexity* εξαρτάται και από το σύνολο δεδομένων εκπαίδευσης και πιθανώς ένα ίδιο σύνολο εκπαίδευσης να φέρει πιο σταθερά αποτελέσματα. Πιο αναλυτικά, τα *IliadSLM*'s, όταν βαθμολογούνται με τις ραψωδίες της *Ιλιάδας*, σε μία επανάληψη δέκα φορές του πειράματος, φέρνουν πρώτες<sup>22</sup> τις ραψωδίες: **Ο** και **Χ**. Τα *OdysseySLM*'s, όταν βαθμολογούνται με τις

<sup>22</sup> Σε αυτό το σημείο με τη λέξη «πρώτος» εννοούμε ότι εμφανίζεται μεγαλύτερη γλωσσική ομοιότητα.

ραψωδίες της *Οδύσσειας*, τις ραψωδίες: **β**, **γ**, και **τ**. Ακολούθως, το *IliadSLM*, όταν βαθμολογείται με τις ραψωδίες της *Οδύσσειας*, φέρνει πρώτες τις ραψωδίες: **γ**, **δ**, και **ο**. Αντίστοιχα το *OdysseySLM*, όταν βαθμολογείται με τις ραψωδίες της *Ιλιάδας*, τις ραψωδίες: **Τ** και **Κ**. Ωστόσο, κατά την επανάληψη δέκα φορές του πειράματος με τους *Ομηρικούς ύμνους*, ο ύμνος «*Εἰς Ἀφροδίτην*» βρίσκεται πάντα στην πρώτη θέση μεταξύ των άλλων ύμνων.

### **5.1.2 Ταξινόμηση ομηρικών κειμένων μέσω Στατιστικών Γλωσσικών Μοντέλων (SLM) και Νευρωνικών Γλωσσικών Μοντέλων Μακράς - Βραχείας Μνήμης (LSTM)**

Η υλοποίηση του δεύτερου πειράματος αφορούσε στη σύγκριση *Στατιστικών Γλωσσικών Μοντέλων (SLM)*, *Νευρωνικών Γλωσσικών Μοντέλων Μακράς - Βραχείας Μνήμης (Long Short-Term Memory - LSTM)* και ανθρώπων-επισημειωτών (*human-annotators*). Οι κρίσεις των ανθρώπων-επισημειωτών συλλέχθηκαν μέσω ερωτηματολογίου (βλ. Υποενότητα 4.3). και αντιπαραβλήθηκαν με τις κρίσεις των *SLM* και *LSTM* γλωσσικών μοντέλων για τα *Ιλιαδικά* και *Οδυσσειακά* αποσπάσματα του ερωτηματολογίου. Στόχος αυτού του πειράματος ήταν να αξιολογηθεί η απόδοση των γλωσσικών μοντέλων προκειμένου να διερευνηθεί η ικανότητά ανίχνευσής του ποσοστού, κατά το οποίο τα δοθέντα αποσπάσματα ανήκουν ή όχι στην *Ιλιάδα* και αντίστοιχα στην *Οδύσσεια*.

Η λογική πάνω στην οποία βασίστηκε το εν λόγω πείραμα ήταν να δημιουργηθούν δύο γλωσσικά μοντέλα, ένα για την *Ιλιάδα* και ένα για την *Οδύσσεια*. Το *Ιλιαδικό γλωσσικό μοντέλο*, ομοίως με το πείραμα για την απόδοση συγγραφικής πατρότητας μέσω *SLM*, εκπαιδεύτηκε στους πρώτους 9.600 χαρακτήρες από κάθε ραψωδία της *Ιλιάδας* και αντίστοιχα το *Οδυσσειακό μοντέλο* στους πρώτους 9.600 χαρακτήρες από κάθε ραψωδία της *Οδύσσειας*. Στόχος της εν λόγω εκπαίδευσης ήταν να χρησιμοποιηθούν γλωσσικά μοντέλα που έχουν εκπαιδευτεί στο εκάστοτε έπος (*Ιλιάδα/Οδύσσεια*), ώστε να αναλυθούν και υπολογιστικά τα περισσότερα αποσπάσματα που απευθύνθηκαν στους ανθρώπους-επισημειωτές μέσω του ερωτηματολογίου. Στο σημείο αυτό πρέπει να αναφερθεί ότι αφαιρέθηκαν από τα δεδομένα εκπαίδευσης των γλωσσικών μοντέλων οι ραψωδίες εκείνες, αποσπάσματα των οποίων εμπεριέχονταν στο ερωτηματολόγιο. Αναλυτικότερα, από την *Ιλιάδα*

αφαιρέθηκαν οι ραψωδίες:  $\Gamma$ ,  $\Theta$ ,  $K$ ,  $\Lambda$ ,  $O$  και  $\Phi$ , ενώ από την *Οδύσσεια* οι ραψωδίες:  $\gamma$ ,  $\zeta$ ,  $\kappa$ ,  $\nu$ ,  $\sigma$  και  $\varphi$ .

Κατ’ αυτόν τον τρόπο υλοποιήθηκαν δύο κατηγορίες γλωσσικών μοντέλων (βλ. Πίνακας 3): η πρώτη αφορούσε στην ταξινόμηση ομηρικών κειμένων βάσει *SLM* και η δεύτερη στην ταξινόμηση ομηρικών κειμένων βάσει *LSTM*. Η πρώτη κατηγορία αποτελούταν από ένα *Ιλιαδικό μοντέλο IliadSLM* και ένα *Οδυσσειακό μοντέλο OdysseySLM* και η δεύτερη κατηγορία αποτελούταν από ένα *IliadLSTM* και ένα *OdysseyLSTM*. Όλα τα γλωσσικά μοντέλα και των δύο κατηγοριών αξιολογήθηκαν μέσω της μέτρησης *Perplexity* με τα αποσπάσματα του ερωτηματολογίου της *Ιλιάδας* και της *Οδύσσειας*.

Προκειμένου να αξιολογήσουμε την απόδοση των γλωσσικών μοντέλων, ώστε να διερευνηθεί η ικανότητα ανίχνευσής του ποσοστού κατά το οποίο τα δοθέντα αποσπάσματα ανήκουν ή όχι στην *Ιλιάδα* και αντίστοιχα στην *Οδύσσεια*, κατασκευάσαμε έναν *ταξινομητή (classifier)*. Η λειτουργία αυτού του ταξινομητή ήταν να κατατάσσει κάθε ένα από τα αποσπάσματα στην *Ιλιάδα* ή την *Οδύσσεια*, με κριτήριο τις μετρήσεις *Perplexity* από τα αντίστοιχα γλωσσικά μοντέλα. Έτσι, ορίσαμε δύο κλάσεις: “0” για την κλάση της *Οδύσσειας* και “1” για την κλάση της *Ιλιάδας*.

---

**Αλγόριθμος 2: Δυαδικός ταξινομητής (Binary classifier)**

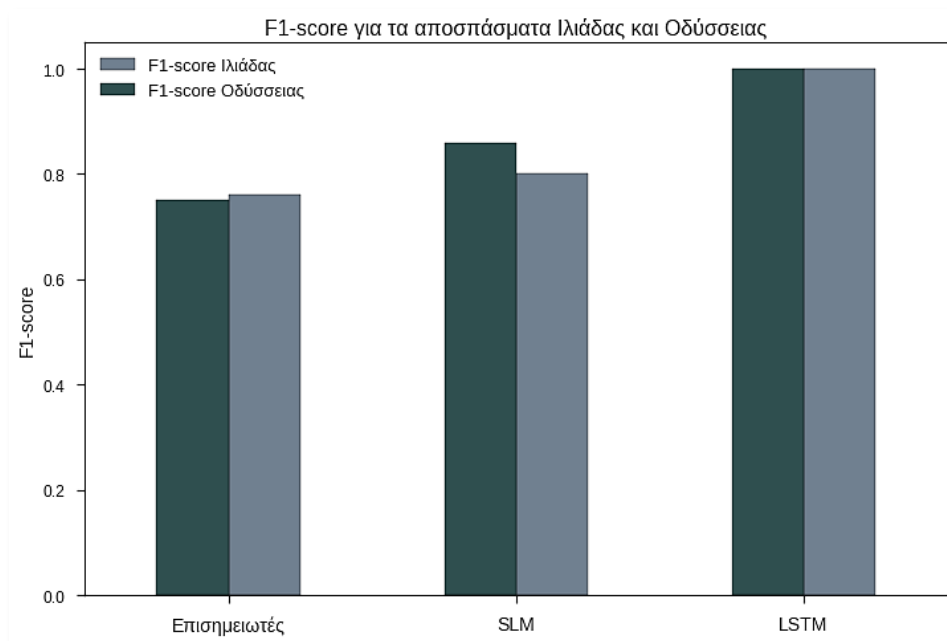
---

Η συνάρτηση αυτή επιστρέφει μία ετικέτα 0 ή 1, ανάλογα την κλάση που προβλέφθηκε ότι ανήκει το δοθέν απόσπασμα.

```
1. Function classify(Iliad_model, Odyssey_model, text):  
2.   Set PPL_O equal to Perplexity(Odyssey_model, text)  
3.   Set PPL_I equal to Perplexity(Iliad_model, text)  
4.   if PPL_O is less than PPL_I:  
5.     return “0”  
6.   else  
7.     return “1”
```

---

Η επίδοση του *ταξινομητή* αξιολογήθηκε με τη μέτρηση *F1-score* (βλ. Υποενότητα 4.4.2) για τα αποσπάσματα της *Ιλιάδας* και της *Οδύσσειας* που περιλαμβάνονταν στο ερωτηματολόγιο. Παραλλήλως αξιολογήθηκαν και οι κρίσεις των *επισημειωτών* με την ίδια μέτρηση αξιολόγησης για τα αποσπάσματα αυτά. Στο διάγραμμα που ακολουθεί αποτυπώνεται η μέτρηση *F1-score* των γλωσσικών συστημάτων και των *επισημειωτών*.



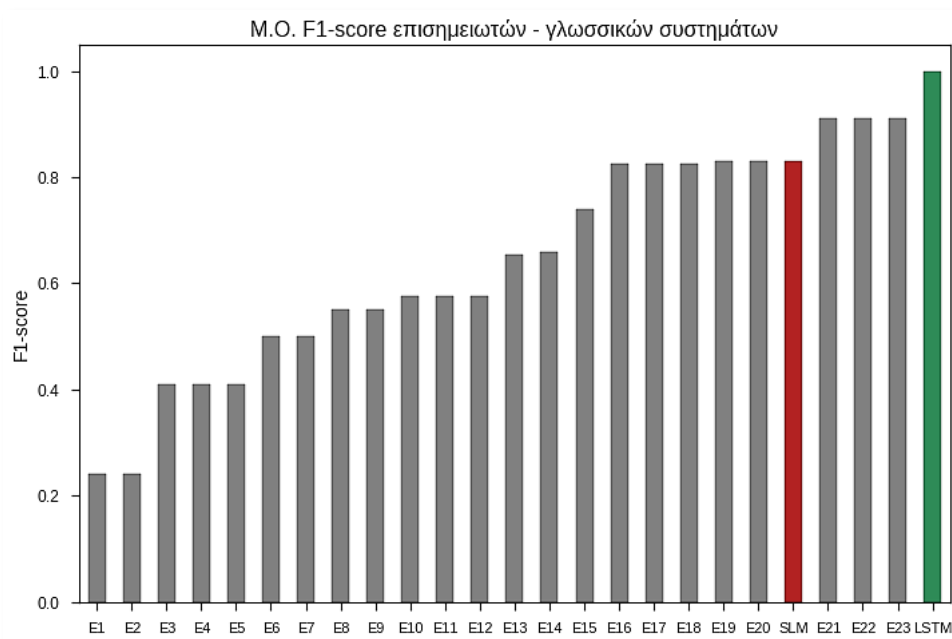
Διάγραμμα 20. F1-score για τα αποσπάσματα της Ιλιάδας και Οδύσσειας του ερωτηματολογίου

Στον παρακάτω πίνακα μπορούμε να δούμε αναλυτικότερα τις μετρήσεις *F1-score* για τα αποσπάσματα της *Ιλιάδας* και αντίστοιχα της *Οδύσσειας* του ερωτηματολογίου.

	F1-score Ιλιάδας	F1-score Οδύσσειας
<b>LSTM</b>	1.00	1.00
<b>SLM</b>	0.80	0.86
<b>Επισημειωτές</b>	0.76	0.75

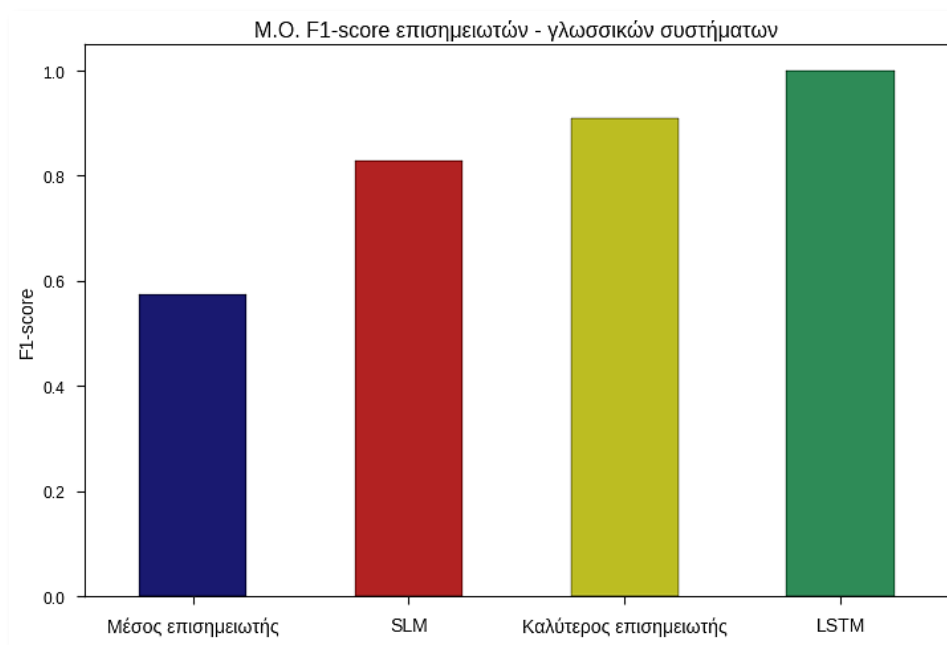
Πίνακας 12. F1-score για τα αποσπάσματα της Ιλιάδας και της Οδύσσειας του ερωτηματολογίου

Στο πιο κάτω διάγραμμα φαίνεται αναλυτικά η συνολική επίδοση, μέσω της μέτρησης *F1-score*, κάθε *επισημειωτή* στα *Ιλιαδικά* και τα *Οδυσειακά* αποσπάσματα συγκριτικά με την επίδοση των γλωσσικών συστημάτων (*SLM* και *LSTM*).



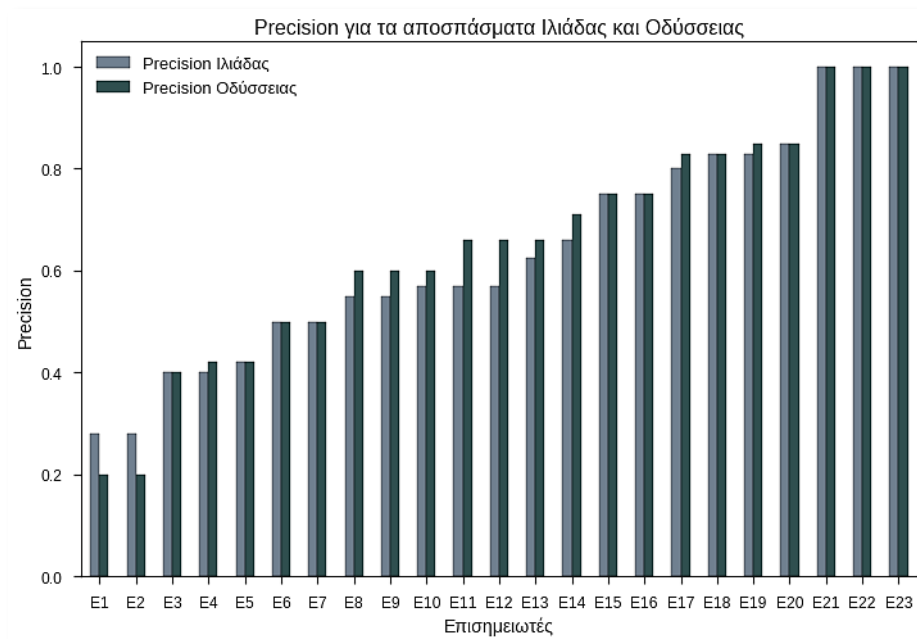
Διάγραμμα 21. Συνολική επίδοση *F1-score* επισημειωτών, *SLM* και *LSTM* για τα *Ιλιαδικά* και *Οδυσειακά* αποσπάσματα του ερωτηματολογίου

Ολοκληρώνοντας την περιγραφή των αποτελεσμάτων του δεύτερου πειράματος, στο παρακάτω διάγραμμα παρουσιάζουμε την επίδοση *F1-score* του μέσου και του καλύτερου επισημειωτή, καθώς και την επίδοση των γλωσσικών συστημάτων με τους εν λόγω επισημειωτές στα *Ιλιαδικά* και τα *Οδυσσειακά* αποσπάσματα του ερωτηματολογίου.

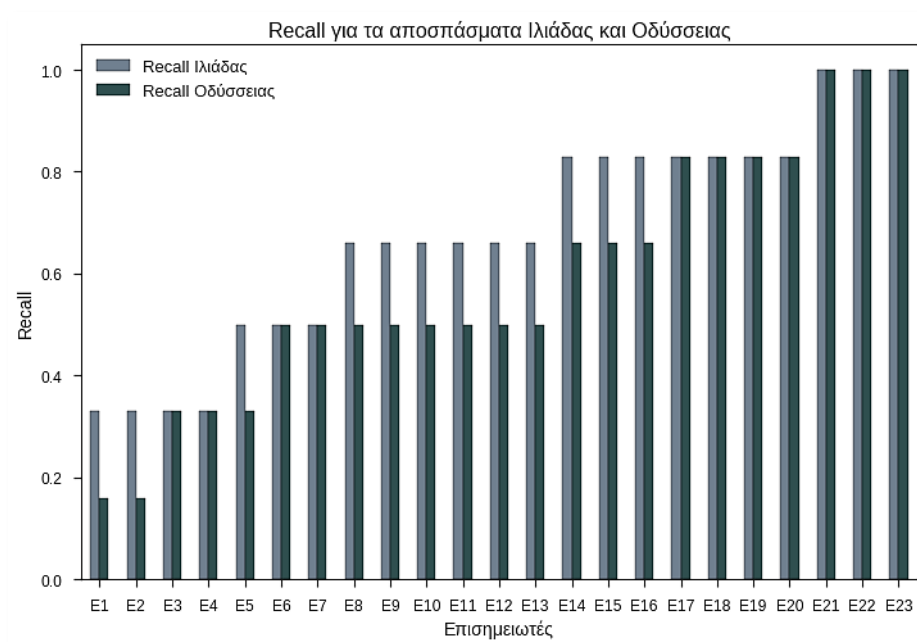


Διάγραμμα 22. Συνολική επίδοση *F1-score* μέσου επισημειωτή, καλύτερου επισημειωτή καθώς και των γλωσσικών συστημάτων, *SLM* και *LSTM*, για τα *Ιλιαδικά* και *Οδυσσειακά* αποσπάσματα του ερωτηματολογίου

Από τα παραπάνω διαγράμματα, καταλήγουμε στο συμπέρασμα, ότι σε επίπεδο γλωσσικών μοντέλων, ο ταξινομητής που βασίστηκε σε *LSTM* γλωσσικά μοντέλα σημείωσε τις καλύτερες επιδόσεις εις ότι αφορά την ταξινόμηση των αποσπασμάτων του ερωτηματολογίου απ' ότι το *SLM* και οι επισημειωτές. Συγκεκριμένα, ο ταξινομητής που βασίστηκε στα νευρωνικά γλωσσικά μοντέλα *LSTM* σημείωσε *f1-score* ίσο με **1.00**, βαθμολογία που φανερώνει ότι κατέταξε όλα τα αποσπάσματα σωστά στις αντίστοιχες κλάσεις (0 = *Οδύσσεια* | 1 = *Ιλιάδα*). Το *SLM* απέτυχε στην ορθή κατηγοριοποίηση δύο αποσπασμάτων με *f1-score* **0.83** και οι επισημειωτές, σημείωσαν *f1-score* ίσο με **0.755**, μικρότερο ποσοστό και από τα δύο γλωσσικά μοντέλα. Με άλλα λόγια, το *LSTM*, κατηγοριοποίησε ορθώς τα αποσπάσματα τόσο της *Ιλιάδας* όσο και της *Οδύσσειας*. Το *SLM* σημείωσε καλύτερη επίδοση στα *Οδυσσειακά* αποσπάσματα του ερωτηματολογίου, ενώ οι επισημειωτές απέδωσαν με μεγαλύτερη επιτυχία στα *Ιλιαδικά* αποσπάσματα, όπως μπορούμε να παρατηρήσουμε και από τις μετρήσεις *Precision* και *Recall* στα ακόλουθα διαγράμματα.



Διάγραμμα 23. Precision επισημειωτών για τα Ιλιαδικά και Οδυσειακά αποσπάσματα του ερωτηματολογίου



Διάγραμμα 24. Recall επισημειωτών για τα Ιλιαδικά και Οδυσειακά αποσπάσματα του ερωτηματολογίου

## 6. Συζήτηση και μελλοντικές κατευθύνσεις

Ανακεφαλαιώνοντας, στην παρούσα εργασία με αφετηρία τα *ομηρικά έπη* και τους καλούμενους *Ομηρικούς ύμνους* προβήκαμε σε δύο πειράματα με κύριο στόχο την αξιοποίηση της *γλωσσικής μοντελοποίησης* στο έργο της *απόδοσης συγγραφικής πατρότητας*. Εφαρμόσαμε δύο προσεγγίσεις από τη μοντελοποίηση φυσικής γλώσσας: τα *στατιστικά γλωσσικά μοντέλα* και τα *νευρωνικά γλωσσικά μοντέλα LSTM*. Τα εν λόγω γλωσσικά μοντέλα προσέγγισαν τα δεδομένα σε επίπεδο χαρακτήρα. Τα γλωσσικά μοντέλα επιπέδου χαρακτήρα, όπως αναφέρθηκε στο Κεφάλαιο 2, έχουν εφαρμοστεί σε αρκετές έρευνες [19], [20] και έχουν σημειώσει καλύτερα αποτελέσματα από τα γλωσσικά μοντέλα επιπέδου λέξεων για το ζητούμενο της *απόδοσης συγγραφικής πατρότητας*.

Στο πείραμα για την *απόδοση συγγραφικής πατρότητας* μέσω *στατιστικών γλωσσικών μοντέλων (SLM)* εξετάσαμε το συσχετισμό κάθε ραψωδίας με το έπος που υπάγεται, αλλά και με το συγγενές — θα λέγαμε — έπος (*Ιλιάδα* με *Οδύσσεια* και *τούμπαλιν*). Αναφορικά με τα κυκλικά γλωσσικά μοντέλα *SLM* και λαμβάνοντας ως δεδομένη την τυχαία λήψη των βαθμολογημένων αποσπασμάτων από κάθε ραψωδία και τη σύγκρισή τους με το σύνολο του ιδίου έπους, αλλά και του συγγενούς έπους, κατατείνουμε στο συμπέρασμα ότι σε επίπεδο αξιολόγησης γλωσσικής συγγένειας ραψωδιών με το ίδιο έπος υπαγωγής τους, προκύπτει μεγάλη ομοιότητα των αποσπασμάτων της ραψωδίας **Ο** της *Ιλιάδας* με το σύνολο του έργου, καθώς παρουσιάζει τον χαμηλότερο μέσο όρο από όλες τις άλλες ραψωδίες. Ομοίως, τα αποσπάσματα της ραψωδίας **Β** από την *Οδύσσεια* παρουσιάζουν μεγάλη ομοιότητα με το συνολικό έργο της *Οδύσσειας*. Σχετικά με τα χιαστί γλωσσικά μοντέλα *SLM*, κατατείνουμε στο συμπέρασμα ότι η ραψωδία **Δ** της *Οδύσσειας* ομοιάζει αρκετά σε γλωσσικό επίπεδο με το σύνολο της *Ιλιάδας*, καθώς παρουσιάζει το χαμηλότερο μέσο όρο μεταξύ των υπολοίπων ραψωδιών. Αντίστοιχα, η ραψωδία **Τ** της *Ιλιάδας* παρουσιάζει γλωσσική ομοιότητα με το σύνολο της *Οδύσσειας*. Από την άλλη πλευρά είδαμε ορισμένες ραψωδίες των *ομηρικών επών* καθώς και κάποιους *Ομηρικούς ύμνους* να παρεκκλίνουν από το γλωσσικό ύφος των *ομηρικών επών*. Αναλυτικότερα, στα κυκλικά μοντέλα της *Ιλιάδας* παρατηρήσαμε ότι η ραψωδία **Ε** της *Ιλιάδας* παρουσιάζει γλωσσική παρέκκλιση με το ύφος της *Ιλιάδας*. Στα κυκλικά μοντέλα της *Οδύσσειας* τη ραψωδία **μ** της *Οδύσσειας* με το σύνολο της *Οδύσσειας*. Έπειτα, στα χιαστί μοντέλα



της *Ιλιάδας* ομοίως είδαμε ότι η ραψωδία **μ** της *Οδύσσειας* και ο *Ομηρικός ύμνος «Εἰς Ἑρμῆν»* εμφανίζουν γλωσσική παρέκκλιση από το ύφος της *Ιλιάδας*. Εἰς ὅτι αφορά τα χιαστί μοντέλα *Οδύσσειας* παρατηρήσαμε πώς η ραψωδία **N** της *Ιλιάδας* και ο *Ομηρικός ύμνος «Εἰς Ἀπόλλωνα»* παρεκκλίνουν από το γλωσσικό ύφος της *Οδύσσειας*.

Στο σημείο αυτό πρέπει να αναφερθεί, ὅτι τόσο τα κυκλικά ὅσο και τα χιαστί γλωσσικά μοντέλα *SLM* εκπαιδεύτηκαν σε διαφορετικό πλήθος δεδομένων και γι' αὐτόν το λόγο δε θα πρέπει να θεωρήσουμε ὅτι οι χαμηλότερες τιμές της μέτρησης αξιολόγησης *Perplexity*, στις βαθμολογίες του μέσου ὁρου, εἶναι μεταξύ τους συγκρίσιμες. Πιο συγκεκριμένα, κάθε ένα κυκλικό μοντέλο *SLM* εκπαιδεύτηκε σε διαφορετικό υποσύνολο 23 ραψωδιῶν του ἐκάστοτε ἔπους κάθε φορά, ἐνῶ τα χιαστί γλωσσικά μοντέλα *SLM* (*Ιλιαδικό/Οδυσσειακό*) εκπαιδεύτηκαν σε 24 ραψωδίες το κάθε ένα. Προκειμένου να εξετάσουμε αὐτόν τον ισχυρισμό προχωρήσαμε σε ένα ἀκόμη πείραμα, στο οποίο αξιολογήσαμε τα *Ιλιαδικά στατιστικά γλωσσικά μοντέλα* με κάθε ραψωδία της *Οδύσσειας* και τα *Οδυσσειακά στατιστικά γλωσσικά μοντέλα* με κάθε ραψωδία της *Ιλιάδας* και είδαμε πώς οι μετρήσεις του μέσου ὁρου *Perplexity* αυξάνονταν κατὰ πολὺ. Ολοκληρώνοντας το πείραμα για την ἀπόδοση συγγραφικῆς πατρότητας μέσω στατιστικῶν γλωσσικῶν μοντέλων (*SLM*), αξιολογήσαμε τους *Ομηρικούς ὕμνους* με τα γλωσσικά μοντέλα της *Ιλιάδας* (*IliadSLM*) και της *Οδύσσειας* (*OdysseySLM*) και συμπεράναμε ὅτι τη μεγαλύτερη γλωσσική ομοιότητα πρὸς την *Ιλιάδα* και την *Οδύσεια*, ἀπὸ τους ἄλλους τρεῖς ὕμνους, παρουσιάζει ὁ ὕμνος «*Εἰς Ἀφροδίτην*». Ὁ ὕμνος «*Εἰς Ἀφροδίτην*» συχνὰ χαρακτηρίζεται ἀπὸ τους μελετητές ὡς ὁ πιο *Ομηρικός* μεταξύ των *Ομηρικῶν ὕμνων* [18].

Ἐν συνεχείᾳ, στο πείραμα για την ταξινόμηση *ομηρικῶν κειμένων* μέσω στατιστικῶν γλωσσικῶν μοντέλων και νευρωνικῶν γλωσσικῶν μοντέλων μακράς – βραχείας μνήμης, εφαρμόσαμε *SLM* και *LSTM*, ὥστε να εξετάσουμε συγκριτικά την ικανότητα κατηγοριοποίησής των γλωσσικῶν μοντέλων σε κείμενα, που φέρουν την πατρότητα του Ομήρου. Οι ἐπιδόσεις των γλωσσικῶν μοντέλων ἀντιπαραβλήθηκαν με τις κρίσεις των ἀνθρώπων-ἐπισημειωτῶν για τα κείμενα αὐτά. Με το πείραμα συμπεράναμε ὅτι γλωσσικά μοντέλα, κατηγορίας *SLM* και *LSTM*, δύνανται να ταξινομοῦν με μεγάλη ἐπιτυχία *Ομηρικά κείμενα* στο ἀντίστοιχο ἔργο υπαγωγῆς τους. Συγκεκριμένα, το *Νευρωνικό Γλωσσικό Μοντέλο Μακράς - Βραχείας μνήμης (LSTM)* σημείωσε την υψηλότερη ἐπίδοση *f1-score* (**1.00**) ἀπὸ ἐκείνη του *SLM* (**0.83**) και των ἀνθρώπων-ἐπισημειωτῶν (**0.755**) εἰς ὅτι αφορά τα *Ιλιαδικά* και τα *Οδυσσειακά*

αποσπάσματα του ερωτηματολογίου. Ο καλύτερος *επισημειωτής* σε κατάταξη αποσπασμάτων βρίσκεται ανάμεσα στα δύο γλωσσικά μοντέλα, ενώ ο μέσος *επισημειωτής* παρουσιάζει βαθμολογία *f1-score* μικρότερη από το *SLM*. Δεδομένων των μετρήσεων *F1-score*, είδαμε ότι *Νευρωνικά γλωσσικά μοντέλα*, όπως το *LSTM*, σημειώνουν αξιοσημείωτες επιδόσεις στην ταξινόμηση αρχαίων κειμένων ανάμεσα στα δύο *ομηρικά έπη* και από παραδοσιακά *Στατιστικά Γλωσσικά Μοντέλα (SLM)* και από τους *ανθρώπους-επισημειωτές*, που είναι ως ένα βαθμό εξοικειωμένοι με τα *Ομηρικά κείμενα*.

Επιχειρώντας να απαντήσουμε στα ερωτήματα που θέσαμε στην αρχή της παρούσας εργασίας είδαμε ότι:

1. Πράγματι, υπάρχουν ραψωδίες τόσο της *Ιλιάδας* όσο και της *Οδύσσειας* που εμφανίζουν μεγαλύτερη γλωσσική συγγένεια απ' ότι άλλες με το σύνολο του αντίστοιχου έπους.
2. Τα γλωσσικά μοντέλα φαίνεται να ξεχωρίζουν κάποιες ραψωδίες που έχουν μεγαλύτερη απόκλιση από το γλωσσικό ύφος των επών. Δίνεται, έτσι, έναυσμα για περαιτέρω έρευνα προκειμένου να διαπιστώσουμε, αν οι αποκλίσεις είναι αρκετά σημαντικές, ώστε να αποτελούν ενδείξεις διαφορετικής πατρότητας.
3. Ο *Ομηρικός ύμνος «Εἰς Ἀφροδίτην»* εμφανίζει τη μεγαλύτερη γλωσσική συγγένεια με το σύνολο της *Ιλιάδας* και της *Οδύσσειας* μεταξύ των ύμνων που μελετήσαμε.
4. Τεχνητά γλωσσικά μοντέλα μπορούν να κατηγοριοποιούν με μεγαλύτερη επιτυχία *Ιλιαδικά* και *Οδυσσειακά* αποσπάσματα στο αντίστοιχο έργο υπαγωγής τους απ' ότι η ανθρώπινη κρίση.

Ορισμένες μελλοντικές κατευθύνσεις θα ήταν, αρχικά, ο εμπλουτισμός του ερωτηματολογίου με περισσότερα αποσπάσματα. Έπειτα, η διερεύνηση του ζητήματος και με κατηγορίες νευρωνικών γλωσσικών μοντέλων τεχνολογίας αιχμής (*state-of-the-art*). Τέλος, η ταξινόμηση *Ομηρικών αποσπασμάτων* ανάμεσα και σε άλλους αρχαίους συγγραφείς (π.χ. *Ησίοδος*).

Εν κατακλείδι, πραγματοποιήσαμε μία ποσοτική μελέτη, με υπολογιστικές μεθόδους, στο έργο της *απόδοσης συγγραφικής πατρότητας* σε *Ομηρικά κείμενα*, που εστιάζει σε τεχνικές παραμέτρους της γλώσσας και παρουσιάσαμε τα αποτελέσματα αυτά με τη μορφή διαγραμμάτων και πινάκων. Θα μπορούσε να αναρωτηθεί εύλογα

κάποιος, μελετώντας την παρούσα εργασία, ποια η αξία των ποσοτικών μελετών στη λογοτεχνία και εν γένει στις ανθρωπιστικές επιστήμες. Οι ανθρώπινες συμπεριφορές, τα συναισθήματα και οι πεποιθήσεις δεν κατατάσσονται σε διαγράμματα και πίνακες, ωστόσο με τη συνοδοιπορία των ανθρωπιστικών επιστημών και της υπολογιστικής επιστήμης, υπάρχει αισιοδοξία ότι η χρήση των ποσοτικών δεδομένων θα οδηγήσει σε σημαντικές διαπιστώσεις. Το παρόν πόνημα ελπίζεται να αξιοποιηθεί ως ένα εφαλτήριο στη συνεχιζόμενη μακρόχρονη πορεία με προορισμό την *απόδοση συγγραφικής πατρότητας* στα αρχαία κείμενα, που κατά την κρατούσα άποψη, αποδίδονται στον επικό ποιητή Όμηρο.

## Βιβλιογραφία

### Ξενόγλωση Βιβλιογραφία

- [1] Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, 3(Feb), 1137-1155.
- [2] Chaski, C. E. (2005). Who's at the keyboard? Authorship attribution in digital evidence investigations. *International journal of digital evidence*, 4(1), 1-13.
- [3] Doraisamy, S., & Rüger, S. (2003). Robust polyphonic music retrieval with n-grams. *Journal of Intelligent Information Systems*, 21(1), 53-70.
- [4] Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- [5] Gollub, T., Potthast, M., Beyer, A., Busse, M., Rangel, F., Rosso, P., Stamatatos, E., and Stein, B. (2013). Recent trends in digital text forensics and its evaluation plagiarism detection, author identification and author profiling. In *Proceedings of Conference and Labs of the Evaluation Forum, CLEF*, pages 282–302, Valencia, Spain.
- [6] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, p. 2). Cambridge: MIT press.
- [7] Holmes, D. I., & Forsyth, R. S. (1995). The Federalist revisited: New directions in authorship attribution. *Literary and Linguistic computing*, 10(2), 111-127.
- [8] Jurafsky, D. and Martin, J. H. (2000). Speech and Language Processing: An Introduction to Natural Language Processing, *Computational Linguistics*, and Speech Recognition. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition.
- [9] Kestemont, M., Stover, J., Koppel, M., Karsdorp, F., & Daelemans, W. (2016). Authenticating the writings of Julius Caesar. *Expert Systems with Applications*, 63, 86-96.
- [10] Kimler, M. (2003). *Using style markers for detecting plagiarism in natural language documents*. Institutionen för datavetenskap.
- [11] Lesky, A. (1954). *Mündlichkeit und Schriftlichkeit im homerischen Epos*. Berger.
- [12] Love, H. (2002). *Attributing authorship: An introduction*. Cambridge University Press.

- [13] Luong, M. T., Kayser, M., & Manning, C. D. (2015, July). Deep neural language models for machine translation. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning* (pp. 305-309).
- [14] Manousakis, N., & Stamatatos, E. (2018). Devising Rhesus: A strange ‘collaboration’ between Aeschylus and Euripides. *Digital Scholarship in the Humanities*, 33(2), 347-361.
- [15] Mantegna, R. N., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Peng, C. K., Simons, M., & Stanley, H. E. (1995). Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics. *Physical Review E*, 52(3), 2939.
- [16] Meurers, D. (2012). Natural language processing and language learning. *The Encyclopedia of Applied Linguistics*.
- [17] Mikolov, T., Karafiát, M., & Burget, L. (2010). Jan ěernocký, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association* (pp. 1045-1048).
- [18] Morris, I., & Powell, B. B. (1997). *A new companion to Homer*. Brill.
- [19] Peng, F., Schuurmans, D., Keselj, V., & Wang, S. (2003, April). Language independent authorship attribution with character level n-grams. In *10th Conference of the European Chapter of the Association for Computational Linguistics*.
- [20] Sari, Y. (2018). *Neural and Non-neural Approaches to Authorship Attribution (Doctoral dissertation, University of Sheffield)*.
- [21] Schwenk, H., Déchelotte, D., & Gauvain, J. L. (2006, July). Continuous space language models for statistical machine translation. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions* (pp. 723-730).
- [22] Scott, S. and Matwin, S. (1999). Feature engineering for text classification. In *Proceedings of ICML-99, 16th International Conference on Machine Learning (Bled, SL, 1999)*, pp. 379–388.
- [23] Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, 60(3), 538-556.

[24] Witten, I. H., Bray, Z., Mahoui, M., & Teahan, B. (1999, March). Text mining: A new frontier for lossless compression. In *Proceedings DCC'99 Data Compression Conference (Cat. No. PR00096)* (pp. 198-207). IEEE.

### **Ελληνόγλωσση Βιβλιογραφία**

[25] Latacz, J. (2000). *ΟΜΗΡΟΣ, ο θεμελιωτής της ευρωπαϊκής λογοτεχνίας*. Ε. Σιστάκου, επιμ. Αν. Ρεγκάκος. Αθήνα: Εκδ. Παπαδήμας Δημ. Ν.

[26] Διαμαντάρας, Κ. και Μπότσης, Δ. (2019). *Μηχανική Μάθηση*. Αθήνα: Κλειδάριθμος.

[27] Μαρωνίτης, Ν. Δ., Πόλκας Λ. (2007). *Αρχαϊκή Επική Ποίηση: Από την Ιλιάδα στην Οδύσσεια*. Αθήνα: Ινστιτούτο Νεοελληνικών Σπουδών [Ίδρυμα Τριανταφυλλίδη].

[28] Μικρός, Γ. Κ. (2015). *Υπολογιστική Υφολογία*. Αθήνα: Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα - [www.kallipos.gr](http://www.kallipos.gr).

# Παράρτημα

## Ερωτηματολόγιο

Το ερωτηματολόγιο που δημιουργήθηκε για την παρούσα διπλωματική εργασία μπορεί να αναζητηθεί στον ακόλουθο σύνδεσμο: <https://forms.gle/WQv3zDP9Z3m2f3RWA>.

## Σκοπός ερωτηματολογίου

Το ερωτηματολόγιο απευθύνθηκε σε φοιτητές και αποφοίτους φιλολογικών τμημάτων. Σκοπός του ερωτηματολογίου ήταν να διερευνηθεί η ικανότητά των ανθρώπων-επισημειωτών να εντοπίζουν — χωρίς καμία αναζήτηση πηγών, όπως για παράδειγμα σε βιβλία ή στο διαδίκτυο — σωστά τα αποσπάσματα που αποδίδονται στον επικό ποιητή Όμηρο. Παραλλήλως, εφαρμόστηκαν στατιστικά και νευρωνικά γλωσσικά μοντέλα, που εκπαιδεύτηκαν σε κείμενα του Ομήρου, ώστε να ερευνηθεί η ικανότητα τους να ανιχνεύουν το ποσοστό κατά το οποίο αρχαία κείμενα ανήκουν ή όχι στον ποιητή Όμηρο. Σε επόμενο στάδιο, οι κρίσεις των επισημειωτών αντιπαραβλήθηκαν με τις κρίσεις των γλωσσικών μοντέλων για τα δοθέντα αποσπάσματα.

## Αποσπάσματα ερωτηματολογίου

- 6 αποσπάσματα από την *Ιλιάδα*.
- 6 αποσπάσματα από την *Οδύσσεια*.
- 8 αποσπάσματα από τέσσερις *Ομηρικούς ύμνους*. Συγκεκριμένα:
  - 2 αποσπάσματα από τον *Ομηρικό ύμνο «Εἰς Δημήτραν»*.
  - 2 αποσπάσματα από τον *Ομηρικό ύμνο «Εἰς Απόλλωνα»*.
  - 2 αποσπάσματα από τον *Ομηρικό ύμνο «Εἰς Ἑρμῆν»*.
  - 2 αποσπάσματα από τον *Ομηρικό ύμνο «Εἰς Ἀφροδίτην»*.

Τα αποσπάσματα του ερωτηματολογίου κυμαίνονταν από 580 έως 610 χαρακτήρες.

## Οδηγίες ερωτηματολογίου προς τους επισημειωτές

Στους επισημειωτές ζητήθηκε αρχικά να διαβάσουν προσεκτικά κάθε απόσπασμα και έπειτα να επιλέξουν για κάθε απόσπασμα μία από τις ακόλουθες κατηγορίες:

- α. «ΝΑΙ (Ιλιάδα)», αν θεωρούν ότι το δοθέν απόσπασμα ανήκει στο ομηρικό έπος 'Ιλιάδα'.

**β.** «ΝΑΙ (Οδύσσεια)», αν θεωρούν ότι το δοθέν απόσπασμα ανήκει στο ομηρικό έπος ‘Οδύσσεια’.

**γ.** «ΝΑΙ», αν θεωρούν ότι το δοθέν απόσπασμα ανήκει στα ομηρικά έπη, αλλά δεν είναι σε θέση να προσδιορίσουν σε ποιο από τα δύο έπη (Ιλιάδα/Οδύσσεια) ανήκει.

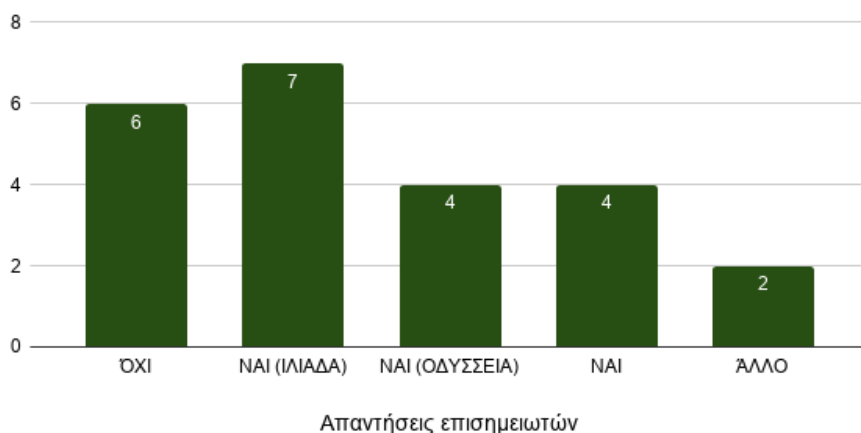
**δ.** «ΟΧΙ», αν θεωρούν ότι το δοθέν απόσπασμα δεν ανήκει στα ομηρικά έπη (Ιλιάδα/Οδύσσεια), αλλά θεωρούν ότι θα μπορούσε να αποδοθεί στον ποιητή Όμηρο λόγω γλωσσικών, μυθολογικών, ιστορικών ή και άλλων στοιχείων που τους οδήγησε σε αυτήν την απάντηση.

**ε.** «ΆΛΛΟ», αν θεωρούν ότι το απόσπασμα δεν ανήκει σε κάποια από τις παραπάνω επιλογές.



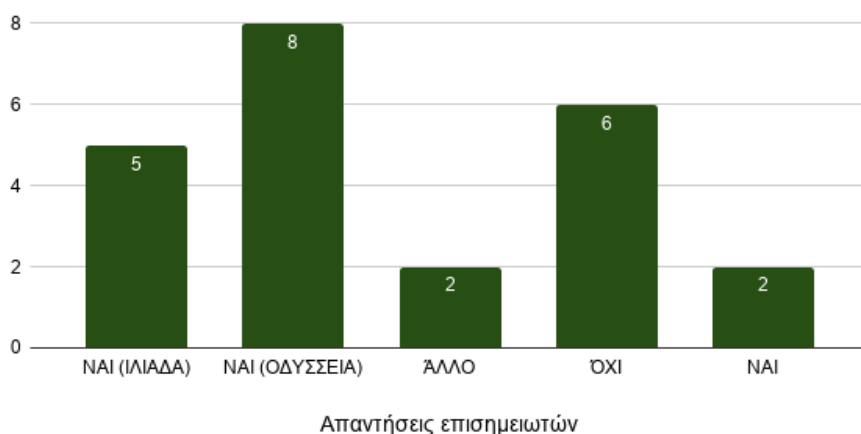
## Αναλυτική επίδοση επισημειωτών στα αποσπάσματα του ερωτηματολογίου

1. Ὡς φάτο, τήν δ' Ὑπεριονίδης ἡμείβετο μύθῳ· Ρείης ἡϋκόμου θυγάτηρ Δήμητερ ἄνασσα εἰδήσεις· δὴ γὰρ μέγα σ' ἄζομαι ἢ δ' ἐλεαίρῳ ἀχνυμένην περὶ παιδὶ ταυυσφύρῳ· οὐδέ τις ἄλλος αἴπιος ἀθανάτων εἰ μὴ νεφεληγερέτα Ζεὺς, ὅς μιν ἔδωκ' Ἀΐδῃ θαλερὴν κεκλήσθαι



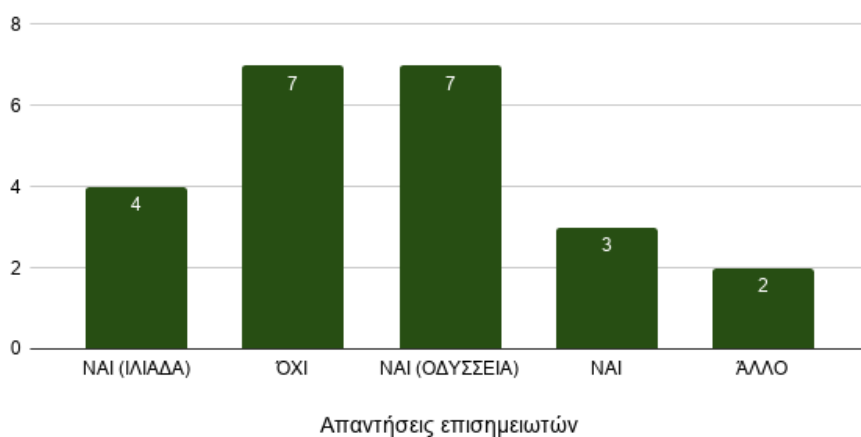
Διάγραμμα 1. Απόσπασμα από τον Ομηρικό ύμνο Εἰς Δημήτραν

2. δ' Ἴριν προὔπεμψαν ἐυκτιμένης ἀπὸ νήσου ἀξέμεν Εἰλείθειαν, ὑποσχόμεναι μέγαν ὄρμον χρυσείοισι λίνοισιν ἐερμένον, ἐννεάπηχυν· νόσφιν δ' ἦνωγον καλέειν λευκωλένου Ἥρης μὴ μιν ἔπειτ' ἐπέεσσιν ἀποστρέψειεν ἰοῦσαν. αὐτὰρ ἐπεὶ τό γ' ἄκουσε ποδὴννεμος



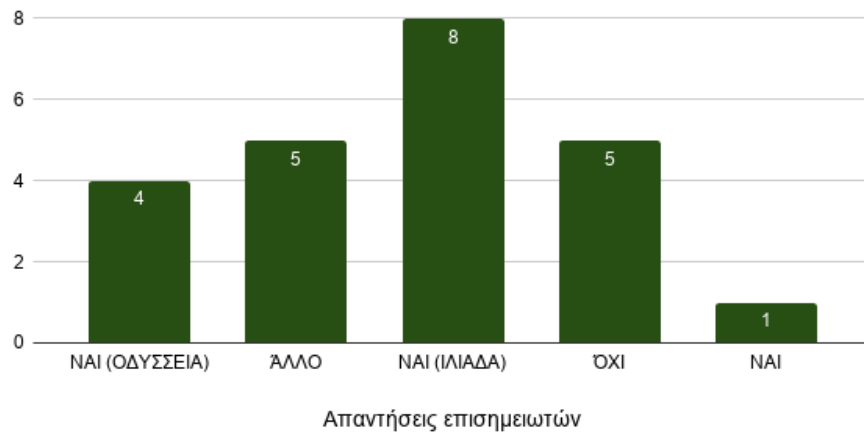
Διάγραμμα 2. Απόσπασμα από τον Ομηρικό ύμνο Εἰς Ἀπόλλωνα

3. Ὡ γέρον Ὀγχηστοῖο βατοδρόπτε ποιήεντος βοῦς ἀπὸ Πιερὶς διζήμενος ἐνθάδ' ἱκάνω πάσας θηλείας, πάσας κεράεσσιν ἐλικτάς, ἐξ ἀγέλης· ὁ δὲ ταῦρος ἐβόσκετο μοῦνος ἀπ' ἄλλων κυάνεος· χαροποὶ δὲ κύνες κατόπισθεν ἔποντο τέσσαρες, ἥύτε φῶτες ὁμόφρονες·



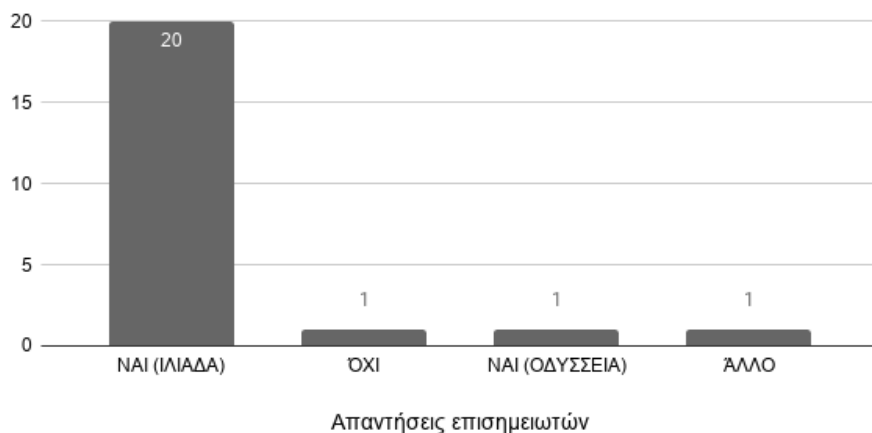
Διάγραμμα 3. Απόσπασμα από τον Ομηρικό ύμνο Εἰς Ἑρμῆν

4. Αὐτίκα σ' ὡς τὰ πρῶτα θεὰ ἴδον ὀφθαλμοῖσιν ἔγνω· ὡς θεὸς ἦσθα· σὺ δ' οὐ νημερτὲς ἔειπες· ἀλλὰ σε πρὸς Ζηνὸς γουνάζομαι αἰγιόχοιο μὴ με ζῶντ' ἀμενηνὸν ἐν ἀνθρώποισιν ἔασης ναῖεν, ἀλλ' ἐλέαιρ'· ἐπεὶ οὐ βιοθάλμιος ἀνὴρ γίγνεται ὃς τε θεαῖς εὐνάζεται



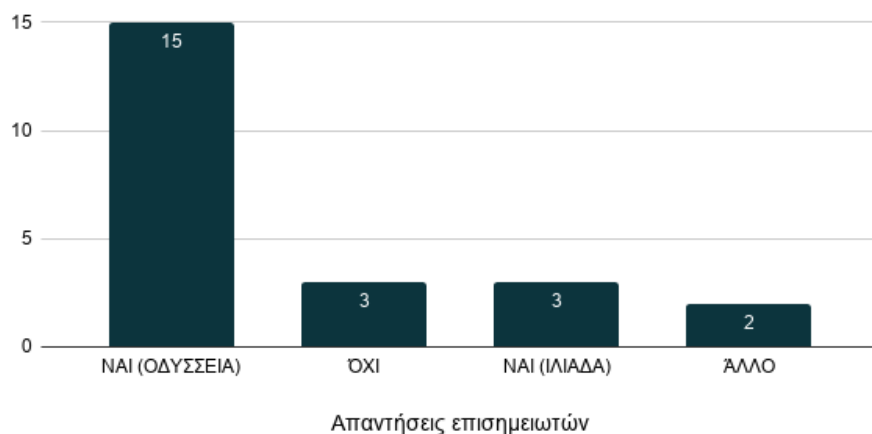
Διάγραμμα 4. Απόσπασμα από τον Ομηρικό ύμνο Εἰς Ἀφροδίτην

5. Τὴν δὲ χολωσαμένη προσεφώνεε δὴ Ἀφροδίτη· «μὴ μ' ἔρεθε σχετλίη, μὴ χωσαμένη σε μεθείω, τῶς δὲ σ' ἀπεχθήρῃ ὡς νῦν ἔκπαγλ' ἐφίλησα, μέσσω δ' ἀμφοτέρων μητίσομαι ἔχθεα λυγρὰ Τρώων καὶ Δαναῶν, σὺ δὲ κεν κακὸν οἶτον ὀλῃαι. Ὡς ἔφατ', ἔδεισεν δ' Ἑλένη



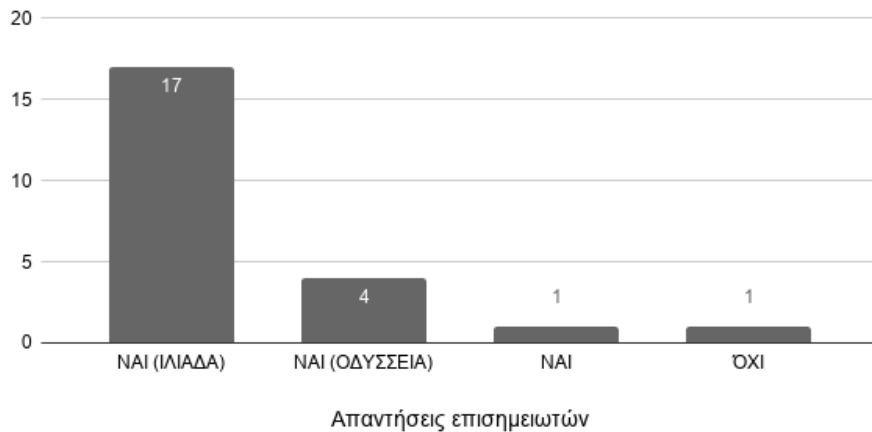
Διάγραμμα 5. Απόσπασμα από τη Ραψωδία Γ της Ιλιάδας

6. ἐν δὲ γυνὴ ταμίη σῖτον καὶ οἶνον ἔθηκεν ὄψα τε, οἷα ἔδουσι διοτρεφεὲς βασιλῆες· ἂν δ' ἄρα Τηλέμαχος περικαλλέα βήσετο δῖφρον· πᾶρ δ' ἄρα Νεστορίδης Πεισίστρατος, ὄρχαμος ἀνδρῶν, ἐς δῖφρον τ' ἀνέβαινε καὶ ἡνία λάζετο χερσὶ μάστιξεν δ' ἐλάαν, τῷ δ' οὐκ



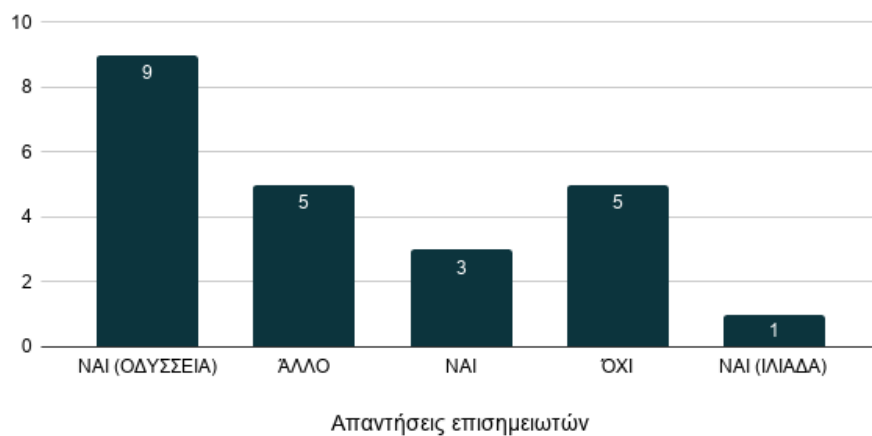
Διάγραμμα 6. Απόσπασμα από τη Ραψωδία γ της Οδύσσειας

7. φάθ', ὁ δ' ἐξ ὕπνοιο μάλα κραιπνῶς ἀνόρουσε, καί μιν φωνήσας ἔπεα πτερόεντα προσηύδα· «σχετλίος ἔσσι γεραῖε· σὺ μὲν πόνου οὐ ποτε λήγεις· οὐ νυ καὶ ἄλλοι ἔασι νεώτεροι υἱές Αἰχαιῶν, οἳ κεν ἔπειτα ἕκαστον ἐγείρειαν βασιλῆων πάντη ἐποικόμενοι· σὺ δ'



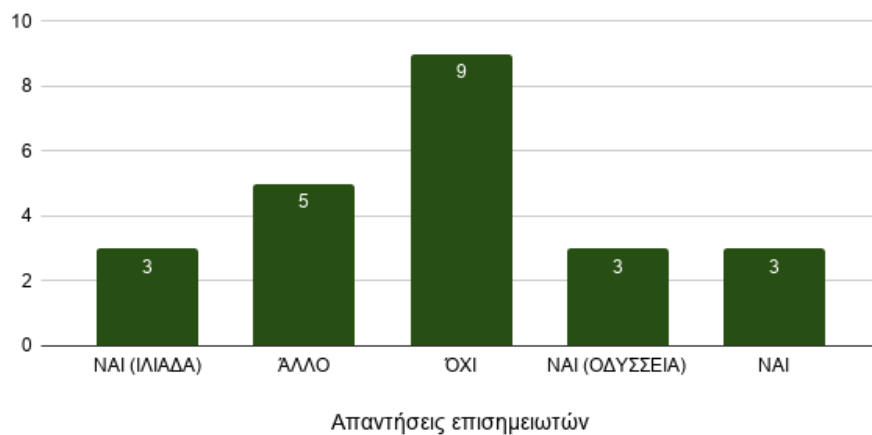
Διάγραμμα 7. Απόσπασμα από τη Ραψωδία Κ της Ιλιάδας

8. ἢ δ' ὅτε δὴ θάλαμον τὸν ἀφίκετο διὰ γυναικῶν οὐδὸν τε δρυῖνον προσεβήσето, τὸν ποτε τέκτων ξέσσαν ἐπισταμένως καὶ ἐπὶ στάθμην ἴθυνεν, ἐν δὲ σταθμούς ἄρσε, θύρας δ' ἐπέθηκε φαεινὰς, αὐτίκ' ἄρ' ἢ γ' ἰμάντα θοῶς ἀπέλυσε κορώνης, ἐν δὲ κληῖδ' ἦκε, θυρέων



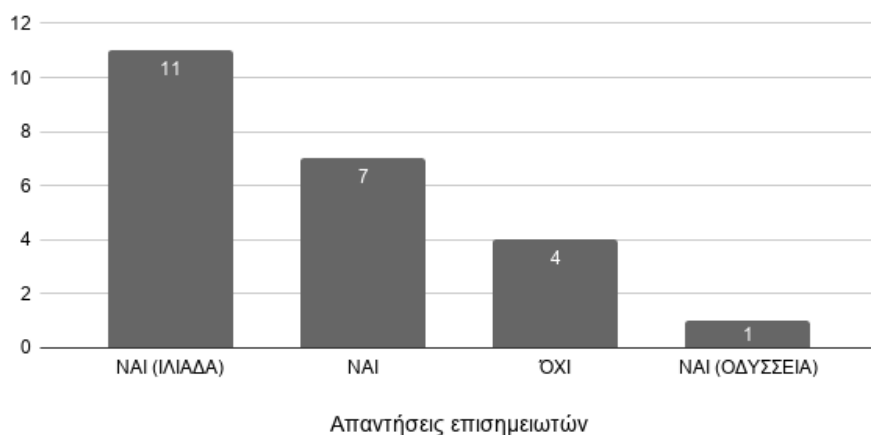
Διάγραμμα 8. Απόσπασμα από τη Ραψωδία φ της Οδύσσειας

9. εἰπὼν ἵπποισιν ἐκέκλετο, τοῖ δ' ὑπ' ὁμοκλής ῥίμφ' ἔφερον θοὸν ἄρμα τανύπτεροι ὥς τ' οἰωνοί· τὴν δ' ἄχος αἰνότερον καὶ κύντερον ἴκετο θυμόν· χωσαμένη δ' ἔπειτα κελαINEΦΕῖ Κρονίωνι νοσφισθεῖσα θεῶν ἀγορὴν καὶ μακρὸν Ὀλυμπον ὥχετ' ἐπ' ἀνθρώπων πόλιν



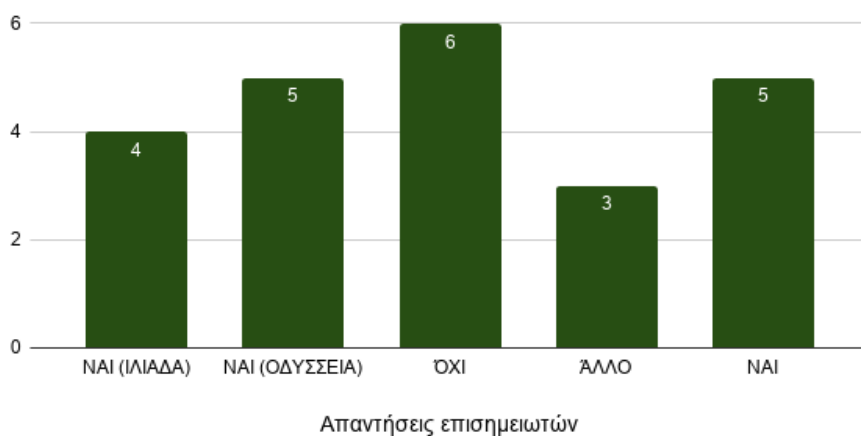
Διάγραμμα 9. Απόσπασμα από τον Ομηρικό ὕμνο Εἰς Δημήτραν

10. φωνήσασα πάλιν τρέπε μώνυχας ἵππους· τῆσιν δ' ὦραι μὲν λῦσαν καλλίτριχας ἵππους, καὶ τοὺς μὲν κατέδησαν ἐπ' ἀμβροσίῃσι κάπησιν, ἄρματα δ' ἔκλιναν πρὸς ἐνώπια παμφανόνωντα· αὐταὶ δὲ χρυσέοισιν ἐπὶ κλισμοῖσι κάθιζον μίγδ' ἄλλοισι θεοῖσι, φίλον



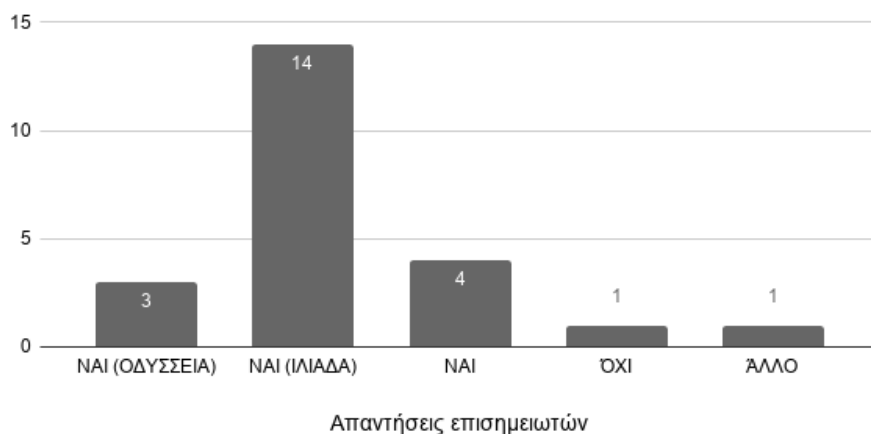
Διάγραμμα 10. Απόσπασμα από τη Ραψωδία Θ της Ιλιάδας

11. Τόσσον φᾶς συνέσευε βοῶν ἴφθιμα κάρηνα. πολλὰ δ' ὄρη σκιάοντα καὶ αὐλώνας κελαδεινοὺς καὶ πεδί' ἀνθεμόνεντα διήλασε κύδιμος Ἑρμῆς. ὀρφναίη δ' ἐπίκουρος ἐπαύετο δαιμονίη νύξ ἢ πλείων, τάχα δ' ὄρθρος ἐγίνετο δημοεργός· ἡ δὲ νέον σκοπιῆν



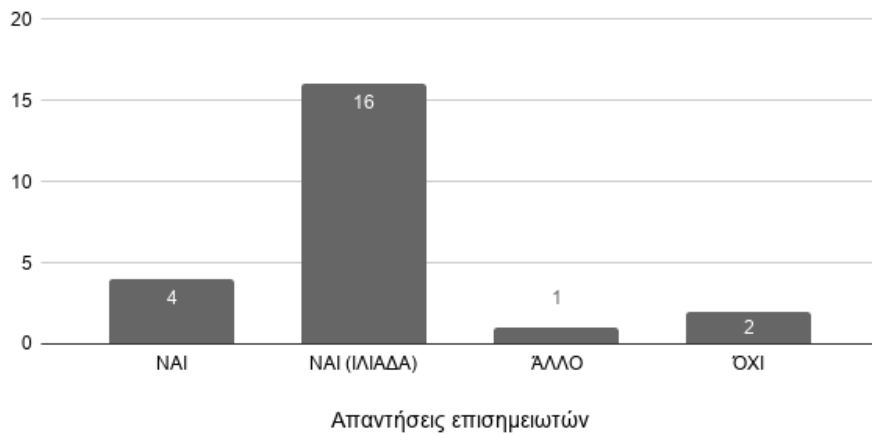
Διάγραμμα 11. Απόσπασμα από τον Ομηρικό ύμνο Εἰς Ἑρμῆν

12. προσαιδήτην βασιλῆα μελιχίοις ἐπέεσσιν· ἀμείλικτον δ' ὅπ' ἄκουσαν· «εἰ μὲν δὴ Ἀντιμάχοιο δαΐφρονος υἱέες ἐστόν, ὅς ποτ' ἐνὶ Τρώων ἀγορῇ Μενέλαον ἄνωγεν, ἀγγελίην ἐλθόντα σὺν ἀντιθέῳ Ὀδυσῆϊ αὐθι κατακτεῖναι μῆδ' ἐξέμεν ἄψ ἔς Ἀχαιοὺς, νῦν μὲν δὴ τοῦ



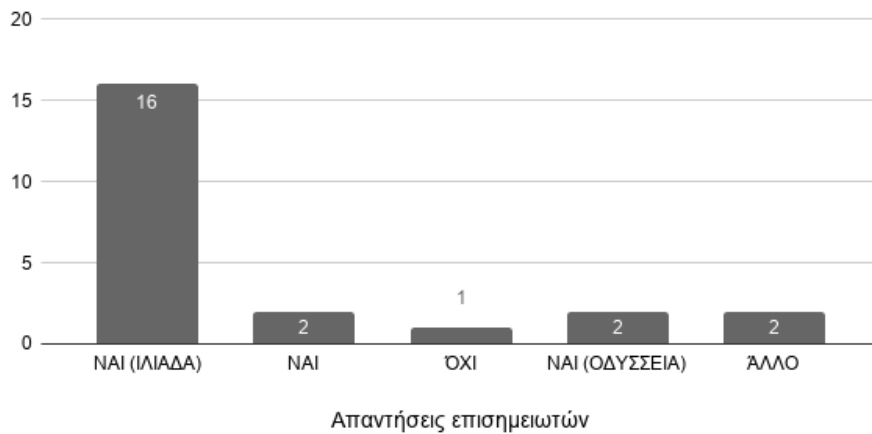
Διάγραμμα 12. Απόσπασμα από τη Ραψωδία Α της Ιλιάδας

13. Ὡς εἰπὼν ἔμπνευσε μένος μέγα ποιμένι λαῶν. ὥς δ' ὅτε τις στατὸς ἵππος ἀκοστήσας ἐπὶ φάνη, δεσμὸν ἀπορρήξας θεῖη πεδίοιο κροαίνων, εἰωθὼς λούεσθαι εὐρὺ ῥεῖος ποταμοῖο, κυδιόων· ὑψοῦ δὲ κάρη ἔχει, ἀμφὶ δὲ χαῖται ὤμοις αἴσσονται· ὁ δ' ἀγλαΐηφι



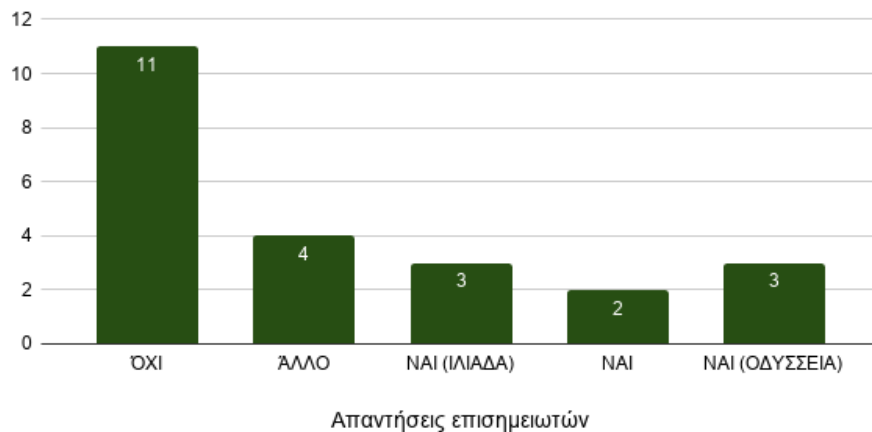
Διάγραμμα 13. Απόσπασμα ἀπὸ τὴ Ραψωδία Ο τῆς Ἰλιάδας

14. κρύπτων ἐν δίνῃσι βαθείησιν μεγάλῃσι. δεινὸν δ' ἀμφ' Ἀχιλῆα κυκῶμενον ἴστατο κῦμα, ὥθει δ' ἐν σάκει πίπτων ῥόος· οὐδὲ πόδεσσιν εἶχε στηρίξασθαι· ὁ δὲ πτελέην ἔλε χερσὶν εὐφυέα μεγάλην· ἥ δ' ἐκ ῥιζῶν ἐριποῦσα κρημνὸν ἅπαντα διώσεν, ἐπέσχε δὲ καλὰ ῥέεθρα



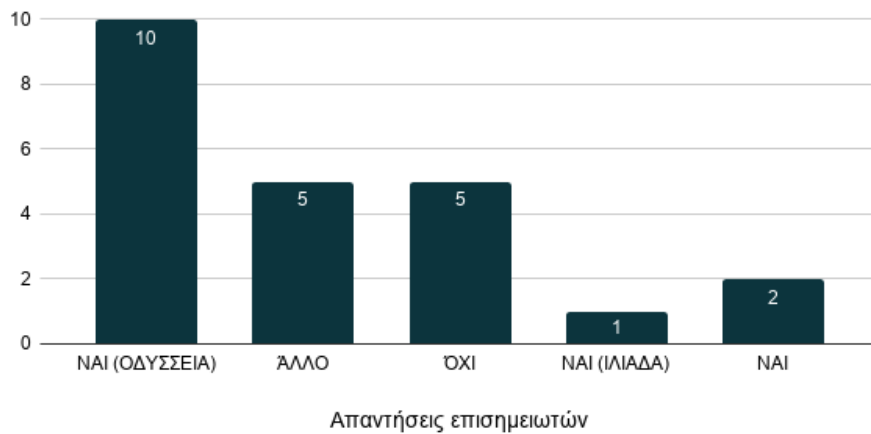
Διάγραμμα 14. Απόσπασμα ἀπὸ τὴ Ραψωδία Φ τῆς Ἰλιάδας

15. Αὐτὰρ ἐπεὶ ῥ' ὁμοσέν τε τελευτήσεν τε τὸν ὄρκον, Δῆλος μὲν μάλα χαῖρε γόνῳ ἐκάτοιο ἄνακτος, Λητῷ δ' ἐννῆμαρ τε καὶ ἐννέα νύκτας ἀέλπτοις ὠδίνεσσι πέπαρτο. θεαὶ δ' ἔσαν ἐνδοθὶ πᾶσαι ὅσσαι ἄρισται ἔσαν, Διώνη τε Ῥεῖη τε Ἰχναίη τε Θέμις καὶ ἀγάστονος



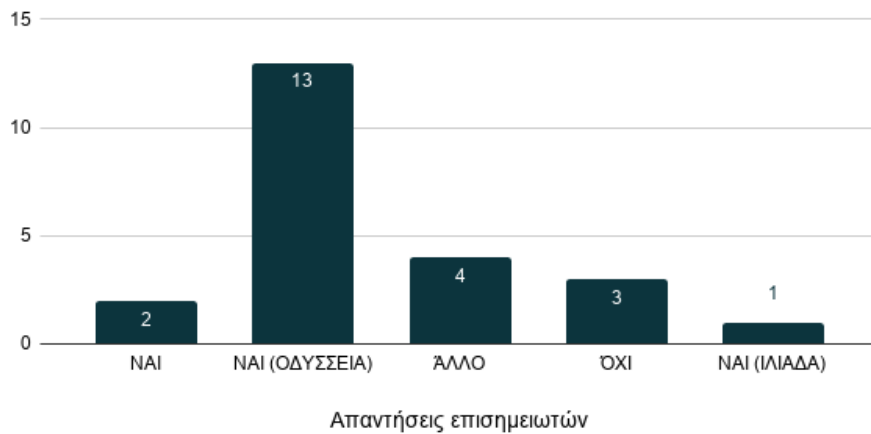
Διάγραμμα 15. Απόσπασμα ἀπὸ τὸν Ὀμηρικὸ ὕμνο Εἰς Ἀπόλλωνα

16. δμώεσσιν ἐκέκλετο, τοῖ δ' ἐπίθοντο. οἱ μὲν ἄρ' ἐκτὸς ἄμαξαν ἐϋτροχον ἡμιονεῖην ὄπλεον ἡμιόνους θ' ὕπαγον ζευξάν θ' ὑπ' ἀπήνη· κούρη δ' ἐκ θαλάμοιο φέρεν ἐσθῆτα φαεινὴν. καὶ τὴν μὲν κατέθηκεν εὐξέστω ἐπ' ἀπήνη· μήτηρ δ' ἐν κίστῃ ἐτίθει μενοεικέ'



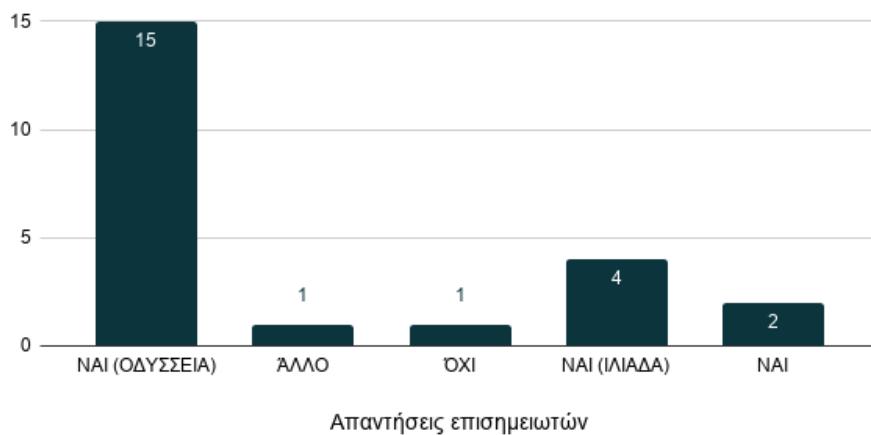
Διάγραμμα 16. Απόσπασμα από τη Ραψωδία ζ της Οδύσσειας

17. δὲ κακὴ νίκησεν ἐταίρων· ἀσκὸν μὲν λῦσαν, ἄνεμοι δ' ἐκ πάντες ὄρουσαν, τοὺς δ' αἶψ' ἄρπάξασα φέρεν πόντονδε θύελλα κλαίοντας, γαίης ἄπο πατρίδος. αὐτὰρ ἐγὼ γε ἐγρόμενος κατὰ θυμὸν ἀμύμονα μερμήριξα, ἥε πεσὼν ἐκ νηὸς ἀποφθίμην ἐνὶ πόντῳ, ἧ



Διάγραμμα 17. Απόσπασμα από τη Ραψωδία κ της Οδύσσειας

18. προσέειπε θεὰ γλαυκῶπις Ἀθήνη· «νήπιός εἰς, ὦ ξεῖν', ἧ τηλόθεν εἰλήλουθας, εἰ δὴ τήνδε τε γαῖαν ἀνείρεαι. οὐδὲ τι λίην οὕτω νώνυμός ἐστιν· ἴσασι δὲ μιν μάλα πολλοί, ἡμὲν ὅσοι ναῖουσι πρὸς ἠῷ τ' ἠελίῳ τε, ἧδ' ὅσοι μετόπισθε ποτὶ ζόφον ἡερόεντα. ἧ τοι μὲν



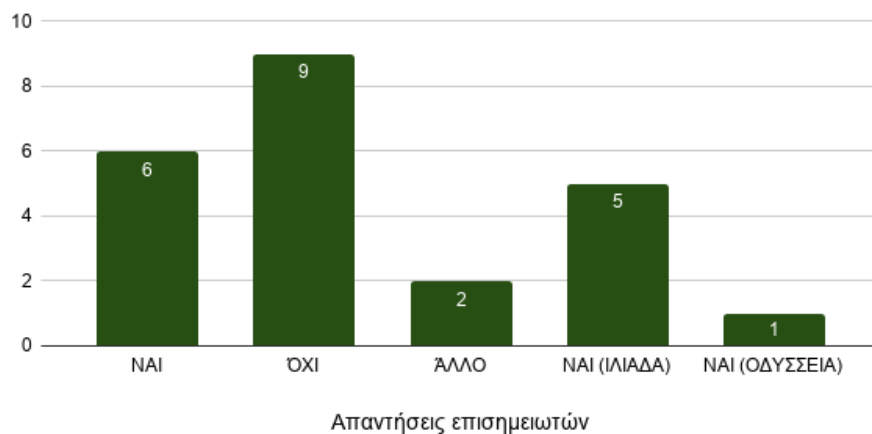
Διάγραμμα 18. Απόσπασμα από τη Ραψωδία ν της Οδύσσειας

19. πῶ τι πάρος τοιοῦτον ἐτύχθη, οἷην τερπωλὴν θεὸς ἤγαγεν ἐς τόδε δῶμα· ὁ ξεῖνός τε καὶ ἴρος ἐρίζετον ἀλλήλοισιν χερσὶ μαχέσσασθαι· ἀλλὰ ξυνελάσσομεν ὥκα.» ὥς ἔφαθ', οἱ δ' ἄρα πάντες ἀνήϊξαν γελῶντες, ἀμφὶ δ' ἄρα πτωχοὺς κακοεῖμονας ἡγερέθοντο. τοῖσιν δ'



Διάγραμμα 19. Απόσπασμα από τη Ραψωδία σ της Οδύσσειας

20. θεὰ γλυκὺν ἴμερον ἔμβαλε θυμῷ. Ἀγχίσην δ' ἔρος εἶλεν, ἔπος τ' ἔφατ' ἔκ τ' ὀνόμαζεν· Εἰ μὲν θνητὴ τ' ἐσσί, γυνὴ δέ σε γείνατο μήτηρ, Ὅτρεὺς δ' ἐστὶ πατὴρ ὀνομακλυτός, ὥς ἀγορεύεις, ἀθανάτου δὲ ἔκhti διακτόρου ἐνθάδ' ἱκάνεις Ἑρμέω, ἐμὴ δ' ἄλοχος κεκλήσεται



Διάγραμμα 20. Απόσπασμα από τον Ομηρικό ύμνο Εἰς Ἀφροδίτην