



ΕΘΝΙΚΟΝ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟΝ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

«ΔΙΔΑΣΚΑΛΙΑ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΩΣ ΔΕΥΤΕΡΗΣ/ΞΕΝΗΣ ΓΛΩΣΣΑΣ»

ΦΙΛΟΣΟΦΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΦΙΛΟΛΟΓΙΑΣ

ΤΟΜΕΑΣ ΓΛΩΣΣΟΛΟΓΙΑΣ

Τσιπουριάρη Μαρία

Διερεύνηση έξι δεικτών λεξιλογικού «πλούτου» σε Σώμα Κειμένων Μαθητών/-τριών της ελληνικής ως Δεύτερης Γλώσσας (Γ2) και συσχέτιση με τα επίπεδα γλωσσομάθειας

ΑΘΗΝΑ 2020

ΕΘΝΙΚΟΝ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟΝ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

«ΔΙΔΑΣΚΑΛΙΑ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΩΣ ΔΕΥΤΕΡΗΣ/ΞΕΝΗΣ ΓΛΩΣΣΑΣ»

ΦΙΛΟΣΟΦΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΦΙΛΟΛΟΓΙΑΣ

ΤΟΜΕΑΣ ΓΛΩΣΣΟΛΟΓΙΑΣ

Τσιπουριάρη Μαρία, Α.Μ.: 404

Διερεύνηση έξι δεικτών λεξιλογικού «πλούτου» σε Σώμα Κειμένων Μαθητών/-τριών της ελληνικής ως Δεύτερης Γλώσσας (Γ2) και συσχέτιση με τα επίπεδα γλωσσομάθειας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΠΙΒΛΕΠΩΝ:

Μαρκόπουλος Γεώργιος

ΕΠΙΤΡΟΠΗ:

Ιακώβου Μαρία

Αναπληρώτρια Καθηγήτρια Εφαρμοσμένης Γλωσσολογίας, ΕΚΠΑ

Μαρκόπουλος Γεώργιος

Αναπληρωτής Καθηγητής Υπολογιστικής Γλωσσολογίας, ΕΚΠΑ

Μιχάλης Αθανάσιος

Επίκουρος Καθηγητής στον Τομέα Παιδαγωγικής, Τμήμα ΦΠΨ, ΕΚΠΑ

ΑΘΗΝΑ 2020

ΕΥΧΑΡΙΣΤΙΕΣ

Το συγκεκριμένο μεταπτυχιακό μου προσέφερε πολλά γνωστικά αλλά και ψυχικά εφόδια, καθώς με έφερε σε επαφή με σπουδαίες διδάσκουσες και διδάσκοντες, με ένα γοητευτικό και μέχρι πρότινος ανεξερεύνητο αντικείμενο αλλά και με πολύ καλές φίλες. Γι' αυτό θα ήθελα να ευχαριστήσω τις/τους διδάσκουσες/-οντες που μας παραστάθηκαν αυτά τα δύο έτη και μας μετέδωσαν τις γνώσεις τους. Ιδιαίτερα τον καθηγητή μου, κύριο Μικρό Γεώργιο, που στάθηκε αφορμή και συνοδοιπόρος στη διερεύνηση του κλάδου της υφομετρίας. Ένα μεγάλο ευχαριστώ θα ήθελα να πω και στον επιβλέποντα καθηγητή μου, κύριο Μαρκόπουλο Γεώργιο, τον οποίο γνώριζα κυρίως μέσα από το αξιόλογο επιστημονικό του έργο, αλλά η συγκεκριμένη εργασία στάθηκε αφορμή να γνωριστούμε και σε διαπροσωπικό επίπεδο και χαίρομαι πολύ γι' αυτό. Τις ευχαριστίες μου θα ήθελα να εκφράσω και στα άλλα δύο μέλη της επιτροπής επίβλεψης της εργασίας μου, τον κύριο Μιχάλη Θανάση και την κυρία Ιακώβου Μαρία, τους οποίους γνώρισα και μέσα από τις παραδόσεις των μαθημάτων και εκτιμώ ιδιαίτερα. Πέρα από τους καθηγητές/-τριες, θα ήθελα να ευχαριστήσω το βοηθητικό προσωπικό και το προσωπικό της βιβλιοθήκης του Πανεπιστημίου, ιδιαίτερα τον κύριο Λάμπρο Σιαπέρα. Ευχαριστώ πολύ και τις καλές φίλες που μου χάρισε το μεταπτυχιακό πρόγραμμα, ιδιαίτερα την Ιωάννα και την Κατερίνα, για την υποστήριξη και τις όμορφες στιγμές που μοιραστήκαμε και συνεχίζουμε να μοιραζόμαστε. Είμαι ευγνώμων για τα πραγματικά αξιόλογα άτομα που γνώρισα σε αυτό τον διετή κύκλο σπουδών.

Πέρα από το επιστημονικό και βοηθητικό προσωπικό της σχολής, θα ήθελα να ευχαριστήσω θερμά κάποια πολύ αγαπημένα μου πρόσωπα, χωρίς την υποστήριξη των οποίων δε θα τα είχα καταφέρει. Πρόκειται για την μητέρα μου, που μου παρείχε τόσο διακριτικά αμέριστη υποστήριξη, τον πατέρα μου και την πολυαγαπημένη μου αδερφή Ανθή. Θα ήθελα να πω ένα μεγάλο ευχαριστώ και στην Αθηνά, η οποία αναγνώριζε την προσπάθειά μου και ήταν εκεί όποτε τη χρειαζόμουν σε γνωστικό και συναισθηματικό επίπεδο και στον Άγγελο που πίστευε σε μένα. Ευχαριστώ θα ήθελα να πω και στους κοντινούς μου φίλους και συγγενείς που με στήριζαν πολύ. Τέλος, θέλω να ευχαριστήσω από καρδιάς τον αγαπημένο μου σύντροφο Κωνσταντίνο, ο οποίος κατανοούσε ότι πρόκειται για μια απαιτητική για εμένα περίοδο και με υποστήριξε σε ψυχικό αλλά και πρακτικό επίπεδο.

ΠΕΡΙΛΗΨΗ

Η συγκεκριμένη έρευνα βασίζεται σε ένα ηλεκτρονικό σώμα 224 μαθητικών κειμένων (ΗΣΜΚ) έκτασης 180-300 λέξεων τα οποία αποτελούν απαντήσεις σε θέματα επιπέδου γλωσσομάθειας B2 στο πλαίσιο των εξετάσεων του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας. Οι συγκεκριμένες Παραγωγές Γραπτού Λόγου (ΠΓΛ) ανήκουν σε αλλόγλωσσους μαθητές/-τριες οι οποίοι/-ες είχαν παρακολουθήσει ένα οκτάμηνο πρόγραμμα εντατικής διδασκαλίας της ελληνικής ως Δεύτερης Γλώσσας (Γ2) περί τις 480 ώρες και κρίθηκαν προβιβαστέου/-εες. Η συλλογή των κειμένων αυτών διενεργήθηκε στο πλαίσιο του προγράμματος «Πυθαγόρας Ι» με φορέα υλοποίησης το Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών (2004 - 2007).

Παρ' ότι τυπικά οι υπό εξέταση ΠΓΛ ανήκουν στο B2 επίπεδο, παρουσιάζουν σημαντικές αποκλίσεις μεταξύ τους και δεν είναι τυχαίο το γεγονός ότι παράγονται από μαθητές/-τριες που προέρχονται και από τα τρία επίπεδα γλωσσομάθειας (αρχάριο, μέσο, προχωρημένο επίπεδο) της ελληνικής ως Γ2. Αξιοποιώντας τη γλωσσολογική ανάλυση βάσει των ΗΣΜΚ, η παρούσα εργασία επικεντρώνεται στη λεξιλογική ικανότητα του μαθητικού κοινού, η οποία θα διερευνηθεί μέσα από την ποσοτική ανάλυση συγκεκριμένων δεικτών λεξιλογικού «πλούτου» στα γραπτά των αλλόγλωσσων μαθητών/-τριών. Πρόκειται για το Ποσοστό Άπαξ Λεγομένων ή Harax Legomenon Percentage (HL), το R1, το Lambda (Λ), το h-point (h), την Εντροπία ή Entropy (H) και το Μέσο Μήκος Λεξικού Τύπου ή Average Tokens length (ATL). Σκοπός είναι να διερευνηθεί η συσχέτιση των δεικτών αυτών με τα τρία επίπεδα γλωσσομάθειας, στα οποία ανήκαν οι συμμετέχοντες/-ουσες.

Προκειμένου να εντοπιστούν πιθανές συσχετίσεις, μεθοδεύθηκε η ποσοτικοποίηση των ΠΓΛ και η απόδοση των δεικτών λεξιλογικού «πλούτου» σε αριθμούς μέσω του εργαλείου ποσοτικοποίησης QUITA. Για την ποσοτική και ποιοτική ανάλυση των δεδομένων, όπως αυτά αποτυπώθηκαν στο QUITA, χρησιμοποιήθηκε το λογισμικό Microsoft Excel 2010. Πέρα από αυτά τα δύο λογισμικά, χρησιμοποιήθηκε το στατιστικό πρόγραμμα SPSS. Για τη σύγκριση των τριών επιπέδων, προκειμένου να διαπιστωθεί ποιοι δείκτες σε ποια επίπεδα εμφανίζουν στατιστικά σημαντική διαφορά, αξιοποιήθηκε το παραμετρικό στατιστικό κριτήριο one way ANOVA (Analysis of Variance – ANOVA). Επίσης, αξιοποιήθηκε η προβλεπτική εξίσωση

πολλαπλής παλινδρόμησης (Multiple Regression Analysis- MRA), για να εξεταστεί αν οι δείκτες που κρίθηκαν στατιστικά σημαντικοί λήφθηκαν υπόψη κατά τη βαθμολόγηση του λεξιλογίου.

Τα αποτελέσματα των παραπάνω μετρήσεων ανέδειξαν την ανεπάρκεια και τις αδυναμίες του υπάρχοντος πλαισίου βαθμολόγησης. Τρεις από τους έξι υπό εξέταση δείκτες λεξιλογικού «πλούτου», το Μέσο Μήκος Λεξικού Τύπου (ATL), το Lambda (Λ) και το Ποσοστό Άπαξ Λεγομένων (HL), κρίθηκαν στατιστικά σημαντικοί. Το γεγονός αυτό πιστοποιεί την ανομοιογένεια μεταξύ των μαθητών/-τριών που τυπικά ανήκουν στο Β2 Επίπεδο και επιβεβαιώνει τη συσχέτιση των δεικτών αυτών με τα επίπεδα γλωσσομάθειας. Τα πορίσματα της έρευνας κατέδειξαν, επίσης, τον αποκλεισμό των δεικτών λεξιλογικού «πλούτου» από τα βαθμολογικά κριτήρια.

Τα δεδομένα αυτά αποτέλεσαν έναυσμα για την εξέταση του ισχύοντος πλαισίου βαθμολόγησης του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας και του πλαισίου βαθμολόγησης του Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς για τη δεύτερη γλώσσα (ΚΕΠΑ). Αν και από τα δύο υιοθετείται η αναλυτική βαθμολόγηση, από το πρώτο δεν υπάρχει σαφής καθορισμός της κατανομής των μονάδων που αντιστοιχούν στο λεξιλόγιο (6 από τις 20 μονάδες) και από το δεύτερο δεν έχει διαλευκανθεί η ποιότητα και η ποσότητα του λεξιλογίου που αντιστοιχεί σε κάθε επίπεδο γλωσσομάθειας. Καταδεικνύεται, λοιπόν, η ανάγκη αναθεώρησης και συγκεκριμενοποίησης του ισχύοντος πλαισίου αξιολόγησης και η συμπερίληψη-τουλάχιστον- των τριών δεικτών λεξιλογικού «πλούτου» που κρίθηκαν στατιστικά σημαντικοί.

ABSTRACT

This research is based on an Learner Corpus of 224 180-300 word texts (ISMKs) which are answers to B2 Proficiency Level topics in the context of Modern Greek Language Teaching Center 's examinations in greek as a Second Language (L2). These written productions belong to language learners who had attended an eight-month intensive teaching program of Greek as a Second Language (L2) for approximately 480 hours and were deemed eligible. The collection of these texts was carried out in the framework of the program "Pythagoras I" implemented by the National and Kapodistrian University of Athens (2004 - 2007)

Although typically these texts belong to the B2 Proficiency Level, they differ significantly. This deviance associated with the fact that they are produced by students from three proficiency levels (beginner, intermediate, advanced) of Greek as L2. Drawing on the linguistic analysis based on ISMKs, the present work focuses on the students' vocabulary competence, which will be explored through the quantitative analysis of specific lexical "richness" indices in the texts of the illiterate students. These are Hapax Legomenon Percentage (HL), R1, Lambda (L), h-point (h), Entropy (H) and Average Tokens length (ATL). The aim is to investigate the correlation of these indices with Three Proficiency Levels to which the participants belonged.

In order to identify possible associations, the quantification of student's output and the rendering of lexical "richness" indices in numbers were performed using the QUITA quantification tool. The Microsoft Excel 2010 software was used for the quantitative and qualitative analysis of the data. In addition to these two software, SPSS statistical software was used. The one-way ANOVA (Analysis of Variance - ANOVA) was used to compare Three Proficiency Levels in order to determine which indices showed statistically significant differences. The Multiple Regression Analysis (MRA) was also used to examine whether the indicators which are considered statistically significant were taken into account for the students's vocabulary assessment.

The results of the above measurements highlighted the inadequacy and weaknesses of

the existing scoring framework. Three of six lexical “richness” indices, the Average Tokens length (ATL), the Lambda (L) and the Percentage of Hapax Legomena (HL), were considered statistically significant. This fact confirms the heterogeneity among students who typically belong to B2 level and confirms the correlation of these indices with proficiency levels. The findings of the survey, also, indicated that three statistically significant indices were not taken into account in students’ vocabulary scoring.

These data have prompted the investigation of the Modern Greek Language Teaching Center’s current scoring framework and the Common European Framework of Reference’s (CEFR) grading framework. Although analytical scoring is adopted by both, there is no clear definition of vocabulary credits’ allocation (6 out of 20 points) and vocabulary’s quality and quantity corresponding to each Proficiency Level. Therefore, it demonstrates the need to revise and refine the current evaluation framework in order to include - at least - the three vocabulary «richness» indices that were considered statistically significant.

Περιεχόμενα

ΕΥΧΑΡΙΣΤΙΕΣ	iii
ΠΕΡΙΛΗΨΗ	iv
ABSTRACT	vi
Κατάλογος Πινάκων.....	3
Κατάλογος σχημάτων.....	4
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ	5
1.1. Επιστημονικό πεδίο.....	5
1.2. Ερευνητικά ερωτήματα και στόχοι.....	9
1.3. Δομή της διατριβής.....	11
ΚΕΦΑΛΑΙΟ 2: ΘΕΩΡΗΤΙΚΟ ΠΛΑΙΣΙΟ ΤΗΣ ΕΡΕΥΝΑΣ.....	13
2.1. Βασικές έννοιες της θεωρίας της κατάκτησης της δεύτερης γλώσσας	13
2.2. Η σημαντικότητα του λεξιλογίου για τους μαθητές της Γ2	15
2.2.1. Οι διαστάσεις του λεξιλογίου	16
2.2.2. Σχέση επιπέδων ΚΕΠΑ και λεξιλογίου	17
2.2.3. Η έννοια του λεξιλογικού πλούτου	20
2.2.4. Οι έξι δείκτες λεξιλογικού «πλούτου»	21
2.3. Γλωσσολογία βάσει ΗΣΚ.....	25
2.3.1. Ορισμός των ΗΣΚ.....	26
2.3.2. Προδιαγραφές των ΗΣΚ.....	26
2.3.3. Αντιπροσωπευτικότητα των ΗΣΚ	27
2.3.4. Περιορισμοί ΗΣΚ.....	28
2.3.5. Είδη ηλεκτρονικών σωμάτων κειμένων	28
ΚΕΦΑΛΑΙΟ 3: ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ	36
3.1. Περιγραφή του Σώματος Ελληνικών Μαθητικών Κειμένων (ΣΕΜΚ)	36
3.1.1. Περιεχόμενο και σχεδιαστικά κριτήρια.....	37
3.1.2. Υπολογιστικά εργαλεία επεξεργασίας του ΣΕΜΚ.....	56
3.2. Μέθοδοι διερεύνησης των ερευνητικών ερωτημάτων και υποθέσεων	61
3.2.1. Ερώτημα 1 ^ο : Οι στατιστικά σημαντικοί δείκτες λεξιλογικού πλούτου ανά επίπεδο γλωσσομάθειας.....	62
3.2.2. Ερώτημα 2 ^ο : Δείκτες λεξιλογικού «πλούτου» ισχύον πλαίσιο αξιολόγησης του μαθητικού κοινού	63

ΚΕΦΑΛΑΙΟ 4: ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΡΕΥΝΑΣ.....	64
4.1. Παραμετρικό στατιστικό κριτήριο one way ANOVA	64
4.1.1. Ποσοτικά αποτελέσματα	65
4.1.2. Ποιοτική επεξεργασία δεδομένων.....	69
4.2. Προβλεπτική εξίσωση πολλαπλής παλινδρόμησης.....	70
4.2.1. Ποσοτικά αποτελέσματα	70
4.2.2. Ποιοτική ανάλυση αποτελεσμάτων.....	73
ΚΕΦΑΛΑΙΟ 5: ΔΙΔΑΚΤΙΚΕΣ ΠΡΟΕΚΤΑΣΕΙΣ	75
5.1. Ισχύον πλαίσιο βαθμολόγησης του ΣΕΜΚ	75
5.1.1. Αξιολόγηση του ισχύοντος πλαισίου βαθμολόγησης του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας.....	76
5.1.2. Αξιολόγηση του ισχύοντος πλαισίου βαθμολόγησης του Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς για τη γλώσσα (ΚΕΠΑ).....	77
5.2. Συμπερίληψη στη βαθμολόγηση των στατιστικά σημαντικών δεικτών λεξιλογικού πλούτου	78
5.2.1. Η φύση των στατιστικά σημαντικών λεξιλογικών δεικτών και ο τρόπος συμπερίληψής τους στη βαθμολόγηση του λεξιλογίου	79
5.3. Αποτίμηση των προτάσεων	82
ΚΕΦΑΛΑΙΟ 6: ΣΥΜΠΕΡΑΣΜΑΤΑ.....	83
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	87
ΠΑΡΑΡΤΗΜΑ Α	106
ΠΑΡΑΡΤΗΜΑ Β	112
ΠΑΡΑΡΤΗΜΑ Γ.....	114

Κατάλογος Πινάκων

Πίνακας 1: Εμπειρική εκτίμηση από τον Milton του μεγέθους του λεξιλογίου για κάθε επίπεδο γλωσσομάθειας.....	19
Πίνακας 2: Γλωσσικές ομάδες των πληροφορητών/-τριών του ΣΕΜΚ και χώρα προέλευσης	41
Πίνακας 3: Χώρες προέλευσης των ερευνητικών υποκειμένων και το πλήθος των 224 ΠΓΛ που αντιστοιχούν σε κάθε χώρα.....	42
Πίνακας 4: Διαθέσιμες και μη διαθέσιμες μεταβλητές των πληροφορητών/-τριών και των ΠΓΛ του ΣΕΜΚ.....	45
Πίνακας 5: Βασικά ποσοστικά δεδομένα του αρχικού ΣΕΜΚ.....	46
Πίνακας 6: Βασικά ποσοστικά δεδομένα του ΣΕΜΚ των 224 ΠΓΛ.....	47
Πίνακας 7: Κατανομή των δεδομένων του ΣΕΜΚ των 224 ΠΓΛ αναφορικά με το επίπεδο αρχικής κατάταξης, το φύλο και τη χώρα προέλευσης.....	48
Πίνακας 8: Ποσοτικά δεδομένα των ευρύτερων γλωσσικών ομάδων του ΣΕΜΚ.....	49
Πίνακας 9: Κατανομή του γλωσσικού υλικού του ΣΕΜΚ ανά κειμενικό γένος.....	51
Πίνακας 10: Ποσοτική κατανομή των δεδομένων του ΣΕΜΚ ανά θέμα ΠΓΛ.....	53
Πίνακας 11: Δεδομένα QUIPA: Μέσος Όρος δεικτών ανά επίπεδο γλωσσομάθειας ...	65
Πίνακας 12: Ο δείκτης σημαντικότητας (sig.) για κάθε παράγοντα λεξιλογικού «πλούτου».....	66
Πίνακας 13: Post Hoc Tests: για τον δείκτη ATL.....	67
Πίνακας 14: Post Hoc Tests: για τον δείκτη Λ.....	68
Πίνακας 15: Post Hoc Tests: για τον δείκτη HL.....	68
Πίνακας 16: Η τιμή R^2 στα τρία επίπεδα γλωσσομάθειας.....	71
Πίνακας 17: Οι τιμές F και Significance για τα τρία επίπεδα γλωσσομάθειας.....	72
Πίνακας 18: Η τιμή Beta για τους έξι δείκτες.....	73
Πίνακας 19: Οι τιμές t & sig. ανά επίπεδο γλωσσομάθειας.....	73

Κατάλογος σχημάτων

Σχήμα 1: Τυπολογία των κειμενικών γενών των ΠΓΛ του ΣΕΜΚ.....50

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Στο συγκεκριμένο κεφάλαιο καθορίζονται το επιστημονικό πεδίο στο οποίο ανήκει η παρούσα εργασία και το ερευνητικό αντικείμενο. Αρχικά, παρουσιάζεται μία σύντομη ανασκόπηση της προϋπάρχουσας σχετικής έρευνας (ενότητα 1.1.), ακολουθούν εν συντομία τα ερευνητικά ερωτήματα και οι στόχοι της εργασίας (ενότητα 1.2.) και, τέλος, η δομή της (ενότητα 1.3.).

1.1. Επιστημονικό πεδίο

Το επιστημονικό πεδίο στο οποίο εντάσσεται η εργασία ανήκει στον ερευνητικό τομέα της γλωσσολογίας, γνωστό ως έρευνα βάσει ηλεκτρονικών σωμάτων μαθητικών κειμένων (Learner Corpus) (στο εξής ΗΣΜΚ)¹. Ο συγκεκριμένος τομέας συνδέει τη σωματοκειμενική γλωσσολογία (Corpus Linguistics)² και την έρευνα για την κατάκτηση της δεύτερης γλώσσας (Second Language Acquisition Research), δεδομένου ότι τα εν λόγω μαθητικά κείμενα ανήκουν σε μαθητές/-τριες των οποίων η ελληνική γλώσσα δεν είναι μητρική.

Η ενασχόληση των ερευνητών/-τριών με τη δυνατότητα αξιοποίησης των ΗΣΚ στην εφαρμοσμένη γλωσσολογία (Applied Linguistics) ξεκίνησε μόλις στα τέλη της δεκαετίας του '80, ύστερα από την αναγνώριση της αναγκαιότητας των Ηλεκτρονικών Σωμάτων Κειμένων (ΗΣΚ)³ σε εφαρμοσμένους τομείς της γλωσσολογίας, όπως η λεξικογραφία (Χαλισιανή, 2008).

¹ Ως Ηλεκτρονικά Σώματα Μαθητικών Κειμένων (ΗΣΜΚ) ορίζονται εκτεταμένες ψηφιοποιημένες συλλογές μαθητικών κειμένων, οι οποίες αξιοποιούνται τόσο για ερευνητικούς όσο και για εκπαιδευτικούς σκοπούς (Pravec, 2002· Granger, 2002· Granger, 2003· Τζιμώκας, 2010). Στην παρούσα έρευνα υιοθετήθηκε ο όρος Ηλεκτρονικά Σώματα Μαθητικών Κειμένων ή ΗΣΜΚ, χωρίς η εν λόγω επιλογή να συνιστά τελική πρόταση, καθώς ο όρος είναι πολυλεξικός. Ένας κοινά αποδεκτός και δηλωτικός όρος αποτελεί ακόμη ζητούμενο.

² Η γλωσσολογική έρευνα με ΗΣΚ εδραιώθηκε ως ιδιαίτερη μεθοδολογία επιστημονικής προσέγγισης της γλώσσας με την ονομασία corpus linguistics. Όσον αφορά την απόδοση του corpus linguistics υιοθετήθηκε ο όρος σωματοκειμενική γλωσσολογία που αποτελεί πρόσφατη πρόταση της Ακαδημίας Αθηνών.

³ Τα ηλεκτρονικά σώματα κειμένων (ΗΣΚ) αποτελούν δομημένες συλλογές κειμένων, αποθηκευμένες σε ηλεκτρονική μορφή και προσπελάσιμες με τη χρήση εργαλείων γλωσσικής τεχνολογίας. Υπάρχουν διαφορετικά είδη ΗΣΚ ανάλογα με τον όγκο τους, το συγχρονικό ή διαχρονικό τους χαρακτήρα, τα είδη λόγου που εκπροσωπούν, τα κριτήρια με βάση τα οποία δομούνται σε υποσύνολα (subcorpora), το

Την πρώτη συστηματική πρόταση για την αξιοποίηση τους στον διδακτικό κλάδο διατύπωσε ο Tim Johns, συνεργάτης του Sinclair στο COBUILD project, ο οποίος τόνισε ότι είναι καλό οι μαθητές/-τριες να έρχονται απευθείας αντιμέτωποι/-ες με τα αυθεντικά δεδομένα λόγου και να υιοθετούν το ρόλο του/της ερευνητή/-τριας της γλώσσας («to make the learner a linguistic researcher» Johns, 2002).

Με αυτόν τον τρόπο επιτυγχάνεται η αποδέσμευση της γλωσσικής διδασκαλίας από κανονιστικές προσεγγίσεις ή από προσεγγίσεις που βασίζονται σε διαισθήσεις (intuitions) των μελετητών ή των δασκάλων της γλώσσας («introspective views» Sinclair, 1997) και ενθαρρύνεται η αυτονομία του/της μαθητή/-τριας (learner autonomy) (Braun, 2005· Kaltenböck & Mehlmauer-Larcher, 2005). Τα οφέλη που προκύπτουν από τα ΗΣΚ μεγιστοποιούνται με τη συνδρομή των ΗΣΜΚ, δηλαδή ΗΣΚ τα οποία έχουν παραχθεί από μαθητές/-τριες (M). Από το πρωτογενές αυτό υλικό δεν επωφελούνται μόνο οι μαθητές/-τριες αλλά και οι ερευνητές/-τριες αφού αποτελεί την κυριότερη πηγή πληροφόρησης για τις διαγλώσσες του μαθητικού κοινού (interlanguages),⁴ παρέχοντας πληροφορίες για τους τρόπους και τα διαδοχικά στάδια κατάκτησης/εκμάθησης μιας γλώσσας αλλά και για τον σχεδιασμό και την αξιολόγηση της εκπαιδευτικής διαδικασίας (Κοιλιάρη, Αρχάκης & Τσάκωνα, 2017)

Τα ΗΣΜΚ αναδεικνύονται πολύ βοηθητικά και για τους/τις ίδιους/-ες τους/τις εκπαιδευτικούς ως μέσα ανίχνευσης των αναγκών των μαθητών/-τριών, προκειμένου να διαμορφώσουν τις διδακτικές τους παρεμβάσεις (Granger, 1997). Η χρησιμότητα των ΗΣΜΚ είναι πλέον αναγνωρισμένη και στην εκπαίδευση των εκπαιδευτικών οι οποίοι/-ες καλούνται να διδάξουν μία γλώσσα ως Γ2 (ή και ως Γ1), αφού η σχετική βιβλιογραφία τα αναγνωρίζει ως έναν από τους σημαντικότερους τρόπους εξοικείωσης των επιμορφούμενων εκπαιδευτικών με τις πραγματικές ανάγκες και δυσκολίες των μαθητών/-τριών. (O’Keefe & Farr, 2003· Mukherjee, 2004· Römer,

αν προέρχονται από το λόγο φυσικών ή μη φυσικών ομιλητών, αν είναι μονόγλωσσα ή δίγλωσσα κ.ά. (McEnery, Xiao & Tono, 2006: 1-76).

⁴Οι μαθητές μιας ξένης γλώσσας, πριν καταλήξουν σε μια επαρκή γνώση του συστήματος της Γ2 και αναπτύξουν ικανοποιητικά την επικοινωνιακή τους ικανότητα σε αυτή, περνούν από ποικίλες ενδιάμεσες μορφές αυτού του συστήματος-στόχου, οι οποίες αποτελούν μεν συνεχώς μεταβαλλόμενους και συχνά ιδιοσυγκρασιακούς, ταυτόχρονα δε εξίσου συστηματικούς κώδικες. Αυτοί οι κώδικες μπορούν να προσεγγιστούν όχι ως ελλειμματικές μορφές της Γ2, αλλά ως μια σειρά συστηματοποιήσεων που περιέχουν αθροίσματα μορφών, διαδικασιών και στρατηγικών στις οποίες καταφεύγει ο μαθητής κατά την εκμάθηση της πρόσθετης γλώσσας (James, 1998:7). Πρόκειται, στην πραγματικότητα, για υποθέσεις του μαθητή, των οποίων η ορθότητα επιβεβαιώνεται με τη δοκιμή τους (Τοκατλίδου, 1986: 93-94).

2008). Τα παραπάνω είναι μόνο κάποια από τα πολλά οφέλη και τις χρήσεις των ΗΣΜΚ (Granger, 1998· Milton, 1998· Römer, 2008· Τζιμώκας, 2010).

Ευνόητα, λοιπόν, όπως σημειώνουν οι Γούτσος & Κουτσουλέλου-Μίχου (2009), τα πλεονεκτήματα της χρήσης αυθεντικού και όχι επινοημένου γλωσσικού υλικού είναι πλέον γενικώς αποδεκτά από τη διδακτική κοινότητα.

Αξιοποιώντας τη γλωσσολογική ανάλυση βάσει των ΗΣΜΚ, η παρούσα εργασία επικεντρώνεται στο γραπτό γλωσσικό εξαγόμενο ενήλικων μαθητών/-τριών της νέας ελληνικής (NE) ως δεύτερης γλώσσας (Γ2), με ποικίλες μητρικές γλώσσες (στο εξής Γ1) οι οποίοι/-ες επιχειρούν καθοδηγούμενη εκμάθηση, δηλαδή παρακολουθούν γλωσσικά μαθήματα ελληνικών στην Ελλάδα. Ειδικότερα, εστιάζει στην ποσοτική ανάλυση συγκεκριμένων δεικτών λεξιλογικού «πλούτου» όπως παρουσιάζονται σε σχετικά σύντομα γραπτά κείμενα (180 έως 300 λέξεων). Τα κείμενα αυτά μπορούν τυπικά να θεωρηθούν επιπέδου B2, αφού αποτελούν απαντήσεις σε εξεταστικά θέματα αυτού του επιπέδου και επιλέχθηκαν μόνο όσα είχαν συγκεντρώσει το απαιτούμενο ποσοστό (τουλάχιστον 60%) που καθιστά κάποιον/-α προβιβάστέο/-α. Ωστόσο, μεταξύ αυτών παρουσιάζονται σημαντικές αποκλίσεις και δεν είναι τυχαίο το γεγονός ότι παράγονται από μαθητές/-τριες που προέρχονται και από τα τρία επίπεδα γλωσσομάθειας (αρχάριο, μέσο, προχωρημένο επίπεδο)⁵ της NEΓ2.

Κίνητρο για την ενασχόληση με το παρόν ερευνητικό αντικείμενο αποτέλεσε η απουσία συστηματικής καταγραφής και ανάλυσης των δεικτών λεξιλογικού «πλούτου» αναφορικά με τα γραπτά μαθητών/-τριών της ελληνικής ως Γ2 με τη χρήση εργαλείων Επεξεργασίας Φυσικής Γλώσσας και τα ασαφή κριτήρια βαθμολόγησης των ΗΣΜΚ, ειδικά αναφορικά με το λεξιλόγιο. Αυτοί οι δείκτες λεξιλογικού «πλούτου» (lexical richness indices) ή δείκτες λεξιλογικής διαφοροποίησης (lexical diversity indices) στοχεύουν με μια αθροιστική τιμή να αποτυπώσουν πόσο διαφορετικό λεξιλόγιο χρησιμοποιεί ένας/μία συγγραφέας στα έργα του/της και, επομένως, να προσεγγίσουν τον βαθμό του λεξιλογικού «πλούτου» που τον/την χαρακτηρίζει (Μικρός, 2015). Τα συγκεκριμένα λεξιλογικά γνωρίσματα εντάσσονται σ' έναν ευρύτερο διεπιστημονικό κλάδο,⁶ την υφομετρία, μέσω της

⁵ Απαιτείται η εντατική συνεργασία επιστημών όπως η Γλωσσολογία, η Επεξεργασία Φυσικής Γλώσσας, η Λογοτεχνική Ανάλυση, η Στατιστική, η Ανάκτηση Πληροφορίας, η Τεχνητή Νοημοσύνη και, ειδικότερα, ο κλάδος της Μηχανικής Μάθησης (Μικρός, 2015).

οποίας επιχειρείται ο αυτόματος προσδιορισμός του/της συγγραφέα ενός κειμένου. Παρά το γεγονός ότι οι δείκτες λεξιλογικού «πλούτου» είναι πολυάριθμοι και τα οφέλη που θα μπορούσαν να προσφέρουν στη διδακτική έρευνα είναι πολλά, δεν έχουν μελετηθεί και αξιοποιηθεί επαρκώς σε αυτόν τον κλάδο. Μια σημαντική, βέβαια, συμβολή των συγκεκριμένων δεικτών στον τομέα της διδακτικής είναι τα λογισμικά αναγνωσιμότητας, χάρη στα οποία κρίνεται η καταλληλότητα των κειμένων σε σχέση με τα επίπεδα γλωσσομάθειας και είναι ιδιαίτερα βοηθητικά για τους/τις διδάσκοντες/-ουσες.⁷

Δεδομένης της καίριας σημασίας της γνώσης του λεξιλογίου σε κάθε γλώσσα, η εξαγωγή ποσοτικών συμπερασμάτων που αφορούν παράγοντες λεξιλογικού «πλούτου» και στα τρία επίπεδα γλωσσομάθειας θα ενίσχυε την αξιοποίηση των δεικτών αυτών στο πλαίσιο της διδακτικής. Το συγκεκριμένο εγχείρημα βασίστηκε σε κείμενα μαθητών/-τριών της ελληνικής ως Γ2 που συνελέγησαν στο πλαίσιο του προγράμματος «Πυθαγόρας Ι» με φορέα υλοποίησης το Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών (2004 - 2007). Το εν λόγω πόνημα αποτελεί πρόσφορο έδαφος για περαιτέρω μελέτη δεδομένης της ελληνικής πραγματικότητας αναφορικά με τα ΗΣΚ. Αρκεί μια σύντομη εξέταση των υπαρχόντων ΗΣΚ για την ελληνική γλώσσα, όπως ο ΕΘΕΓ και το ΣΕΚ για να διαπιστωθεί ότι κατά κανόνα περιλαμβάνουν γλωσσικό υλικό προερχόμενο από φυσικούς ομιλητές της ελληνικής (Γούτσος, 2002· Γούτσος, 2003· Γούτσος, 2007· Ιακώβου, Μαρκόπουλος & Μικρός, 2008· Γούτσος, Σηφianού & Γεωργακοπούλου, 2005· Κουτσουλέλου & Μικρός, 2005· Σαββίδου, 2012). Πέραν τούτου, οι περισσότερες μελέτες που έχουν γίνει κατά την τελευταία εικοσαετία με υποκείμενα μη φυσικούς ομιλητές της ελληνικής, μαθητές/-τριες και μη, αφορούν διαφορετικά κάθε φορά γραμματικά φαινόμενα και τις περισσότερες φορές βασίζονται σε εμπειρικά δεδομένα που συλλέγονται ad hoc, δεν προέρχονται από ΗΣΚ που έχουν συγκροτηθεί με βάση συγκεκριμένα κριτήρια και με συστηματικό τρόπο (Μαγγανά, 1999· Θώμου, 2001· Αντωνοπούλου, Βαλετόπουλος, Καρακύργιου, Μουμτζή & Παναγιωτίδου, 2006· Κόντα, 2012). Η παραπάνω έλλειψη, βέβαια, δεν αποτελεί μόνο ελληνικό φαινόμενο, καθώς η ανάπτυξη σχετικών ΗΣΚ ακόμη και για ευρέως χρησιμοποιούμενες και διδασκόμενες γλώσσες, όπως η αγγλική, μόλις τα τελευταία δεκαπέντε με είκοσι χρόνια έχει προχωρήσει

⁷ Χαρακτηριστικό παράδειγμα εργαλείου αναγνωσιμότητας για την ελληνική είναι αυτό που προτείνεται από το Κέντρο Ελληνικής Γλώσσας (ΚΕΓ) στον ακόλουθο σύνδεσμο: <http://www.greek-language.gr/certification/readability/index.html>

συστηματικότερα (Granger, Gilquin & Meunier, 2013). Όσον αφορά τα νέα ελληνικά, πρέπει να σημειωθεί ότι την τρέχουσα δεκαετία αναδύθηκαν οι πρώτες αξιόλογες προσπάθειες ανάπτυξης αντίστοιχων ΗΣΜΚ⁸. Μια τέτοια προσπάθεια συνιστά και το ΗΣΜΚ που αξιοποιείται στην παρούσα εργασία

Επομένως, με την παρούσα εργασία επιχειρείται η προώθηση της γλωσσολογικής έρευνας και των εκπαιδευτικών εφαρμογών αυτής. Επιδιώκοντας τη διερεύνηση της πιθανής συσχέτισης των δεικτών λεξιλογικού «πλούτου» και των επιπέδων γλωσσομάθειας, από τους πολυάριθμους παράγοντες λεξιλογικού «πλούτου», η έρευνα επικεντρώθηκε σε έξι παράγοντες για τους οποίους θα γίνει εκτενέστερη αναφορά στη συνέχεια: το Ποσοστό Άπαξ Λεγομένων ή Harax Legomenon Percentage (HL), το R1, το Lambda (Λ), το h-point (h), την Εντροπία ή Entropy (H) και το Μέσο Μήκος Λεξικού Τύπου ή Average Tokens length (ATL). Σε δεύτερο χρόνο, θα επιδιωχθεί να διαπιστωθεί αν οι δείκτες που αποκαλύφθηκε ότι σχετίζονται με το επίπεδο γλωσσομάθειας λήφθηκαν υπόψη κατά τη βαθμολόγηση του λεξιλογίου του συγκεκριμένου ΗΣΜΚ. Βάσει των αποτελεσμάτων από την επεξεργασία των δεδομένων, θα ακολουθήσει σχετική διδακτική πρόταση.

1.2. Ερευνητικά ερωτήματα και στόχοι

Στο πλαίσιο της παρούσας έρευνας θα αξιοποιηθούν τα δεδομένα από την ανάλυση έξι δεικτών λεξιλογικού «πλούτου» που ανιχνεύονται σε ΗΣΜΚ της ΝΕΓ2, προκειμένου να διερευνηθούν δύο βασικά ερευνητικά ερωτήματα που σχετίζονται άμεσα με την γλωσσολογική έρευνα και τη διδασκαλία.

Το πρώτο ερευνητικό ερώτημα αφορά το ενδεχόμενο οι επιλεγμένοι δείκτες λεξιλογικού «πλούτου» να σχετίζονται με τα επίπεδα γλωσσομάθειας από τα οποία

⁸ Όσον αφορά τα ελληνικά ως δεύτερη ή ξένη γλώσσα, η ανάπτυξη ΗΣΜΚ βρίσκεται προς το παρόν σε σχετικά πρώιμο στάδιο. Συγκεκριμένα, μέχρι στιγμής έχουν δημιουργηθεί τέσσερα ΗΣΜΚ της ΝΕΓ2 στο πλαίσιο αντίστοιχων ερευνητικών προγραμμάτων: α) το ΗΣΚ εκμάθησης της ελληνικής γλώσσας του προγράμματος «Πυθαγόρας Ι» με φορέα υλοποίησης το Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών (2004 - 2007), β) το ΗΣΚ των παιδιών της μουσουλμανικής μειονότητας της Θράκης του προγράμματος «Εκπαίδευση των παιδιών της μουσουλμανικής μειονότητας στη Θράκη» (Τζεβελέκου κ.ά., 2008), γ) το «Ελληνικό Σώμα Κειμένων Μαθητών (ΕΣΚΕΙΜΑΘ)» με ΠΓΛ παιδιών μεταναστών και παλιννοστούντων του προγράμματος «Εκπαίδευση αλλοδαπών και παλιννοστούντων μαθητών» (Παπαδοπούλου & Τάντος, 2014· Τάντος κ.ά., 2015) και δ) το «Σώμα Εξελικτικών Παραγωγών Μαθητών της Ελληνικής ως Γ2 (ΣΕΠΑΜΕ2)» με παραγωγές γραπτού και προφορικού λόγου από μαθητές οι οποίοι φοιτούν στο Διδακταλείο της Νέας Ελληνικής ως δεύτερης/ξένης γλώσσας του ΕΚΠΑ και συμμετέχουν στα ειδικά τμήματα ενίσχυσης των παραγωγικών τους δεξιοτήτων (βλ. Ιακώβου κ.ά., 2017) (Τζιμώκας, 2018).

προέρχονται οι μαθητές/-τριες. Στόχος είναι να εντοπιστούν οι δείκτες που παρουσιάζουν στατιστικά σημαντική διαφορά μεταξύ των τριών επιπέδων (αρχάριο, μέσο, προχωρημένο) και κατ' επέκταση τα επίπεδα μεταξύ των οποίων παρουσιάζεται η απόκλιση για κάθε δείκτη. Σύμφωνα με το Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς (ΚΕΠΑ) για την Εκμάθηση, τη Διδασκαλία και την Αξιολόγηση των Γλωσσών (Council of Europe 2008), η γλωσσομάθεια σε μια δεύτερη γλώσσα διαβαθμίζεται σε έξι επίπεδα, αλλά κατ' ουσίαν πρόκειται για τρία,⁹ με τις αντίστοιχες αποδόσεις τους στην Ελληνική.¹⁰ Στην ίδια κλίμακα του ΚΕΠΑ διαβαθμίζονται και οι επιδόσεις των μαθητών/-τριών ως προς την ανάπτυξη του λεξιλογίου στη Γ2 βάσει δύο μαθησιακών παραμέτρων: η πρώτη σχετίζεται με το εύρος της λεξικής γνώσης που είναι σε θέση να κατέχει ένας/μία μαθήτρια/-τρια ανά επίπεδο γλωσσομάθειας αναφορικά με τη γλώσσα-στόχο και η δεύτερη με το βάθος της γνώσης¹¹ που είναι σε θέση να ελέγξει και να ενεργοποιήσει ανά επίπεδο γλωσσομάθειας για την κάλυψη των επικοινωνιακών αναγκών που προκύπτουν από αυτό (Ιακώβου, 2009). Διερευνώντας τους συγκεκριμένους δείκτες λεξιλογικού «πλούτου», προβλέπεται να εμπλουτίσουμε τις γνώσεις μας για το εύρος της λεξικής γνώσης ανά επίπεδο γλωσσομάθειας και να συσχετίσουμε τα ευρήματα με τις

⁹ Σύμφωνα με το Κοινό Ευρωπαϊκό Πλαίσιο (ΚΕΠ) για την Εκμάθηση, τη Διδασκαλία και την Αξιολόγηση των Γλωσσών (Council of Europe 2008), η γλωσσομάθεια σε μια δεύτερη γλώσσα κλιμακώνεται σε έξι επίπεδα. Στην ουσία πρόκειται για τρία επίπεδα (Α,Β,Γ) με δύο διαβαθμίσεις στο έκαστο. Στο Α επίπεδο (αρχάριο) ο χρήστης θεωρείται αρχάριος και διακρίνεται σε Α1 (στοιχειώδες επίπεδο) και Α2 (εισαγωγικό επίπεδο). Στο Β επίπεδο (μέσο) ο χρήστης χαρακτηρίζεται ανεξάρτητος και χωρίζεται σε Β1 (βασικό επίπεδο) και Β2 (επίπεδο επάρκειας). Τέλος, στο Γ επίπεδο (προχωρημένο) πλέον ο χρήστης θεωρείται ικανός και πάλι χωρίζεται σε δύο επιμέρους επίπεδα, το Γ1 (προχωρημένο) και το Γ2 (επίπεδο επάρκειας). Αντίστοιχα, καθορίζεται και το επίπεδο γνώσης σε κάθε επίπεδο : Α1 (στοιχειώδης γνώση), Α2 (βασική γνώση), Β1 (μέτρια γνώση), Β2 (καλή γνώση), Γ1 (πολύ καλή γνώση), Γ2 (άριστη γνώση). Σχετικά με τα επίπεδα γλωσσομάθειας στην ελληνική ως δεύτερη γλώσσα, βλ. Βαρλοκώστα & Τριανταφυλλίδου (2011).

¹⁰ Σχετικά με την ορολογία των επιπέδων γλωσσομάθειας στην Ελληνική, βλ. Ιακώβου (2003) και Κοντός κ.ά. (2001).

¹¹ Σχετικά με τις έννοιες του 'εύρους' (width/ breadth) και του 'βάθους' (depth) της λεξικής γνώσης, βλ. Henricksen (1999). Στην πρώτη περίπτωση, για καθεμία πλευρά της λεξικής γνώσης χωριστά (π.χ. προφορά/ σημασία/ χρήση...) ο κάθε μαθητής κινείται πάνω σε ένα συνεχές που κυμαίνεται από το 0 (απόλυτη άγνοια), περνάει από το μερικώς ακριβές (μερική γνώση) για να φθάσει στο 1, το απόλυτα ακριβές (πλήρης γνώση). Η δεύτερη περίπτωση σχετίζεται με τον βαθμό που ελέγχει σε έναν κατακόρυφο άξονα ο μαθητής όλες τις πλευρές της λεξικής γνώσης στο σύνολό τους (π.χ. και προφορά και σημασία και χρήση...) προκειμένου να τις ενεργοποιήσει στη γλωσσική του παραγωγή και να να αποφυγεί σε μικρότερο ή μεγαλύτερο βαθμό τα σχετικά λεξικά λάθη. Η διάκριση αυτή μπορεί να θεωρηθεί ότι συνάδει με τη 'γνώση' και τον 'έλεγχο' για τα οποία κάνουν για πρώτη φορά λόγο οι Bialystock & Sharwood Smith (1985). Η μεν 'γνώση' αναφέρεται στην πρώτη περίπτωση, καθώς σχετίζεται με τον τρόπο με τον οποίο το γλωσσικό σύστημα αναπαρίσταται στον νου του μαθητή (διευρυμένη ή μη γλωσσική ποικιλία), ενώ ο έλεγχος είναι 'το σύστημα επεξεργασίας που είναι σε θέση να ελέγξει τη γνώση στη διάρκεια της γλωσσικής της πραγμάτωσης' (απόλυτος ή μερικός έλεγχος/ εις βάθος γνώση).

προσδοκίες που υπαγορεύονται από το ΚΕΠΑ αναφορικά με το λεξιλόγιο σε κάθε επίπεδο γλωσσομάθειας.

Το δεύτερο ερευνητικό ερώτημα απορρέει από τα ευρήματα του πρώτου και σχετίζεται άμεσα με τη διδακτική πραγματικότητα. Εφόσον αξιολογηθούν οι επιλεγμένοι δείκτες λεξιλογικού «πλούτου» σε κάθε επίπεδο γλωσσομάθειας, κάποιος ενδέχεται να κριθούν στατιστικά σημαντικοί, να παρουσιάσουν, δηλαδή, διαφοροποίηση ανά επίπεδο. Σε αυτή την περίπτωση, ζητούμενο είναι να εξετασθεί αν οι στατιστικά σημαντικοί δείκτες λήφθηκαν υπόψη στη βαθμολόγηση του λεξιλογίου των μαθητικών κειμένων. Εάν αποδειχθεί ότι δε λήφθηκαν υπόψη, κρίνεται αναγκαία η αναθεώρηση του ισχύοντος πλαισίου, προκειμένου η βαθμολόγηση του λεξιλογίου να είναι πιο αντικειμενική και αντιπροσωπευτική των δυνατοτήτων των μαθητών/-τριών σε σχέση με το γνωστικό τους επίπεδο.

Φυσικά, η διερεύνηση των έξι δεικτών λεξιλογικού «πλούτου» δεν αξιώνει τη διαμόρφωση ενός ολοκληρωμένου πλαισίου αξιολόγησης. Θα μπορούσε, ωστόσο, να αποτελέσει πρόδρομο για μία ολιστική προσέγγιση των δεικτών αυτών, συμβάλλοντας στην επανεξέταση του προβλεπόμενου για κάθε επίπεδο λεξιλογίου και στη διαμόρφωση σαφέστερων κριτηρίων αξιολόγησης.

1.3. Δομή της διατριβής

Στο πρώτο εισαγωγικό κεφάλαιο που προηγήθηκε παρουσιάστηκε το πεδίο εστίασης (1.1.) και τα βασικά ερευνητικά ερωτήματα (1.2.). Το δεύτερο κεφάλαιο που ακολουθεί συνιστά το θεωρητικό μέρος της παρούσας εργασίας. Σε αυτό αναλύονται βασικές έννοιες στις οποίες στηρίχθηκε η παρούσα έρευνα όπως είναι η διαγλώσσα, η υφομετρία, οι δείκτες λεξιλογικού «πλούτου» και δίνεται έμφαση στη γλωσσολογία βάσει ΗΣΚ, στη χρήση εργαλείων Επεξεργασίας Φυσικής Γλώσσας και στις πρακτικές εφαρμογές των ΗΣΜΚ. Στο τρίτο κεφάλαιο ξεκινά το ερευνητικό μέρος της εργασίας, όπου περιγράφονται το ΣΕΜΚ, οι σχεδιαστικές προδιαγραφές του και τα υπολογιστικά εργαλεία που χρησιμοποιήθηκαν για τον εντοπισμό των δεικτών λεξιλογικού «πλούτου» και την επεξεργασία των ευρημάτων. Επίσης, παρουσιάζονται αναλυτικά οι μέθοδοι διερεύνησης των ερευνητικών ερωτημάτων της εργασίας. Ακολουθεί το τέταρτο κεφάλαιο όπου παρατίθενται τα αποτελέσματα των μετρήσεων που διενεργήθηκαν στο ερευνώμενο ΗΣΜΚ στο πλαίσιο της εξέτασης

των δύο ερευνητικών ερωτημάτων που προηγήθηκαν. Σε αυτό γίνεται εμφανές ποιους λεξιλογικοί παράγοντες διαφοροποιούνται ανά επίπεδο γλωσσομάθειας και διαπιστώνεται αν λήφθηκαν υπόψη από τους αξιολογητές. Στο πέμπτο κεφάλαιο αποτυπώνονται κάποιες προτάσεις αναφορικά με τον διδακτικό τομέα βάσει των αποτελεσμάτων που προέκυψαν από την επεξεργασία του ΗΣΜΚ. Τέλος, στο έκτο κεφάλαιο, με το οποίο ολοκληρώνεται η διπλωματική εργασία, καταγράφονται τα κυριότερα συμπεράσματα που συνάγονται από τα αποτελέσματα της ποσοτικής και ποιοτικής ανάλυσης των ερευνητικών δεδομένων και επισημαίνεται η σημασία τους για την έρευνα και τη διδασκαλία της ελληνικής ως Γ2.

ΚΕΦΑΛΑΙΟ 2: ΘΕΩΡΗΤΙΚΟ ΠΛΑΙΣΙΟ ΤΗΣ ΕΡΕΥΝΑΣ

Η πρώτη ενότητα του συγκεκριμένου κεφαλαίου (2.1.) αναφέρεται στη διασάφηση ορισμένων θεμελιωδών εννοιών σχετικών με τη γλωσσική κατάκτηση/εκμάθηση και κατ' επέκταση με τη γλωσσική διδασκαλία. Στη δεύτερη ενότητα (2.2.), επισημαίνεται η σημαντικότητα του λεξιλογίου για το μαθητικό κοινό, αποσαφηνίζεται η έννοια του λεξιλογικού «πλούτου» και περιγράφονται οι έξι δείκτες λεξιλογικού «πλούτου» στους οποίους εστιάζει η παρούσα εργασία. Στο επόμενο κεφάλαιο (2.3.), γίνεται λόγος για τη συμβολή των Ηλεκτρονικών Σώματων Κειμένων (ΗΣΚ) στη Γλωσσολογία, ενώ εκτενής είναι η αναφορά στα Ηλεκτρονικά Σώματα Μαθητικών Κειμένων (ΗΣΜΚ), υποσύνολο των ΗΣΚ, για τη συμβολή τους τόσο στην έρευνα όσο και στη διδασκαλία της δεύτερης γλώσσας.

2.1. Βασικές έννοιες της θεωρίας της κατάκτησης της δεύτερης γλώσσας

Για την ουσιαστικότερη κατανόηση της έννοιας του λεξιλογικού πλούτου στο πλαίσιο των ΗΣΜΚ, θα ήταν χρήσιμο να γίνει σύντομη αναφορά σε μερικές βασικές έννοιες του κλάδου της κατάκτησης της δεύτερης γλώσσας. Η Μπέλλα (2011:21-26) προσδιορίζει εννέα βασικές έννοιες της θεωρίας της κατάκτησης της δεύτερης γλώσσας, από τις οποίες οι πέντε σχετίζονται άμεσα με το αντικείμενο της παρούσας έρευνας και παρατίθενται κάτωθι. Οι έννοιες αυτές είναι οι εξής:

α) Δεύτερη και ξένη γλώσσα: η διάκριση μεταξύ δεύτερης και ξένης γλώσσας είναι κατά βάση γεωγραφική. Δεύτερη χαρακτηρίζεται συνήθως μία γλώσσα η οποία κατακτάται σε περιβάλλον όπου χρησιμοποιείται ως κύρια γλώσσα, ενώ ξένη θεωρείται μία γλώσσα που κάποιος κατακτά στο περιβάλλον της μητρικής του. Δεδομένης της ασάφειας των γεωγραφικών ορίων μιας γλώσσας, καθώς σε έναν τόπο μπορεί να ομιλούνται πολλές διαφορετικές κυρίαρχες γλώσσες (π.χ. Βέλγιο, Ελβετία), η Μπέλλα (2011:21) καταλήγει στη χρήση του όρου «δεύτερη» (στο εξής Γ2) για την «εκμάθηση οποιασ(οποίων)δήποτε γλώσσας(ών) πέραν της μητρικής». Το πρωτογενές υλικό της παρούσας έρευνας συνιστά προϊόν εκμάθησης μιας Γ2 εντός

ενός γεωγραφικού χώρου όπου συνιστά την κυρίαρχη γλώσσα. Πρόκειται για την εκμάθηση ελληνικών στην Ελλάδα.

β) Κατάκτηση και εκμάθηση γλώσσας: η πρώτη έννοια αναφέρεται στην εξοικείωση με μια δεύτερη γλώσσα μέσα από τη φυσική επαφή με αυτήν κατά την καθημερινή γλωσσική επικοινωνία, ενώ η εκμάθηση σημαίνει τη διαδικασία εξοικείωσης με έναν γλωσσικό κώδικα, πέραν του μητρικού, μέσω συνειδητής διδασκαλίας. Και σε αυτή την περίπτωση, η Μπέλλα (2011:21-22) προτείνει τη χρήση του όρου «κατάκτηση» και για τις δύο περιπτώσεις. Η επιλογή αυτή υιοθετείται και σε αυτή την εργασία, όπου γίνεται λόγος μόνο για γραπτά κείμενα μαθητών που παράγονται στο πλαίσιο της τελικής αξιολόγησής τους μετά από την παρακολούθηση εντατικού προγράμματος εκμάθησης της ΝΕΓ2.

γ) Γλωσσική ικανότητα και γλωσσική πραγμάτωση: η πρώτη έννοια αφορά τις νοητικές αναπαραστάσεις κανόνων που συνιστούν την εσωτερική γραμματική του/της κάθε ομιλητή/-τριας και αντανακλάται στις διαισθήσεις του/της σχετικά με τη γραμματικότητα των προτάσεων. Η γλωσσική πραγμάτωση, από την άλλη, συνιστά τη χρήση αυτής της εσωτερικής γραμματικής για την κατανόηση και την παραγωγή της γλώσσας. Ως χρήση υπόκειται στους περιορισμούς της μνήμης, την κόπωση, τις κοινωνικές συμβάσεις και τον χαρακτήρα των επικοινωνιακών περιστάσεων. Η παρούσα έρευνα εστιάζει στη γλωσσική πραγμάτωση και συγκεκριμένα σε ορισμένους δείκτες λεξιλογικού «πλούτου» οι οποίοι εντοπίζονται σε γραπτά αλλόγλωσσων.

δ) Γλωσσικό εισαγόμενο και αφομοιωμένο γλωσσικό εισαγόμενο: το γλωσσικό εισαγόμενο συνιστά όλα τα στοιχεία στα οποία εκτίθεται ο/η μαθητής/-τρια μιας Γ2 είτε κατά την αυθόρμητη επικοινωνία είτε κατά τη συστηματική καθοδήγηση, ενώ το αφομοιωμένο γλωσσικό εισαγόμενο αναφέρεται στο μέρος του γλωσσικού εισαγόμενου που κατακτάται και ενσωματώνεται στο γλωσσικό σύστημα του/της μαθητή/-τριας. Πρόκειται για φαινόμενο που συμβάλλει αποφασιστικά στην πορεία της γλωσσικής κατάκτησης και στη διαμόρφωση της διαγλώσσας. Κατ' αναλογία με το γλωσσικό εισαγόμενο χρησιμοποιείται ο όρος γλωσσικό εξαγόμενο, για να γίνει αναφορά στη γλώσσα που τελικά παράγει ο/η μαθητής/-τρια της Γ2. Στη δεδομένη έρευνα η εστίαση γίνεται στο γλωσσικό εξαγόμενο, όπως αποτυπώνεται στα γραπτά των μαθητών/-τριών.

ε) Ατομικές διαφορές: σχετίζονται με τους ποικίλους γνωστικούς ή συναισθηματικούς παράγοντες που επηρεάζουν την πορεία κατάκτησης, την ταχύτητα κατάκτησης και την επιτυχία. Οι κυριότεροι γνωστικοί παράγοντες είναι η ηλικία και η έφεση, ενώ οι κυριότεροι συναισθηματικοί παράγοντες είναι τα κίνητρα του/της μαθητή/-τριας και η στάση του/της απέναντι στη Γ2 και στους φορείς της¹². Στην παρούσα έρευνα, δεν θα λέγαμε ότι υπάρχουν εντυπωσιακές ατομικές διαφορές μεταξύ των υποκειμένων του δείγματος, καθώς η συντριπτική τους πλειοψηφία είναι ενήλικες γύρω στα 20 έτη με ιδιαίτερα ισχυρό κίνητρο για την εκμάθηση της ελληνικής που θα τους βοηθήσει στη συνέχιση των σπουδών τους σε ελληνικά πανεπιστήμια¹³.

2.2. Η σημαντικότητα του λεξιλογίου για τους μαθητές της Γ2

Όπως επισημαίνεται από τους Kilgarriff, Charalabopoulou et al. (2014), μέχρι πρόσφατα το λεξιλόγιο δεν αποτελούσε προτεραιότητα στον τομέα της διδασκαλίας της δεύτερης γλώσσας. Αντίθετα, θεωρούνταν κυρίως δευτερεύουσα δραστηριότητα, αφού το διδακτικό προσωπικό έδινε έμφαση στην ανάπτυξη της επικοινωνιακής και γλωσσικής, δηλαδή γραμματικής, ικανότητας. Πλέον, ο πρωταρχικός ρόλος του λεξιλογίου στη Γ2 έχει αναγνωριστεί και η γνώση του λεξιλογίου θεωρείται απαραίτητη προϋπόθεση για την απόκτηση των τεσσάρων παραδοσιακών γλωσσικών δεξιοτήτων, δηλαδή την ακρόαση, την ομιλία, την ανάγνωση, τη γραφή αλλά και για τη γραμματική.

Η προτεραιότητα του λεξιλογίου σε σχέση με άλλους τομείς στην εκμάθηση μιας Γ2 επιβεβαιώνεται και από τον Nation (2001), ο οποίος τονίζει πως η κατανόηση και παραγωγή της γλώσσας εξαρτώνται σε μεγάλο βαθμό από το μέγεθος του λεξιλογίου, με 3.000 οικογένειες λέξεων να θεωρείται ένα κρίσιμο όριο. Ως εκ τούτου, μια συστηματική και βασισμένη σε αρχές προσέγγιση του λεξιλογίου έχει ως επακόλουθο την καλύτερη γνώση μιας γλώσσας.

¹² Για μια εκτενή παρουσίαση των παραγόντων που καθορίζουν τις ατομικές διαφορές βλ. Μπέλλα 2011: 25-26.

¹³ Βλ. ενότητα 3.1. για περισσότερες λεπτομέρειες σχετικά με τα δημογραφικά και άλλα χαρακτηριστικά των πληροφορητών της παρούσας έρευνας.

Εντούτοις, σύμφωνα με τον Bonk (2000), δεν υπάρχει απόλυτο κατώτατο λεξικό όριο, όπως προτάθηκε από τον Nation (2001), αλλά, σε σχετική έρευνα, διαπιστώθηκε ότι οι μαθητές/-τριες που γνώριζαν λιγότερο από το 80% του συνόλου των λέξεων που εμπεριείχονταν στα αποσπάσματα στα οποία εξετάστηκαν είχαν σχεδόν πάντα ανεπαρκή κατανόηση, ενώ το 60% των συμμετεχόντων που γνώριζαν περισσότερο από το 90% των λέξεων πέτυχαν ικανοποιητική κατανόηση των κειμένων. Τελικά, τα πορίσματα της έρευνας έδειξαν ότι για την πλειοψηφία των συμμετεχόντων η ικανοποιητική κατανόηση απέρρεε από τη γνώση του 95% ή και περισσότερο των τρεχουσών λέξεων. Σύμφωνα με τον Schmitt (2000), ο αριθμός λέξεων που χρειάζεται ένας/μία μαθητής/-τρια εξαρτάται από τον τελικό του στόχο. Συγκεκριμένα, έχει ανάγκη περίπου από 2000 λέξεις για επικοινωνιακούς σκοπούς, 3000 λέξεις, για να αρχίσει κανείς να διαβάζει αυθεντικά κείμενα, ίσως έως και 10000 λέξεις για να ανταποκριθεί κανείς σε ακαδημαϊκά κείμενα και 15000-20000 λέξεις, για να φτάσει στο επίπεδο ενός μορφωμένου φυσικού ομιλητή.

2.2.1. Οι διαστάσεις του λεξιλογίου

Τα παραπάνω ποσοστά αναδεικνύουν τον καίριο ρόλο του λεξιλογίου στην κατάκτηση μιας Γ2 και τον τεράστιο αριθμό λεξικών στοιχείων που θεωρείται αναγκαίο να μάθει ένας/μία μαθητής/-τρια. Τα ποσοστά αυτά είναι ακόμα πιο υψηλά, αν συνυπολογιστούν και οι λεξικές φράσεις, οι οποίες δεν λήφθηκαν υπόψη (Schmitt, 2004).

Στο σημείο αυτό, κρίνεται απαραίτητο να καθοριστεί τι νοείται ως λεξιλογική γνώση. Όταν γίνεται λόγος για γνώση λεξιλογίου είναι χρήσιμο να ορίζονται οι διαστάσεις του, έτσι όπως τις διακρίνουν οι Anderson και Freebody (1981:77–117). Πρόκειται για το λεξιλογικό εύρος, δηλαδή τον αριθμό των λέξεων που γνωρίζει ένας/μία μαθητής/-τρια, ανεξάρτητα από τη μορφή στην οποία τις γνωρίζει, και το λεξιλογικό βάθος, δηλαδή το πόσο καλά ή πόσο ολοκληρωμένα κατέχει αυτές τις λέξεις. Αυτό συχνά αναφέρεται ως ποιότητα ή βάθος της λεξιλογικής γνώσης (quality or ‘depth’ of vocabulary knowledge) και είναι τόσο σημαντικό όσο και το μέγεθος του (Schmitt, 2008:329-363). Ο Meara (1996) προσθέτει μία τρίτη διάσταση στις ήδη υπάρχουσες, αυτή της ευχέρειας, δηλαδή του αυτοματισμού με τον οποίο ένα άτομο μπορεί να αναγνωρίζει και να χρησιμοποιεί στο λόγο κάποιες λέξεις. Επομένως, πέρα από τον

αριθμό των λέξεων, ένας/μια μαθητής/-τρια πρέπει να γνωρίζει σε βάθος το κάθε λεξικό στοιχείο και να μπορεί να το χρησιμοποιεί με άνεση.

Η σημαντικότητα του λεξιλογίου για τη διδασκαλία μιας Γ2 αναδεικνύεται και από τον Nation (2001), ο οποίος δομεί το μάθημα μέσα από μία ισορροπημένη τετραμερή προσέγγιση (four-strand approach) για την εκμάθηση νέων λεξικών στοιχείων: το γλωσσικό εισαγόμενο που εστιάζει στη σημασία (meaning-focused input), το λεξικό εξαγόμενο που εστιάζει στη σημασία (meaning focused output), την εστίαση στον τύπο (language-focused learning) και την ανάπτυξη της ευχέρειας (fluency development).

Παρά την ανάδειξη της σημαντικότητας του λεξιλογίου στη Γ2, δεν έχει διαλευκανθεί η ποιότητα και η ποσότητα που αντιστοιχεί σε κάθε επίπεδο γλωσσομάθειας. Ακόμα και από το ΚΕΠΑ (Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς) δεν υπάρχει σαφής προσδιορισμός.

2.2.2. Σχέση επιπέδων ΚΕΠΑ και λεξιλογίου

Στις αρχές της δεκαετίας του '90, το Συμβούλιο της Ευρώπης ανέλαβε την πρωτοβουλία για την ανάπτυξη της διδασκαλίας ξένων γλωσσών στα νέα κράτη-μέλη, μέσα από το έργο της «Εκμάθηση γλωσσών για την Ευρωπαϊκή Ιθαγένεια». Η πολιτική της γλωσσικής εκπαίδευσης του Συμβουλίου της Ευρώπης επικεντρώνεται στην ιδέα της πολλαπλο-γλωσσικής και πολλαπλο-πολιτισμικής ικανότητας του ατόμου την οποία ορίζει ως «την εκπαιδευτική αξία που αποτελεί τη βάση για τη γλωσσική ανοχή» και ως «ένα σημαντικό στοιχείο διαπολιτισμικής εκπαίδευσης» (Beacco & Byram, 2007).

Το αρχικό κείμενο εκδόθηκε το 1995 και στην ανανεωμένη έκδοση αναφέρεται χαρακτηριστικά ότι «το Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς (στο εξής ΚΕΠΑ) ¹⁴

¹⁴ Συχνά το ΚΕΠΑ συναντάται ως ΚΕΠ. Η ολοκληρωμένη ονομασία του ΚΕΠΑ είναι «Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς για τις Γλώσσες: Μάθηση, Διδασκαλία, Αξιολόγηση». Δυστυχώς πολλές φορές παραλείπεται ο όρος «Αναφοράς». Δεδομένου ότι ο αρχικός σχεδιασμός του ΚΕΠ αποσκοπούσε στο να λειτουργεί σαν πλαίσιο αναφοράς σύμφωνα με το οποίο να μπορούν να περιγραφούν τα διαφορετικά προσόντα, να μπορούν να αναγνωριστούν διαφορετικοί στόχοι εκμάθησης ξένων γλωσσών και να μπορούν να οριστούν διαφορετικά πρότυπα επίτευξης των σκοπών αυτών, η χρήση του όρου «αναφορά» στην ονομασία του ΚΕΠ κρίνεται απαραίτητη (Morrow, 2004) και συνεπώς η αντικατάσταση του ακρώνυμου ΚΕΠΑ από το ΚΕΠ καθίσταται αυτονόητη.

παρέχει μια κοινή βάση για την ανάπτυξη προγραμμάτων διδασκαλίας γλωσσικών μαθημάτων, οδηγιών για τη σύνταξη σχετικών προγραμμάτων σπουδών, εξετάσεων, διδακτικών εγχειριδίων για ολόκληρη την Ευρώπη. Περιγράφει με περιεκτικό τρόπο τι πρέπει να μάθει να κάνει ο/η μαθητής/-τρια μιας ξένης γλώσσας, για να χρησιμοποιήσει μια γλώσσα για επικοινωνία, καθώς και ποιες γνώσεις και δεξιότητες πρέπει να αναπτύξει ώστε να μπορεί να λειτουργεί αποτελεσματικά» (ΚΕΠΑ, 2008:1)

Με λίγα λόγια, το ΚΕΠΑ επιδιώκει να διαμορφώσει ένα εύχρηστο και κατανοητό εργαλείο για τους δημιουργούς προγραμμάτων διδασκαλίας γλωσσικών μαθημάτων, τους/τις συγγραφείς των διδακτικών εγχειριδίων και άλλου υλικού διδασκαλίας, τους φορείς διεξαγωγής εξετάσεων, τους/τις καθηγητές/-τριες και τους/τις μαθητές/-τριες προκειμένου να εντοπίζουν τις διάφορες μορφές εμπλοκής τους στη διαδικασία εκμάθησης ξένων γλωσσών, η οποία θα βασίζεται σε ένα ενιαίο περιγραφικό πλαίσιο αναφοράς (Heyworth, 2006).

Το σημαντικότερο κομμάτι του ΚΕΠΑ, το οποίο και αφορά άμεσα την παρούσα εργασία, είναι αυτό που αναφέρεται στα κοινά επίπεδα αναφοράς. Πρόκειται για μια ευρεία περιγραφή του τι μπορεί να κάνει ο/η χρήστης/-τρια μιας γλώσσας στα 6 διαφορετικά επίπεδα επίδοσης. Πρόκειται στην ουσία για τρία επίπεδα που κυμαίνονται από το «βασικό» επίπεδο (A1,A2), το επίπεδο «ανεξάρτητου» χρήστη (B1,B2) και αυτό του «ικανού» χρήστη (Γ1,Γ2). Αυτά λειτουργούν ως σημεία αναφοράς τόσο για την περιγραφή των επιπέδων και της επίδοσης, όσο και τον καθορισμό των στόχων. Επίσης, μας επιτρέπουν, σύμφωνα με τον Morrow (2004), να καθορίσουμε το τι θέλουμε από τους/τις μαθητές/-τριες μας να μπορούν να κάνουν στο κάθε ένα από τα επίπεδα αυτά.

Στα αρχικά στάδια δημιουργίας του ΚΕΠΑ για τις γλώσσες περιλαμβάνονταν ολόκληρες λίστες λεξιλογίου που ανταποκρίνονταν στο κάθε ένα από τα περιγραφόμενα στάδια. Όμως, η διαδικασία αυτή της δημιουργίας λιστών λέξεων είχε ένα σοβαρό μειονέκτημα. Περιοριζε την ευελιξία του όλου συστήματος και την ικανότητα να εφαρμόζεται σε μια μεγάλη ποικιλία γλωσσικών μαθημάτων. Για τον λόγο αυτό, από την έκδοση του ΚΕΠΑ του 2001 αφαιρέθηκαν οι λίστες αυτές, με το επιχείρημα ότι οι περιγραφητές πρέπει να παραμένουν ολιστικοί και πράγματι πρόκειται για μία πιο ολιστική προσέγγιση στην περιγραφή των γλωσσών. Αυτό έχει ως αποτέλεσμα να γίνει το πλαίσιο ιδιαίτερα ευέλικτο. Όμως, η υπερβολική αυτή

ευελιξία οδήγησε σε πολλές παρανοήσεις. Έτσι, πολλοί ερευνητές πιστεύουν ότι η ύπαρξη συγκεκριμένων παραγόντων μέτρησης που θα βασίζονται στη γνώση και θα αξιολογούνται αντικειμενικά, όπως είναι το εύρος του λεξιλογίου ανά επίπεδο, θα μπορούσε να βοηθήσει στο να αποφευχθούν τέτοιου είδους παρανοήσεις και ασάφειες. Όπως αναφέρει ο Milton στο *The development of vocabulary breadth across the CEFR levels* (2010:211-232) είναι πολύ σημαντικό να γνωρίζουμε τον αριθμό των λέξεων ανά επίπεδο. Όπως, επίσης, τονίζουν οι Milton και Alexiou (2009), ο προσδιορισμός του μεγέθους του λεξιλογίου ανά επίπεδο θα μπορεί να αποτελέσει ένα πολύ χρήσιμο γλωσσικό εργαλείο και να προβλέπει με ασφάλεια το επίπεδο ΚΕΠΑ του/της κάθε μαθητή/-τριας.

Αν και δεν υπάρχει, πλέον, ένα προκαθορισμένο πλαίσιο στο οποίο να αναγράφεται ξεχωριστά το μέγεθος και το βάθος του λεξιλογίου για το κάθε επίπεδο του ΚΕΠΑ, ο Milton (2010) επιχείρησε εμπειρικά να αντιστοιχήσει το κάθε επίπεδο του ΚΕΠΑ με αριθμούς που να δείχνουν κατά προσέγγιση το μέγεθος του λεξιλογίου που πρέπει ένας/μία μαθητής/-τρια να κατέχει στο επίπεδο αυτό.

Πίνακας 1: Εμπειρική εκτίμηση από τον Milton του μεγέθους του λεξιλογίου για κάθε επίπεδο γλωσσομάθειας

ΕΠΙΠΕΔΑ ΚΕΠΑ	Λέξεις κατά προσέγγιση
A1	Έως 1500
A2	1500-2500
B1	2500-3250
B2	3250-3750
Γ1	3750-4500
Γ2	4500-5000

Εκτιμητές λεξιλογικού μεγέθους ανά επίπεδο του ΚΕΠΑ (Milton, 2010, σλ.24)

Στην έρευνά του, ο Milton (2012:224) ισχυρίζεται ότι «οι μαθητές/-τριες περνώντας από το ένα επίπεδο του ΚΕΠΑ στο άλλο βελτιώνουν το νοητικό λεξικό τους στην ξένη γλώσσα, το οποίο αυξάνεται σε μέγεθος, πολυπλοκότητα και εύρος».

Συνεπώς, η σχέση ανάμεσα στο μέγεθος του λεξιλογίου και στα επίπεδα του ΚΕΠΑ είναι ιδιαίτερα ισχυρή, δεδομένου, βέβαια, ότι μπορεί να υπάρχουν αποκλίσεις ανάλογα με το άτομο. Η ίδια διαπίστωση προκύπτει και από το λεξιλόγιο των περιγραφητών, όπως παρουσιάζονται στο ΚΕΠΑ για τις τέσσερις δεξιότητες.

Είναι εμφανές, λοιπόν, ότι οι μαθητές/-τριες καλλιεργούν ένα διαρκώς αυξανόμενο νοητικό λεξικό, καθώς προάγονται από το ένα επίπεδο στο άλλο. Ο όρος «βασικό λεξιλόγιο» στο επίπεδο Α1 γίνεται «ένα πολύ ευρύ λεξιλογικό ρεπερτόριο» στο επίπεδο Γ2. Επίσης, την ανάπτυξη του λεξιλογικού βάθους τη συναντάμε μόνο στα ανώτερα επίπεδα. Στο ΚΕΠΑ υπονοείται και ότι για να υπάρξει πρόοδος στην επικοινωνία σε μια ξένη γλώσσα απαιτείται η κατάκτηση ολοένα και μεγαλύτερου όγκου λεξιλογίου και ένας πιο ακριβής έλεγχος στη χρήση του λεξιλογίου αυτού (Milton, 2013).

Επομένως, το λεξιλόγιο καταλαμβάνει μία περίοπτη –πλέον- θέση στη διδακτική διαδικασία και η καλλιέργεια του λεξιλογικού «πλούτου» συνδέεται άμεσα με αυτό. Ζητούμενο, βέβαια, είναι να μελετηθούν οι υπό εξέταση δείκτες, ώστε να υπάρχει μια σαφής εικόνα αυτών ανά επίπεδο γλωσσομάθειας.

2.2.3. Η έννοια του λεξιλογικού πλούτου

Στο πλαίσιο της υφομετρικής απόδοσης της συγγραφικής πατρότητας, η έρευνα των λεξιλογικών χαρακτηριστικών έχει επεκταθεί εκτός της χρήσης των λεξικών συχνοτήτων ως υφομετρικών χαρακτηριστικών και σε άλλους σύνθετους δείκτες, οι οποίοι είναι ευρύτερα γνωστοί στη βιβλιογραφία ως δείκτες λεξιλογικού «πλούτου» (lexical richness indices) ή δείκτες λεξιλογικής διαφοροποίησης (lexical diversity indices). Οι συγκεκριμένοι δείκτες αποσκοπούν με μια αριθμητική τιμή να περιγράψουν πόσο διαφορετικό λεξιλόγιο χρησιμοποιεί ένας/μία συγγραφέας στα έργα του/της και, επομένως, να προσεγγίσουν τον βαθμό του λεξιλογικού «πλούτου» που τον/την χαρακτηρίζει (Μικρός, 2015:78).

Στο σημείο αυτό, αξίζει να ειπωθεί πως τα χαρακτηριστικά τα οποία κατά καιρούς έχουν μετρηθεί ως υφομετρικοί δείκτες, από το ξεκίνημα της υφομετρίας μέχρι σήμερα, είναι πολυάριθμα¹⁵ και ανήκουν σε όλο το εύρος των γλωσσικών επιπέδων. Σύμφωνα με τον Bailey (1979), τα υφομετρικά χαρακτηριστικά θα πρέπει να διακρίνονται εύκολα, να έχουν άμεση σχέση με τη γλωσσική δομή, να είναι συχνά και εύκολα μετρήσιμα και να παρουσιάζουν αντίσταση στη συνειδητή χειραγώγηση

¹⁵ Ο Rudman (1997:360) αναφέρει χαρακτηριστικά ότι πάνω από 1.000 υφομετρικά χαρακτηριστικά έχουν προταθεί κατά καιρούς από διάφορους ερευνητές. Αν και η ακρίβεια του συγκεκριμένου αριθμού δεν μπορεί εύκολα να επιβεβαιωθεί, είναι σίγουρο ότι ο αριθμός των χαρακτηριστικών είναι εξαιρετικά μεγάλος και στην τρέχουσα έρευνα συνεχώς προτείνονται νέα.

από το συγγραφέα (Μικρός, 2015:58). Αυτές οι ιδιότητες παραμένουν ακόμα και σήμερα επίκαιρες, αφού τα υφομετρικά χαρακτηριστικά που χρησιμοποιούνται στην τρέχουσα έρευνα κατέχουν τις προαναφερθείσες ιδιότητες..

Παρά το γεγονός ότι η μελέτη των υφομετρικών χαρακτηριστικών ξεκίνησε επιδιώκοντας την απόδοση συγγραφικής πατρότητας σε έργα που ήταν αγνώστου ή αμφισβητούμενου πνευματικού δημιουργού, η αξιοπιστία της μεθόδου είναι πλέον υψηλή και ήδη έχει αρχίσει να χρησιμοποιείται επικουρικά μαζί με άλλες τεχνικές σε κρίσιμους κλάδους όπως είναι η εγκληματολογική γλωσσολογία (forensic linguistics)¹⁶. Από το σύνολο των υφομετρικών γνωρισμάτων, στην παρούσα εργασία, η εστίαση γίνεται στο λεξιλογικό επίπεδο, αφού ο λόγος γίνεται για τον λεξιλογικό «πλούτο». Στη δεδομένη περίπτωση, στόχος δεν είναι η απόδοση της συγγραφικής πατρότητας για ένα έργο, αλλά ο εντοπισμός δεικτών λεξιλογικού «πλούτου» στα γραπτά αλλόγλωσσων μαθητών/-τριών της ελληνικής και η συσχέτιση αυτών με το επίπεδο γλωσσομάθειας στο οποίο βρίσκονται.

2.2.4. Οι έξι δείκτες λεξιλογικού «πλούτου»

Από τους πολυάριθμους παράγοντες λεξιλογικού «πλούτου», η έρευνα επικεντρώθηκε σε έξι καίριους δείκτες: το Harax Legomenon Percentage (HL) ή Ποσοστό Άπαξ Λεγομένων, το R1, το Lambda (Λ), το h-point (h), την Entropy (H) ή Εντροπία και το Average Tokens length (ATL) ή Μέσο Μήκος Λεξικού Τύπου.

2.2.4.1. H-point

Ο Hirsch (2005) προσπάθησε να υπολογίσει τον σωρευτικό αντίκτυπο και τη συνάφεια των αποτελεσμάτων κάθε επιστημονικής έρευνας με αμερόληπτο τρόπο, καθώς, όπως επισημαίνει, η ποσοτικοποίηση είναι συχνά απαραίτητη για σκοπούς αξιολόγησης και σύγκρισης. Για να επιτευχθεί αυτό, πρότεινε τον δείκτη h (h-point). Ο συγκεκριμένος δείκτης δημιουργήθηκε από τον Hirsch για τη μέτρηση απήχησης που έχουν οι επιστημονικές εργασίες που δημοσιεύονται (scientometrics), αλλά αποδείχθηκε πως είναι χρήσιμος και για άλλους επιστημονικούς κλάδους. Σύμφωνα

¹⁶ Solan, L. M., and Tiersma, P. M. (2004) Author identification in american courts, Applied Linguistics 25, 4, 448-465.

με τον ίδιο, οι ερευνητές που δημοσιεύουν ποιοτικό και ποσοτικά υψηλό επιστημονικό έργο έχουν επίσης υψηλό h-point, ενώ επισημαίνει ότι ο δείκτης h στις βιοεπιστήμες είναι πολύ υψηλότερος απ' ό,τι στη φυσική (Hirsch, 2005). Στη γλωσσολογία, το h-point εμφανίστηκε για πρώτη φορά με τον Popescu (Popescu, Altmann, Grzybek, Jayaram, Köhler, Krupa, Mačutek, Pustet, Uhlířová & Vidya, 2009a). Αυτός ο δείκτης είναι ένα σταθερό σημείο που έχει διαφορετικές χρήσεις στην κειμενολογία (textology). Κατά βάση, το h-point χωρίζει το λεξιλόγιο σε δύο ομάδες, τις λέξεις ή φράσεις που λαμβάνουν σημασία όταν πλαισιώνονται από άλλες λέξεις (synsemantics) και τις λέξεις ή φράσεις που φέρουν σημασία ανεξάρτητα από το γλωσσικό περιβάλλον (autosemantics)¹⁷. Οι λέξεις της πρώτης κατηγορίας είναι πάντα περισσότερες σε σχέση με τις λέξεις της δεύτερης. Όσο περισσότερες είναι οι «αυτόνομες» λέξεις και όσο πιο συχνά εμφανίζονται, τόσο μεγαλύτερη είναι η θεματική συγκέντρωση του κειμένου (Popescu, Mačutek & Altmann, 2009b).

2.2.4.2. R1

Το R1 είναι ένας δείκτης λεξιλογικού «πλούτου» που βασίζεται στο h-point (h). Αυτός ο δείκτης μειώνει τον αντίκτυπο του μήκους κειμένου και ενισχύει τη δυνατότητα σύγκρισης των κειμένων. Περιλαμβάνει όχι μόνο τα άπαξ λεγόμενα (hapax legomena), αλλά, στην πραγματικότητα, σχεδόν όλες τις «αυτόνομες» λέξεις (autosemantics) μέχρι το σημείο h, αποδίδοντας μια πιο ρεαλιστική εικόνα του πλούτου του λεξιλογίου. Και ο δείκτης Lambda (Λ) έχει πολύ άμεση σχέση με το R1 (Popescu et al., 2009b:29-34).

2.2.4.3. Ποσοστό Άπαξ Λεγομένων (Hapax Legomenon Percentage /HL)

Το άπαξ λεγόμενο (hapax legomenon) είναι μία λέξη η οποία εμφανίζεται μόνο μία φορά σε ένα κείμενο. Το ποσοστό των άπαξ λεγομένων (Hapax Legomenon Percentage) εκφράζεται ως μία απλή αναλογία ανάμεσα στον συνολικό αριθμό των λέξεων (-N) ενός κειμένου και τον αριθμό των Άπαξ Λεγομένων (Nh). Ο δείκτης HL μπορεί να εκτιμηθεί μόνο αν κάποιος εξαλείψει τον επιβλαβή παράγοντα του μήκους

¹⁷Σύμφωνα με την ιστοσελίδα [https://www.thefreedictionary.com:synsemantic\(adj.\): a word or phrase meaningful only when it occurs in the company of other words και autosemantic \(adj.\): a word or phrase meaningful in isolation, independent of context.](https://www.thefreedictionary.com:synsemantic(adj.): a word or phrase meaningful only when it occurs in the company of other words και autosemantic (adj.): a word or phrase meaningful in isolation, independent of context.)

του κειμένου μέσω κάποιου μετασχηματισμού. Αυτό έχει δοκιμαστεί σε πολλές περιπτώσεις, αλλά οι λύσεις δεν φαίνεται να είναι ικανοποιητικές. Βέβαια, το ίδιο ισχύει και για όλους τους άλλους δείκτες που αφορούν άλλες ιδιότητες του κειμένου και εξαρτώνται από το μήκος κειμένου (Popescu, Čech & Altmann, 2011). Ως δείκτης λεξιλογικού «πλούτου» έχει αρκετές ατέλειες: (i) τα άπαξ λεγόμενα δεν είναι οι αποκλειστικοί δείκτες του πλούτου (ii) πολύ σύντομα κείμενα μπορούν να αποκτήσουν τον μέγιστο πλούτο 1, ενώ πολύ μεγάλα κείμενα στα οποία όλες οι λέξεις επαναλαμβάνονται θα τείνουν να έχουν πλούτο που θα πλησιάζει το 0. Όμως, τα κείμενα του ενδιάμεσου μήκους, παραδείγματος χάρη από το $N = 100$ έως το $N = 100000$ θα εξακολουθούσαν να εξαρτώνται από το N . Επομένως, οι δείκτες τύπου $V1/N$ δεν λειτουργούν σωστά για οποιοδήποτε μήκος κειμένου, ακόμα και αν είναι δυνατό να δημιουργηθεί μια δοκιμή για σύγκριση για όχι πολύ σύντομα και όχι πολύ μεγάλα κείμενα. (Popescu et al., 2009a: 31-34).

2.2.4.4. Lambda (Λ)

Το lambda (Λ) είναι ένας δείκτης ο οποίος ασχολείται με μια δομή συχνοτήτων του κειμένου. Από τη μία πλευρά, το Λ σχετίζεται με τον λεξιλογικό «πλούτο», από την άλλη μεριά λαμβάνει υπόψη τη σχέση μεταξύ των γειτονικών συχνοτήτων (Cech, 2011). Για παράδειγμα, αν έχουμε τρία κείμενα με τον ίδιο αριθμό λέξεων ($N = 20$) και το ίδιο μέγεθος λεξικών τύπων ($V = 10$), ο λόγος των λεξικών τύπων προς τις λέξεις (TTR) είναι επίσης ο ίδιος ($10/20 = 0,5$). Όμως, τα τρία αυτά κείμενα με το ίδιο N και V μπορεί να έχουν διαφορετικές σχέσεις μεταξύ των συχνοτήτων. Ενδέχεται σε μία από τις περιπτώσεις οι λεξικοί τύποι να παρουσιάζουν μεγάλη ανισομέρεια ως προς τις εμφανίσεις τους (π.χ. ένας λεξικός τύπος να εμφανίζεται πέντε φορές) και σε μία άλλη να εμφανίζονται όλοι οι λεξικοί τύποι να σημειώνουν τον ίδιο αριθμό εμφανίσεων (π.χ. οι δέκα λεξικοί τύποι να εμφανίζονται από δύο φορές έκαστος). Ένα από τα μεγαλύτερα πλεονεκτήματα αυτής της μέτρησης είναι η ανεξαρτησία του δείκτη Λ από το μήκος κειμένου, η επίδραση του οποίου πλήττει την αξιοπιστία όλων των άλλων δεικτών λεξιλογικού «πλούτου» (Popescu et al., 2011:1).

Είναι αναγκαίο να τονιστεί ότι το Λ δεν είναι ένας δείκτης που μαρτυρά κατευθείαν την ποιότητα της γραφής του συγγραφέα ή το ταλέντο του. Το Λ εκφράζει τη δομή η

οποία αναδύεται ως αποτέλεσμα της χρήσης της γλώσσας. Ο δείκτης αυτός εκφράζει μια ιδιότητα που βρίσκεται σε συνάρτηση με κάποια ποιοτικά χαρακτηριστικά του κειμένου αλλά από μόνος του δεν αποτελεί "δείκτη ποιότητας". Ως εκ τούτου, το Δ είναι άμεσα ένας δείκτης της διαμόρφωσης συχνότητας του κειμένου και έμμεσα ένας δείκτης λεξιλογικού «πλούτου»¹⁸ (Popescu et al., 2011:1-9).

2.2.4.5. Εντροπία (Entropy /H)

Ο όρος "εντροπία" χρησιμοποιείται σε πολλές επιστήμες και υπάρχουν διάφοροι ορισμοί με διαφορετικές έννοιες. Γενικά, η εντροπία μετράει την ποικιλομορφία ή την αβεβαιότητα. Στη γλωσσολογία, η εντροπία εστιάζει στη μελέτη συχνοτήτων των μορφών λέξεων. Δε συνιστά ένα μέτρο αβεβαιότητας, αλλά ένα μέτρο μη ομοιομορφίας. Εάν η συγκέντρωση συχνοτήτων είναι υψηλή, η εντροπία είναι μικρή. Όσο μικρότερο είναι το H , τόσο μικρότερος είναι ο λεξιλογικός «πλούτος» (Popescu et al., 2009b:157-161). Στο πλαίσιο της ποσοτικής γλωσσολογίας, τα εντροπικά μέτρα χρησιμοποιούνται για την κατανόηση των νόμων στις φυσικές γλώσσες, όπως η σχέση μεταξύ της συχνότητας λέξεων, της προβλεψιμότητας και του μήκους των λέξεων. Μερικά από τα κύρια εννοιολογικά και πρακτικά προβλήματα της εκτίμησης της εντροπίας λέξεων είναι η εξάρτηση από το μέγεθος του κειμένου, το κειμενικό είδος, το ύφος, η μέθοδος εκτίμησης, καθώς και η εξάρτηση των λέξεων από το συγκεκριμένο. Γενικά ισχύει πως η εντροπία μεγαλώνει όταν δεν μπορούμε να προβλέψουμε την εμφάνιση μιας λέξης βάσει των λέξεων που προηγούνται, γεγονός το οποίο, κατά προσέγγιση, ισχύει όλες τις γλώσσες του κόσμου (Bentz, Alikaniotis, Cysouw & Ferrer-i-Cancho, 2017).

2.2.4.6. Μέσο Μήκος Λεξικού Τύπου (Average Tokens Length/ ATL)

Το Μέσο Μήκος του Λεξικού Τύπου σχετίζεται με τον αριθμητικό μέσο όρο των λεξικών τύπων. Η μέτρηση του λεξικού μήκους μπορεί να γίνει τόσο σε χαρακτήρες όσο και σε συλλαβές, αν και οι δεύτερες θεωρούνται περισσότερο αξιόπιστες¹⁹ γιατί

¹⁸ Είναι αναγκαίο να σημειωθεί ότι η δομή συχνοτήτων πανομοιότυπου κειμένου σε εξαιρετικά συνθετικές και εξαιρετικά αναλυτικές γλώσσες αποδίδει πολύ διαφορετικές τιμές του Δ , οι οποίες στην περίπτωση αυτή σηματοδοτούν την ιδιότητα της γλώσσας και όχι την ιδιότητα του κειμένου (Popescu et al., 2011).

¹⁹ Σε ορισμένες γλώσσες όπως τα Νέα Ελληνικά η αυτόματη αναγνώριση και μέτρηση των συλλαβών στις λέξεις είναι από μόνο του ξεχωριστό πρόβλημα Επεξεργασίας Φυσικής Γλώσσας. Στη Νέα

δεν επηρεάζονται από την πολυπλοκότητα του γραφηματικού συστήματος (Grzybek, 2007:18). Οι βασικότεροι δείκτες που χρησιμοποιούνται για την περιγραφή του σχήματός της μέτρησής του είναι ο μέσος όρος που ονομάζεται και πρώτη κεντρική ροπή (first central moment). Για να υπολογιστεί ο μέσος όρος λεξικού μήκους, αθροίζεται το μήκος κάθε λέξης του κειμένου (είτε σε χαρακτήρες είτε σε συλλαβές) και διαιρείται με το κειμενικό μέγεθος, το συνολικό αριθμό, δηλαδή, των λέξεων που απαρτίζουν το κείμενο (Μικρός, 2015:59-64).

2.3. Γλωσσολογία βάσει ΗΣΚ

Η έννοια του σώματος κειμένων (corpus) προσφέρει μια διαφορετική προσέγγιση στα γλωσσικά δεδομένα. Πρόκειται για τη συλλογή μεγάλου όγκου γλωσσικών δεδομένων τα οποία είναι αποθηκευμένα σε ηλεκτρονική μορφή με τρόπο ώστε να ανακαλούνται ταχύτατα και με ακρίβεια. Ένα σώμα κειμένων αποτελεί μια συστηματική συλλογή γλωσσικών δεδομένων που προέκυψαν με φυσικό τρόπο και όχι με πειραματικές διαδικασίες και δε σχετίζονται με τη διαίσθηση του/της ερευνητή/-τριας, αλλά με τη γλωσσική χρήση. Πρόκειται για αρχεία γλωσσικής χρήσης που απαντούν σε προφορικά και γραπτά κείμενα και αποτελούν μαρτυρίες για το τι πραγματικά ειπώθηκε ή γράφτηκε και μάλιστα ως προϊόν της αυθόρμητης γλωσσικής δραστηριότητας των ομιλητών μιας γλώσσας και όχι ως απάντηση σε μια γλωσσολογική έρευνα για το υπό εξέταση στοιχείο (Γούτσος & Φραγκάκη, 2015:13-14). Κατ' επέκταση, η γλωσσολογία σωμάτων κειμένων δεν αφορά έναν συγκεκριμένο κλάδο της γλωσσολογίας ή ένα μόνο επίπεδο γλωσσικής ανάλυσης, αλλά μπορεί να έχει εφαρμογές σε πολλές πτυχές της γλώσσας. Με αυτόν τον τρόπο διαφέρει από τους κλάδους της γλωσσολογίας που εξετάζουν πτυχές του γλωσσικού συστήματος, όπως η φωνολογία, η σύνταξη αλλά και από εκείνους τους κλάδους που αφορούν πτυχές της χρήσης της γλώσσας όπως λ.χ. η κοινωνιογλωσσολογία και η κειμενογλωσσολογία. Η χρήση των σωμάτων κειμένων στη μελέτη της γλώσσας συνιστά, καταρχάς, ένα σύνολο μεθόδων και διαδικασιών που αφορά τα εμπειρικά

Ελληνική υπάρχει το πρόβλημα της συνίζησης, της επιλεκτικής, δηλαδή, ένωσης του άτονου ι με το προηγούμενο γειτονικό του φωνήεν. Έτσι λέξεις όπως «λόγια» ή «χρόνια» μπορούν να έχουν δύο δυνατές συλλαβικές τεμαχίσεις π.χ.: α) «χρό-νια» (έχω 20 χρόνια να τον δω) β) «χρό-νι-α» (χρόνια ασθένεια). Το φαινόμενο αυτό είναι αρκετά συχνό και παράγει πολλά λάθη στον αυτόματο συλλαβισμό της Νέας Ελληνικής. Για το λόγο αυτό στα Νέα Ελληνικά και την έρευνα που θα παρουσιάσουμε εδώ, θα χρησιμοποιήσουμε ως βάση μέτρησης το χαρακτήρα (Μικρός, 2015).

δεδομένα της γλωσσολογίας και, επομένως, σε ένα πρώτο επίπεδο η γλωσσολογία σωμάτων κειμένων αποτελεί μια μεθοδολογία (Γούτσος & Φραγκάκη, 2015:22).

2.3.1.Ορισμός των ΗΣΚ

Τα Ηλεκτρονικά Σώματα Κειμένων (στο εξής ΗΣΚ), ο πυρήνας της σωματο-κειμενικής γλωσσολογίας, σύμφωνα με τους McEnery και Wilson (1996:24), συνιστούν σώματα κειμένων με πεπερασμένο αριθμό κειμένων και λέξεων σε ηλεκτρονική μορφή, τα οποία έχουν επιλεγεί με τέτοιο τρόπο, ώστε να αποτελούν όσο το δυνατόν πιο αντιπροσωπευτικά δείγματα της γλωσσικής ποικιλίας που μελετάται. Όπως αναφέρουν και οι Μικρός & Φουντοπούλου (2002:49), το ΗΣΚ είναι ένα σώμα κειμένων το οποίο είναι κωδικοποιημένο για τυποποιημένες (standardized) και ομοιογενείς εργασίες ανάκτησης γλωσσικής πληροφορίας. Τα γλωσσικά τμήματα που περιέχει περιλαμβάνουν πληροφορίες για την προέλευσή τους και την εξωγλωσσική λειτουργία τους.²⁰ Βασικό στοιχείο των ΗΣΚ θεωρείται και η δυνατότητα γενικεύσεων που μπορεί να προσφέρουν ως πηγή πληροφοριών για την ευρύτερη γλωσσική ποικιλία ή και ολόκληρη τη γλώσσα στην οποία ανήκουν (Γούτσος & Φραγκάκη, 2015).

2.3.2. Προδιαγραφές των ΗΣΚ

Βάσει των παραπάνω ορισμών, ένα ΗΣΚ για να είναι ερευνητικώς αξιοποιήσιμο, είναι αναγκαίο να διαθέτει (Granger, 2002:8-10): α) αυθεντικό γλωσσικό υλικό αντλημένο από ένα ευρύ φάσμα επικοινωνιακών περιστάσεων, το οποίο μπορεί να περιλαμβάνει από καθημερινές αυθόρμητες συνδιαλέξεις μέχρι δραστηριότητες εντός της γλωσσικής τάξης, β) δείγματα από διάφορες γλωσσικές ποικιλίες (π.χ. επίσημο/ανεπίσημο ύφος), γ) μόνο κειμενικά δεδομένα, δηλαδή συνεχή τμήματα λόγου με μια ορισμένη ολότητα και όχι απομονωμένες προτάσεις ή λέξεις, δ) σαφείς σχεδιαστικές προδιαγραφές, ε) τυποποίηση και πλήρη τεκμηρίωση των δεδομένων του (συγκεκριμένη μορφή ηλεκτρονικών δεδομένων, τυποποιημένο και εύχρηστο λογισμικό επισημείωσης, τεκμηρίωση των μεταβλητών των πληροφορητών και των

²⁰ Στην ίδια μελέτη τους οι Μικρός & Φουντοπούλου (2002:49-52) προσεγγίζουν αναλυτικά την έννοια του ΗΣΚ και παρουσιάζουν διάφορα κριτήρια κατηγοριοποίησης των ΗΣΚ παραθέτοντας σε κάθε περίπτωση αρκετά παραδείγματα καθιερωμένων πια ΗΣΚ τόσο γενικής όσο και ειδικής ερευνητικής χρήσης.

δραστηριοτήτων στο πλαίσιο των οποίων παρήχθη το γλωσσικό υλικό), οι λεπτομέρειες της οποίας πρέπει να καταγραφούν για κάθε κείμενο του σώματος και να είναι διαθέσιμες στους ερευνητές είτε ως ενσωματωμένη πληροφορία στο ηλεκτρονικό αρχείο κάθε κειμένου υπό μορφή SGML είτε αποθηκευμένες ξεχωριστά, αλλά συνδεδεμένες με το κάθε αρχείο μέσω ενός συστήματος αναφοράς και τέλος ε) χαρακτηρισμό.²¹ Ο ηλεκτρονικός χαρακτηρισμός ενός ΗΣΚ είναι κεφαλαιώδους σημασίας για την αξιοποιησιμότητά του και συνιστά την προσθήκη ερμηνευτικών -κυρίως γλωσσολογικών- πληροφοριών σε ένα υπάρχον σώμα γραπτού ή/και προφορικού λόγου μέσω κάποιου είδους κωδικοποίησης που είτε είναι συνημμένη είτε διέπει τα ηλεκτρονικά δεδομένα με το πρωτογενές γλωσσικό υλικό (Granger, 2002:16). Η μορφή χαρακτηρισμού που πρέπει να επιλεγεί για ένα ΗΣΚ εξαρτάται πρωτίστως από το υπό εξέταση ερευνητικό ερώτημα. Γίνεται κατανοητός, λοιπόν, ο μεγάλος αριθμός διαφορετικών ειδών χαρακτηρισμού ΗΣΚ που μαρτυρείται στη διεθνή βιβλιογραφία.

2.3.3. Αντιπροσωπευτικότητα των ΗΣΚ

Σημαντικό, επίσης, είναι και το ζήτημα της αντιπροσωπευτικότητας ενός ΗΣΚ. Σύμφωνα με τον Γούτσο (2003:5-6), στη διεθνή πρακτική κατάρτισης ΗΣΚ υπάρχουν δύο είδη διαστρωμάτωσης του υλικού προκειμένου να είναι αντιπροσωπευτικό: α) η διαστρωμάτωση με βάση δημογραφικά στοιχεία και β) η διαστρωμάτωση με βάση την κειμενική τυπολογία. Για τα ελληνικά δεδομένα, τόσο το ΣΕΚ (Γούτσος, 2003) όσο και το ΗΣΜΚ της ΝΕΓ2 (Τζιμόκας, 2010· Τζιμόκας, 2014), που αξιοποιείται στην παρούσα έρευνα, ακολουθούν το δεύτερο είδος διαστρωμάτωσης. Επιπλέον, ο Γούτσος (2003) αναφέρει ότι η αντιπροσωπευτικότητα αναφέρεται σε τρεις διαστάσεις κάθε ΗΣΚ, το μέγεθος, την αυθεντικότητα και τις αναλογίες, τη σχετική «ισορροπία» δηλαδή μεταξύ των κειμενικών ειδών που το απαρτίζουν. Εξάλλου, όπως επισημαίνει ο Μικρός (2007:192), η αντιπροσωπευτικότητα ενός ΗΣΚ συνδέεται άμεσα με το λειτουργικό εύρος του. Ένα ΗΣΚ με μεγάλο λειτουργικό εύρος περιέχει κείμενα όπου εμφανίζονται γλωσσικά χαρακτηριστικά όλων των συχνοτήτων, από τα πολύ συχνά έως τα πολύ σπάνια. Γενικά, ισχύει ότι όσο πιο

²¹ Για τον χαρακτηρισμό των ΗΣΚ μπορεί κανείς να ανατρέξει στους McEnery & Wilson, 1996· McEnery & Rayson, 1997:196-7· Pravec, 2002: 81· Φλώρου, 2009: 21-25.

ευρεία αντιπροσώπευση πετύχουμε, τόσο μεγαλύτερες πιθανότητες θα έχουμε να συμπεριλάβουμε γλωσσικές μονάδες χαμηλής συχνότητας.

Όσον αφορά το ΗΣΜΚ της ΝΕΓ2, λεπτομέρειες για το περιεχόμενό του, τις σχεδιαστικές προδιαγραφές του, την τυποποίηση και την τεκμηρίωσή του καθώς και για τον χαρακτηρισμό και την αντιπροσωπευτικότητά του θα αναφερθούν αναλυτικά στο κεφάλαιο 3.

2.3.4. Περιορισμοί ΗΣΚ

Όσον αφορά τους περιορισμούς των ΗΣΜΚ, η συσχέτιση μεγέθους του ΗΣΚ και εγκυρότητας των ερευνητικών αποτελεσμάτων προκαλεί μεγάλο προβληματισμό. Πέρα από αυτό, αξίζει να ληφθεί υπόψη ότι τα ερευνητικά συμπεράσματα που βασίζονται σε ένα ΗΣΚ καταδεικνύουν τι ισχύει αποκλειστικά για το συγκεκριμένο ΗΣΚ. Κατά συνέπεια, καθίσταται δύσκολη η συναγωγή γενικεύσεων βάσει ενός ΗΣΚ, εκτός και εάν διαθέτει αντιπροσωπευτικότητα τέτοια που να επιτρέπει εύλογες γενικεύσεις (McEnery, Xiao & Topo, 2006). Είναι επόμενο, λοιπόν, τα όρια κάθε έρευνας στην οποία βασίζεται ένα ΗΣΚ να προκαθορίζονται από τις προδιαγραφές του. Μια τέταρτη αδυναμία των ΗΣΜΚ αφορά την απουσία των λεγόμενων «αρνητικών ενδείξεων», δηλαδή στοιχείων που θα μας βοηθούν στην πρόβλεψη του τι είναι και τι δεν είναι δυνατόν να υπάρχει σε μια γλώσσα. Ωστόσο, από ένα αντιπροσωπευτικό ΗΣΚ μπορεί να αντληθούν στοιχεία για τις πιο κυρίαρχες τάσεις και τα πιο τυπικά χαρακτηριστικά μιας γλώσσας.

2.3.5. Είδη ηλεκτρονικών σωμάτων κειμένων

Υπάρχουν διαφορετικά είδη ΗΣΚ ανάλογα με το μέγεθός τους, τον συγχρονικό ή διαχρονικό τους χαρακτήρα, τα είδη λόγου που εκπροσωπούν, τα κριτήρια με βάση τα οποία δομούνται σε υποσύνολα (subcorpora), το αν προέρχονται από τον λόγο φυσικών ή μη φυσικών ομιλητών/-τριών, αν είναι μονόγλωσσα ή δίγλωσσα κ.ά. Σύμφωνα με τους McEnery et al. (2006), τα ΗΣΚ διακρίνονται σε (α) Προφορικά και γραπτά σώματα κειμένων, (β) Γενικά και ειδικά, (γ) Μονόγλωσσα και πολύγλωσσα, (δ) Συγχρονικά και διαχρονικά ή ιστορικά σώματα κειμένων, (ε) Επισημειωμένα και μη επισημειωμένα, (στ) Παιδαγωγικά, (ζ) Σώματα Κειμένων Μαθητών. Τα τελευταία

περιλαμβάνουν κυρίως γραπτά κείμενα μαθητών/-τριών μιας δεύτερης ή ξένης γλώσσας. Τα Σώματα Κειμένων Μαθητών ή Ηλεκτρονικά Σώματα Μαθητικών Κειμένων (ΗΣΜΚ) είναι μια σχετικά πρόσφατη εξέλιξη στο χώρο της χρήσης των ΗΣΚ και στόχο έχουν τη διερεύνηση των γλωσσικών ικανοτήτων των μαθητών/-τριών και των λαθών τους είτε για ερευνητικούς είτε για διδακτικούς σκοπούς.

2.3.5.1. Ηλεκτρονικά Σώματα Μαθητικών Κειμένων (ΗΣΜΚ)

Στην αρχή της δεκαετίας του '90 εμφανίστηκε το ηλεκτρονικό σώμα μαθητικού λόγου (computer learner corpus) ή Ηλεκτρονικό Σώμα Μαθητικών Κειμένων (στο εξής ΗΣΜΚ), το οποίο μπορεί να οριστεί ως μια ηλεκτρονική βάση δεδομένων η οποία περιλαμβάνει αυθεντικά (δια)γλωσσικά δεδομένα που έχουν παραχθεί από μαθητές/-τριες μιας δεύτερης ή ξένης γλώσσας (Leech, 1998· Dagneaux et al. 1998:165). Δύο είναι οι κυριότερες διαφορές ενός τέτοιου σώματος από τις παλαιότερες συλλογές δεδομένων: α) είναι πλήρως μηχανογραφημένο και άρα μπορεί να αναλυθεί γρήγορα και αποτελεσματικά με τη χρήση πλήθους ηλεκτρονικών εργαλείων και β) είναι μεγάλο και επομένως αποτελεί μια πολύ πιο αξιόπιστη βάση για την περιγραφή της διαγλώσσας (Granger, 2003:465). Αυτό το νέο εργαλείο εισάγει στον χώρο της έρευνας και της διδασκαλίας της δεύτερης/ξένης γλώσσας νέες μεθοδολογικές πρακτικές, από τις οποίες όσες εστιάζουν στη διδασκαλία αναγνωρίζονται υπό τον όρο Εκμάθηση Γλώσσας με τη Βοήθεια Υπολογιστή (Computer-Aided Language Learning), ενώ όσες έχουν καθαρά ερευνητική κατεύθυνση αναγνωρίζονται απλώς ως Έρευνα βάσει ΗΣΜΚ (Computer Learner Corpus Research).

Τα πρώτα ΗΣΜΚ εκμάθησης δεύτερης/γλώσσας που εμφανίστηκαν τη δεκαετία του '90 είχαν αντικείμενο την αγγλική ως δεύτερη/ξένη γλώσσα. Σήμερα, η πλειοψηφία των σχετικών ΗΣΜΚ παραμένει στο χώρο της αγγλικής, αλλά παράλληλα εμφανίζονται όλο και περισσότερα ΗΣΜΚ που αφορούν και άλλες γλώσσες. Ενδεικτικά αναφέρεται ότι από τα 128 ΗΣΜΚ που υπάρχουν παγκοσμίως και είναι καταγεγραμμένα στον διαδικτυακό κατάλογο του πανεπιστημίου της Louvain στην ηλεκτρονική διεύθυνση <http://www.uclouvain.be/en-cecl-lcworld.html>, τα 76 (60%) αφορούν τα αγγλικά ως Γ2 και τα 52 (40%) άλλες γλώσσες. Γενικά, μπορούμε να

παρατηρήσουμε μια καθυστέρηση δεκαετίας στα ελληνικά σε σχέση με τις φάσεις ανάπτυξης των σωμάτων κειμένων στα αγγλικά (Renouf, 2007).

2.3.5.1.1. Σώματα Μαθητικών Κειμένων της ελληνικής ως Γ2

Όσον αφορά τα ελληνικά ως δεύτερη ή ξένη γλώσσα, η ανάπτυξη ΗΣΜΚ βρίσκεται προς το παρόν σε σχετικά πρώιμο στάδιο. Συγκεκριμένα, μέχρι στιγμής έχουν δημιουργηθεί τέσσερα ΗΣΜΚ της ΝΕΓ2 στο πλαίσιο αντίστοιχων ερευνητικών προγραμμάτων: α) το ΗΣΚ εκμάθησης της ελληνικής γλώσσας του προγράμματος «Πυθαγόρας Ι» με φορέα υλοποίησης το Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών (2004 - 2007), β) το ΗΣΚ των παιδιών της μουσουλμανικής μειονότητας της Θράκης του προγράμματος «Εκπαίδευση των παιδιών της μουσουλμανικής μειονότητας στη Θράκη»,²² γ) το «Ελληνικό Σώμα Κειμένων Μαθητών» (ΕΣΚΕΙΜΑΘ) με ΠΓΛ παιδιών μεταναστών και παλιννοστούντων του προγράμματος «Εκπαίδευση αλλοδαπών και παλιννοστούντων μαθητών»²³ και δ) το «Σώμα Εξελικτικών Παραγωγών Μαθητών της Ελληνικής ως Γ2 (ΣΕΠΑΜΕ2)» με παραγωγές γραπτού και προφορικού λόγου από μαθητές/-τριες οι οποίοι/-ες φοιτούν στο Διδασκαλείο της Νέας Ελληνικής ως δεύτερης/ξένης γλώσσας του ΕΚΠΑ και συμμετέχουν στα ειδικά τμήματα ενίσχυσης των παραγωγικών τους δεξιοτήτων (Ιακώβου κ.ά., 2017).²⁴ Το ΗΣΜΚ του ΕΚΠΑ αξιοποιείται από την παρούσα έρευνα και θα παρουσιαστεί αναλυτικά στο 3ο κεφάλαιο. Τα άλλα τρία ΗΣΜΚ που αναφέρθηκαν παραπάνω προορίζονται προς το παρόν μόνο για ακαδημαϊκή χρήση, ενώ το γλωσσικό υπόβαθρο των πληροφορητών/-τριών τους περιλαμβάνει διάφορες μητρικές γλώσσες. Όπως είναι αναμενόμενο, στο ΗΣΚ των μουσουλμανοπαίδων κυριαρχούν η τουρκική, η πομακική και η ρομά. Στο ΕΣΚΕΙΜΑΘ συμπεριλαμβάνονται κείμενα από μαθητές/-τριες αλβανικής (47%), ρωσικής (15%),

²² βλ. Τζεβελέκου, Σταμούλη, Κάντζου, Παπαγεωργακόπουλος, Χονδρογιάννη, Βαρλοκώστα, Ιακώβου & Λύτρα, 2008.

²³ βλ. Παπαδοπούλου & Τάντος, 2014· Τάντος κ.ά., 2015.

²⁴ Η ανάπτυξη του ΣΕΠΑΜΕ2 ξεκίνησε στα τέλη του 2014. Πρόκειται για την πρώτη βάση εξελικτικών δεδομένων τόσο από γραπτό όσο και από προφορικό λόγο των μαθητών, η οποία βρίσκεται προς το παρόν στο στάδιο της κατάρτισης και συνοδεύεται από πολύ λεπτομερή τεκμηρίωση τόσο του γλωσσικού υλικού όσο και των δημογραφικών χαρακτηριστικών των πληροφορητών. Περισσότερες πληροφορίες σχετικά με το εν λόγω ΗΣΚ υπάρχουν στην ιστοσελίδα του: <http://sepame2.phil.uoa.gr/>

ελληνικής (15%), γεωργιανής (6%), βουλγαρικής (6%), ρουμανικής (2,4%), τουρκικής (1,8%) και άλλων μητρικών γλωσσών (7,2%). Τέλος, η πλειοψηφία των πληροφορητών/-τριών του ΣΕΠΑΜΕ2 έχουν ως Γ1 τη ρωσική (30,4%) και ακολουθούν οι μαθητές/-τριες με Γ1 την αραβική (9,8%), την ουκρανική (8,6%), την αγγλική (7,4%), τη ρουμανική (4,3%), την ισπανική (3,7%), τη γαλλική (3,3%), την τουρκική (3,1%), την αλβανική (2,9%), την ιταλική (2,9%), την κινεζική (2,6%), τη σερβική (2,6%), τη γεωργιανή (2,4%), ενώ στο υπόλοιπο 15,8% των πληροφορητών συμπεριλαμβάνονται μαθητές/-τριες ποικίλου γλωσσικού υποβάθρου με μικρότερη αντιπροσώπευση. Από το σύνολο των 317 γραπτών του ΕΣΚΕΙΜΑΘ, στην έρευνα αυτή αξιοποιήθηκαν τα 224 ώστε να περιοριστούν οι μεγάλες αποκλίσεις ως προς το μέγεθος των κειμένων, γεγονός που θα έθετε υπό αίρεση την αξιοπιστία των αποτελεσμάτων.

2.3.5.1.2. Μεθοδολογικές πρακτικές της γλωσσολογίας βάσει ΗΣΜΚ

Στην παρούσα ενότητα γίνεται λόγος για γενικότερα ζητήματα που αφορούν τρεις βασικές μεθοδολογικές πρακτικές της έρευνας για την εκμάθηση της Γ2: α) την ποσοτική ανάλυση, β) την ποιοτική ανάλυση και γ) τη διάκριση μεταξύ διατομεακών και διαχρονικών δεδομένων. Σκοπός αυτών των επισημάνσεων είναι να αποσαφηνιστεί πληρέστερα ο μεθοδολογικός προσανατολισμός της παρούσας έρευνας.

2.3.5.1.2.1 Ποσοτική ανάλυση

Ήδη έχει επισημανθεί ότι τα ΗΣΜΚ αποτελούν ένα μεθοδολογικό εργαλείο που παρέχει ποικίλες δυνατότητες ποσοτικής και ποιοτικής ανάλυσης των γλωσσικών δεδομένων. Όσον αφορά την ποσοτική ανάλυση των δεδομένων του ΗΣΜΚ της παρούσας έρευνας, στηρίχθηκε σε ένα από τα πιο διαδεδομένα προγράμματα ποσοτικής ανάλυσης κειμένου, ιδιαίτερα χρήσιμο και δημοφιλές στις ανθρωπιστικές επιστήμες, το QUITA. Δεδομένου ότι το QUITA (Quantitative Indicator Text Analyzer) δεν αφορά αποκλειστικά την ποσοτική γλωσσολογία, περιέχει τους πιο βασικούς και συνήθεις δείκτες. Πρόκειται για ένα πρόγραμμα που συνδυάζει όλα τα σημαντικά μέρη κάθε ποσοτικής έρευνας: απόκτηση αποτελεσμάτων, στατιστικές δοκιμές και γραφική απεικόνιση.

2.3.5.1.2.2. Ποιοτική ανάλυση

Τα δύο βασικά ζητήματα της ποιοτικής ανάλυσης του περιεχομένου των ΗΣΜΚ είναι: α) πώς ενσωματώνονται τα δεδομένα των ΗΣΚ στη γλωσσολογική ανάλυση και β) ποια είναι τα κριτήρια επιλογής της γλωσσικής ποικιλίας που θα αποτελέσει τη γλώσσα-στόχο με την οποία θα αντιπαραβληθούν τα δεδομένα του μαθητικού λόγου. Αναφορικά με το πρώτο ερώτημα, έχει επισημανθεί ότι για να είναι ερευνητικώς αξιοποιήσιμα τα δεδομένα ενός ΗΣΚ, πρέπει να είναι αυθεντικά και μηχανικώς επεξεργάσιμα καθώς και να πληρούν μια σειρά από επιπλέον κριτήρια. Το δεύτερο ερώτημα δε θα εξεταστεί στην παρούσα εργασία καθώς δεν είναι συναφές με τους στόχους της παρούσας έρευνας, η οποία εξετάζει αποκλειστικά διαγλωσσικά δεδομένα από μη φυσικούς ομιλητές της ΝΕ, χωρίς να τα αντιπαραβάλλει με γλωσσικά δεδομένα από φυσικούς ομιλητές της. Συνεπώς, ελλείπει ενός συγκρίσιμου ΗΣΚ φυσικών ομιλητών, το εύρος της ισχύος των πορισμάτων της έρευνας καθορίζεται αυστηρά από τα σχεδιαστικά χαρακτηριστικά και το περιεχόμενο του ΗΣΜΚ που χρησιμοποιεί.

2.3.5.1.2.3. Διατομεακά (cross-sectional) έναντι διαχρονικών (longitudinal) δεδομένων

Μία ακόμη παράμετρος με καταλυτική επίδραση στα πορίσματα των ερευνών για την εκμάθηση της Γ2 είναι η χρονική διάρκεια δειγματοληψίας του πρωτογενούς γλωσσικού υλικού από τους/τις μαθητές/-τριες. Με βάση την παράμετρο αυτή, τα ΗΣΜΚ, όπως και κάθε ΗΣΚ, μπορούν να διακριθούν αφενός σε δειγματικά ή διατομεακά (sample ή cross-sectional), δηλαδή σώματα που περιέχουν ένα και μόνο στιγμιαίο δείγμα διαγλώσσας από κάθε πληροφορητή/-τρια, και σε διαχρονικά (diachronic/longitudinal), τα οποία περιέχουν αρκετά δείγματα της διαγλώσσας κάθε πληροφορητή/-τρια, κάθε ένα από τα οποία απέχει από το άλλο συγκεκριμένο χρονικό διάστημα (McEnery et al., 2006:15). Η πλειονότητα των ερευνών της κατάκτησης της Γ2 στηρίζεται εδώ και πολλά χρόνια σε ΗΣΚ του πρώτου είδους. Όσον αφορά το ΗΣΜΚ στο οποίο βασίστηκε η παρούσα έρευνα, σύμφωνα με τα όσα εκτίθενται στην ενότητα 3.1.1. για τον σχεδιασμό και το περιεχόμενό του, μπορεί να θεωρηθεί ως καθαρά διατομεακού χαρακτήρα, καθώς κάθε περιλαμβανόμενη σε αυτό ΠΓΛ προέρχεται από διαφορετικό/-η μαθητή/-τρια και, επομένως, συνιστά ένα στιγμιαίο δείγμα της διαγλώσσας του.

2.3.5.1.3. Πρακτικές εφαρμογές των ΗΣΜΚ

Παρά τη σύντομη ιστορία των ερευνών βάσει των ΗΣΜΚ, οι εφαρμογές τους είναι ποικίλες και αφορούν τόσο τους/τις ερευνητές/-τριες της Γ2 όσο και τους/τις διδάσκοντες/-ουσες της δεύτερης/ξένης γλώσσας.

2.3.5.1.3.1. Εφαρμογές στην έρευνα για την κατάκτηση της δεύτερης γλώσσας

Όσον αφορά την έρευνα για την κατάκτηση της δεύτερης γλώσσας (Second Language Acquisition), η οποία προσπαθεί να ερμηνεύσει τους μηχανισμούς της κατάκτησης μιας δεύτερης/ξένης γλώσσας, τα δεδομένα των ΗΣΜΚ μπορούν να υποστούν επεξεργασία προκειμένου να διενεργηθούν:

- η αντιπαραβολική ανάλυση δεδομένων μεταξύ α) μη φυσικών και φυσικών ομιλητών/-τριών²⁵ αλλά και β) μεταξύ διαφορετικών ομάδων μη φυσικών ομιλητών/-τριών μόνο όπου οι αποκλίσεις θεωρούνται άμεσα σχετιζόμενες με την εκάστοτε Γ1. Με άλλα λόγια, τα ΗΣΜΚ μπορούν να συμβάλουν στη διάκριση μεταξύ διαγλωσσικών στοιχείων καθολικού χαρακτήρα, τα οποία δεν σχετίζονται με τη μητρική γλώσσα των μαθητών/-τριών και διαγλωσσικών στοιχείων που αποτελούν αποτέλεσμα παρεμβολής από συγκεκριμένη Γ1 (Granger, 2002:13· Granger, 2003:541).
- η υπολογιστική ανάλυση λαθών (ΥΑΛ): η εν λόγω ανάλυση αποτελεί μια από τις κυριότερες επιστημονικές συμβολές των ΗΣΜΚ και χρησιμοποιείται σε ένα ευρύ φάσμα εφαρμογών τόσο στον ερευνητικό όσο και κυρίως στον διδακτικό τομέα, από τον οποίο άλλωστε πρωτοξεκίνησε (Ellis, 1994:48). Ο βασικότερος λόγος για τον κεντρικό χαρακτήρα της ΥΑΛ στον χώρο των μελετών της διαγλώσσας είναι ότι ο γραπτός λόγος του μαθητικού κοινού, ακόμη και σε υψηλά επίπεδα γλωσσικής επάρκειας, χαρακτηρίζεται από πολύ μεγαλύτερες συχνότητες λαθών σε σχέση με τον γραπτό λόγο των φυσικών ομιλητών/-τριών (Ηλιοπούλου, 2009· Granger, 2003b:542). Γενικά, η ΥΑΛ παρέχει πολύτιμες πληροφορίες στη γενικότερη προσπάθεια για την

²⁵ Για περισσότερες πληροφορίες βλέπε Granger, 2002: 12-13· Granger, 2003b:541-542· De Cock κ.α., 1998: 68-69.

κατανόηση της εξέλιξης της διαγλώσσας, συνιστώντας ταυτόχρονα ένα πολύτιμο εργαλείο για την επίτευξη βασικών στόχων της διδασκαλίας της Γ2.

2.3.5.1.3.2. Εφαρμογές στη διδασκαλία της δεύτερης/ξένης γλώσσας

Η εκμάθηση βάσει ηλεκτρονικών δεδομένων (data-driven learning), αν και σχετικά νέα ως πρακτική, ανοίγει πολλές και σημαντικές προοπτικές τόσο για τον/την διδάσκοντα/-ουσα όσο και για τον/την μαθητή/-τρια της τάξης της ξένης γλώσσας. Σύμφωνα με τους Granger & Tribble (1997:200), η εν λόγω διδακτική μεθοδολογία εντάσσεται στη γενικότερη προσπάθεια αντιμετώπισης των προβλημάτων σε θέματα γλωσσικής ακρίβειας που προκύπτουν στο πλαίσιο της επικοινωνιακής διδακτικής προσέγγισης και εστιάζει στον γλωσσικό τύπο χρησιμοποιώντας στην τάξη δεδομένα ηλεκτρονικών συμφραστικών πινάκων αντλημένων από ΗΣΚ φυσικών ομιλητών/-τριών ή μη φυσικών ομιλητών/-τριών που μαθαίνουν μια δεύτερη/ξένη γλώσσα. Μερικά από αυτά σχετίζονται με: α) το επίπεδο γλωσσομάθειας στο οποίο ενδείκνυται η χρήση των δεδομένων των ΗΣΚ, β) τις μαθησιακές στρατηγικές με τις οποίες αλληλεπιδρούν πιο αποτελεσματικά και γ) τα κριτήρια επιλογής και επεξεργασίας των δεδομένων των συμφραστικών πινάκων, ώστε να προσιδιάζουν στο εκάστοτε διδασκόμενο γλωσσικό φαινόμενο και να μην κουράζουν τους/τις μαθητές/-τριες. Οι ίδιοι/-ες ερευνητές/-τριες θεωρούν ότι τα δεδομένα των ΗΣΜΚ είναι προτιμότερο να χρησιμοποιούνται σε προχωρημένο επίπεδο γλωσσομάθειας ως συμπλήρωμα και όχι ως υποκατάστατο διδακτικών πακέτων παραδοσιακού τύπου και μόνο για περιορισμένο αριθμό και συγκεκριμένους τύπους μαθητών (Granger, 2002:26).

Οι κυριότεροι τομείς της διδασκαλίας της δεύτερης/ξένης γλώσσας που επωφελούνται είναι ο σχεδιασμός διδακτικού υλικού, ο σχεδιασμός και η ανάπτυξη προγραμμάτων σπουδών, η γλωσσική αξιολόγηση, η διδακτική μεθοδολογία, η ανάπτυξη λογισμικών γραμματικού και υφολογικού ελέγχου καθώς και η ανάπτυξη υπολογιστικών εργαλείων γλωσσικής εκμάθησης (Granger, 1999:200-201· Granger, 2002:21-26· Granger, 2003b:542-544 ·Granger κ.α., 2002:19-26, Granger & Tribble 1997:199-209 ·Meunier, 2002:119-137):

Η χρήση των ΗΣΚ στη διδασκαλία μπορεί να είναι άμεση ή έμμεση. Η άμεση χρήση των ΗΣΚ αναφέρεται στην απευθείας πρόσβασή τους από τους/τις μαθητές/-τριες μέσω των ηλεκτρονικών υπολογιστών προκειμένου να παρατηρήσουν τη χρήση ενός

γλωσσικού φαινομένου και να απαντήσουν κάποιες ερωτήσεις, δικές τους ή του/της καθηγητή/-τριας τους, ή να κάνουν κάποιες δραστηριότητες που τους έχουν ανατεθεί. Συνήθως οι δραστηριότητες αυτές έχουν στόχο την ενσυνείδηση (awareness raising) (Hunston, 2002). Η έμμεση χρήση των ΗΣΚ είναι πιο ευρεία και επιτρέπει στον/στην καθηγητή/-τρια να κάνει χρήση των ΗΣΚ χωρίς αυτό να γίνεται πάντοτε αντιληπτό από τους/τις μαθητές/-τριες του/της, αλλά πάντοτε με στόχο τη βελτίωση της διδασκαλίας και εκμάθησης της Γ2.

Τα σώματα κειμένων είναι ιδιαίτερα βοηθητικά για τους/τις εκπαιδευτικούς, καθώς προσφέρουν πληθώρα δεδομένων αυθεντικού λόγου και μπορούν να αποτελέσουν μια επιπλέον πηγή πληροφόρησης σε σχέση με τη γραμματικότητα αλλά και τη χρήση κάποιων δομών. Εξάλλου, τα σώματα κειμένων δίνουν τη δυνατότητα στον/στην εκπαιδευτικό αφενός να εκθέτει τους/τις σπουδαστές/-στριες σε ένα μαθησιακό περιβάλλον πλούσιο σε γλωσσικά δεδομένα, και αφετέρου να κατασκευάζει εξειδικευμένες γραμματικές ή λεξιλογικές δραστηριότητες κατάλληλες για το γλωσσικό επίπεδο και τις ανάγκες των σπουδαστών/-τριών του/της (Roemer, 2008:120). Στην παρούσα εργασία, έμφαση θα δοθεί στις εφαρμογές των ΗΣΜΚ στη διδασκαλία της δεύτερης γλώσσας.

ΚΕΦΑΛΑΙΟ 3: ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ

Στην παρούσα έρευνα, όπως έχει ήδη επισημανθεί, διερευνάται, κατ' αρχάς, η πιθανή συσχέτιση έξι δεικτών λεξιλογικού «πλούτου» με τα επίπεδα γλωσσομάθειας στα οποία κατατάχτηκαν οι εξεταζόμενοι/-ες προτού παρακολουθήσουν μαθήματα ελληνικών. Δευτερευόντως, σε περίπτωση που αποδειχθεί η συσχέτιση των/κάποιων δεικτών με τα επίπεδα γλωσσομάθειας, προβλέπεται να εξεταστεί αν λήφθηκαν υπόψη στη βαθμολόγηση του λεξιλογίου στις ΠΓΛ των μαθητών/-τριών και δεν αποκλείεται το ενδεχόμενο αναθεώρησης του ισχύοντος πλαισίου, προκειμένου η βαθμολόγηση του λεξιλογίου να είναι πιο αντικειμενική. Για τη διερεύνηση των παραπάνω, αξιοποιήθηκε ένα ΗΣΜΚ που συμβατικά, στο πλαίσιο της συγκεκριμένης έρευνας, ονομάζεται «Σώμα Ελληνικών Μαθητικών Κειμένων» (στο εξής ΣΕΜΚ), ονομασία που προτείνεται από τον Τζιμώκα (2018) για το σύνολο των συγκεκριμένων ΠΓΛ. Το ΣΕΜΚ αποτελεί απότοκο του ερευνητικού προγράμματος «Πυθαγόρας Ι» και στη συγκεκριμένη εργασία αξιοποιείται ως πρωτογενές υλικό για την εξαγωγή συμπερασμάτων σχετικά με συγκεκριμένους λεξιλογικούς δείκτες και τα επίπεδα ελληνομάθειας. Όσον αφορά τον σχεδιασμό, τη δομή, το μέγεθος και άλλα περιγραφικά στοιχεία του ΣΕΜΚ, καθοριστική στάθηκε η συμβολή του Δημήτρη Τζιμώκα (2018) ο οποίος, στο πλαίσιο της διατριβής του, αναφέρεται λεπτομερώς στην ανάπτυξη και περιγραφή αυτού του ΗΣΜΚ. Στο κεφάλαιο αυτό, καταγράφονται λεπτομερώς το περιεχόμενο και τα σχεδιαστικά του κριτήρια, επισημαίνονται τόσο τα πλεονεκτήματα όσο και οι περιορισμοί του καθώς και οι ανάλογες προεκτάσεις τους στα πορίσματα της έρευνας. Επιπλέον, παρουσιάζεται το πρόγραμμα ποσοτικής ανάλυσης του ΗΣΜΚ το οποίο αξιοποιήθηκε ώστε να εξαχθούν οι προς μελέτη δείκτες λεξιλογικού «πλούτου».

3.1. Περιγραφή του Σώματος Ελληνικών Μαθητικών Κειμένων (ΣΕΜΚ)

Στην ενότητα αυτή θα παρουσιαστούν οι σχεδιαστικές λεπτομέρειες καθώς και τα εργαλεία επεξεργασίας του ΣΕΜΚ, προκειμένου να γίνει αντιληπτός ο χαρακτήρας

του πρωτογενούς υλικού βάσει του οποίου έγιναν οι μετρήσεις και εξήχθησαν τα συμπεράσματα της παρούσας έρευνας αναφορικά με τους δείκτες λεξιλογικού «πλούτου» που εντοπίστηκαν στα γραπτά των μαθητών/-τριών της ελληνικής ως δεύτερης γλώσσας (Γ2).

3.1.1. Περιεχόμενο και σχεδιαστικά κριτήρια

Όπως επισημαίνει και ο Τζιμώκας (2018), το ΣΕΜΚ που παρουσιάζεται εδώ εντάσσεται στο Ηλεκτρονικό Σώμα Μαθητικού Λόγου (Computer Learner Corpus) ή ΗΣΜΚ, αλλά δεν μπορεί να θεωρηθεί ένα επαρκώς τεκμηριωμένο ΗΣΚ. Η αδυναμία αυτή προκύπτει από το γεγονός ότι η δειγματοληψία των ΠΓΛ έγινε πολύ αργότερα από την παραγωγή τους και δεν ήταν δυνατή η χρήση ερωτηματολογίων ή ειδικών εντύπων για την λεπτομερή καταγραφή όσο το δυνατόν περισσότερων μεταβλητών των μαθητών/-τριών και των ΠΓΛ. Αυτό είχε ως αποτέλεσμα, η τεκμηρίωση των μεταβλητών αυτών να βασιστεί κατά κύριο λόγο στα στοιχεία που συμπλήρωναν οι ίδιοι/-ες οι εξεταζόμενοι/-ες και οι βαθμολογητές/-τριες τους στην πρώτη σελίδα του τετραδίου με τα θέματα της γραπτής εξέτασης καθώς και στις σελίδες με τη δραστηριότητα της παραγωγής προφορικού λόγου. Συγκεκριμένα, καταγράφηκαν 18 μεταβλητές των μαθητών/-τριών και των ΠΓΛ για καθένα από τα γραπτά κείμενα του ΣΕΜΚ. Όπως φαίνεται στον Πίνακα 4 στην ενότητα 3.1.1.1, κάποιες από αυτές τις μεταβλητές (μέσο παραγωγής λόγου, κειμενικό γένος, μέσο μέγεθος ΠΓΛ, επίπεδο ελληνομάθειας, πλαίσιο εκμάθησης της ελληνικής γλώσσας) χρησιμοποιήθηκαν ως κριτήρια σχεδιασμού και γι' αυτό είναι κοινές για όλα τα κείμενα του σώματος. Κάθε κείμενο του σώματος διαθέτει κωδικοποιημένο όνομα σε αρχείο λογιστικού φύλλου μέσω του λογισμικού Excel του Microsoft Office, στο οποίο αποθηκεύτηκαν όλες οι μεταβλητές. Κάθε κωδικοποιημένο όνομα είναι ταυτόσημο με το όνομα του αντίστοιχου αρχείου «απλού κειμένου» (.txt) που περιέχει την ΠΓΛ.

3.1.1.1. Μεταβλητές των μαθητών/-τριών

Οι μαθητές/-τριες που συμμετείχαν στη δημιουργία του ΣΕΜΚ μοιράζονται αρκετά κοινά γνωρίσματα στη συντριπτική τους πλειοψηφία. Υπάρχουν, εντούτοις, και ορισμένα διαφοροποιητικά χαρακτηριστικά.

3.1.1.1.1. Κοινά χαρακτηριστικά

Τα δεδομένα των μαθητών/-τριών που συμμετείχαν στη δημιουργία του συγκεκριμένου ΗΣΜΚ αποδεικνύουν ότι πολλά χαρακτηριστικά είναι κοινά. Αυτά σχετίζονται με: α) την ηλικία, β) τις συνθήκες εκμάθησης της ελληνικής, γ) το κίνητρό τους για την εκμάθηση της νέας ελληνικής και δ) το επίπεδο ελληνομάθειας. Συγκεκριμένα, στη συντριπτική τους πλειοψηφία ήταν νεαροί/-ες ενήλικες (από 20 έως 30 ετών περίπου), όλοι/όλες μαθητές/-τριες του Διδασκαλείου Νέας Ελληνικής Γλώσσας της Φιλοσοφικής Σχολής του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών που είχαν παρακολουθήσει οκτάμηνο πρόγραμμα γλωσσικής διδασκαλίας επί 15 ώρες εβδομαδιαίως. Τα παραπάνω συντείνουν στο ότι όλοι/-ες είχαν παρακολουθήσει γύρω στις 480 ώρες εντατικής διδασκαλίας και ότι μάθαιναν τη νέα ελληνική ως δεύτερη και όχι ως ξένη γλώσσα. Ακόμη, παρά τη διαφοροποίηση στον χρόνο εξέτασης,²⁶ όσον αφορά το διδακτικό υλικό, αξίζει να ειπωθεί ότι τους χορηγήθηκε ανάλογο πάντα με το επίπεδο ελληνομάθειας στο οποίο κατατάχθηκαν στην αρχή των μαθημάτων, εφόσον κατά τη διάρκεια της τετραετίας από την οποία προέρχεται το υλικό του ΣΕΜΚ δεν σημειώθηκε κάποια αξιοσημείωτη αλλαγή στα διδακτικά εγχειρίδια του Διδασκαλείου του ΕΚΠΑ²⁷. Πέρα από την ηλικία και τις συνθήκες εκμάθησης, ένα ακόμα κοινό γνώρισμα των μαθητών/-τριών συνιστά η πρόθεση. Σχεδόν όλοι/-ες οι συμμετέχοντες/-ουσες μάθαιναν ελληνικά έχοντας ως στόχο την απόκτηση δεξιοτήτων ανεξάρτητου προς ικανού χρήστη της ελληνικής που να μπορεί να ανταποκρίνεται στις βασικές επικοινωνιακές και γνωστικές απαιτήσεις ενός ακαδημαϊκού περιβάλλοντος και πιο συγκεκριμένα της ελληνικής τριτοβάθμιας εκπαίδευσης. Το ίδιο, άλλωστε, το Διδασκαλείο Νέας Ελληνικής Γλώσσας εξυπηρετεί μεταξύ άλλων και τον σκοπό της προετοιμασίας των αλλοδαπών μαθητών/-τριών για την απόκτηση της βεβαίωσης ελληνομάθειας, η οποία αποτελεί προαπαιτούμενο τυπικό προσόν για την εισαγωγή τους στα ελληνικά Α.Ε.Ι. και Τ.Ε.Ι. Το πλαίσιο αυτό είναι προκαθορισμένο από το αναλυτικό πρόγραμμα σπουδών του Διδασκαλείου αλλά και από το επίπεδο των τελικών εξετάσεων ελληνομάθειας, των οποίων τα κριτήρια έχουν καθοριστεί από το Συμβούλιο της Ευρώπης για το Επίπεδο

²⁶ Από το σύνολο των 317 πληροφορητών 21 εξετάστηκαν τον Ιούνιο του 2001, 35 εξετάστηκαν τον Σεπτέμβριο του 2001, 145 τον Ιούνιο του 2002, 8 τον Σεπτέμβριο του 2002, 48 τον Ιούνιο του 2004, 7 τον Σεπτέμβριο του 2004 και 53 τον Ιούνιο του 2005 (Τζιμώκας, 2018).

Επάρκειας (Vantage Level),²⁸ το οποίο σύμφωνα με το ίδιο το ΚΕΠΑ για τις γλώσσες του Συμβουλίου της Ευρώπης (2008: 54-55) αντιστοιχεί στο επίπεδο B2.

Ως εκ τούτου, όλες οι ΠΓΛ του ΣΕΜΚ μπορούν τυπικά να θεωρηθούν επιπέδου B2. Αυτό βεβαιώνει το γεγονός ότι αποτελούν απαντήσεις σε εξεταστικά θέματα αυτού του επιπέδου και ότι επιλέχθηκαν μόνο όσες είχαν συγκεντρώσει τουλάχιστον το 60% ή ανώτερο της συνολικής βαθμολογίας (12 ή ανώτερο με άριστο το 20), ποσοστό που καθιστά κάποιον/-α προβιβαστέο/-εα και το γραπτό –τυπικά τουλάχιστον- επιτυχές.

3.1.1.1.2. Το επίπεδο γλωσσομάθειας των μαθητών

Παρά την τυπική κατάταξη των μαθητών/-τριών στο B2 επίπεδο, το πραγματικό επίπεδο ελληνομάθειάς τους ποικίλλει και αυτό οφείλεται στη γλωσσική τους αφετηρία κατά την ένταξή τους στον συγκεκριμένο εκπαιδευτικό φορέα αλλά και στην τελική τους εικόνα. Σχετικές πληροφορίες λαμβάνουμε από την πρώτη σελίδα του εξεταστικού τετραδίου. Σε αυτήν, οι εξεταζόμενοι/-ες καλούνταν να δηλώσουν σε ποιο από τα εξής τρία επίπεδα ανήκαν: Α (αρχάριο), Β (μέσο) και Γ (προχωρημένο). Αυτά τα επίπεδα ελληνομάθειας είχαν προκύψει από την κατατακτήρια γραπτή δοκιμασία στην οποία υποβάλλονται όλοι/-ες οι νεοεισερχόμενοι/-ες μαθητές/-τριες στο Διδασκαλείο Νέας Ελληνικής Γλώσσας του ΕΚΠΑ, προκειμένου να εκτιμηθεί ο βαθμός προηγούμενης γνώσης της ελληνικής που έχουν, ώστε να ενταχθούν στο ανάλογο πρόγραμμα σπουδών: αρχαρίων, μέσων, προχωρημένων.²⁹ Η μεταβλητή αυτή δηλωνόταν από τους/τις ίδιους/-ες τους/τις μαθητές/-τριες, όπως έχει καταγραφεί στα μεταδεδομένα του ΣΕΜΚ. Είναι σημαντικό να ληφθεί υπόψη καθώς δεν θεωρείται άσχετη με το επίπεδο ελληνομάθειας του μαθητικού κοινού κατά την τελική του εξέταση, της οποίας προϊόν απετέλεσαν οι ΠΓΛ του ΣΕΜΚ.

²⁸Βλ. στην ιστοσελίδα :

http://www.greekcourses.uoa.gr/fileadmin/greekcourses.uoa.gr/uploads/DOCUMENTS_DIAFORA/TEST_ELLHNOMATHEIAS.pdf του Διδασκαλείου της Νέας Ελληνικής ως Ξένης Γλώσσας, όπου βρίσκεται αναρτημένο το σχετικό έγγραφο, στην προμετωπίδα του οποίου δηλώνεται επισήμως ότι το επίπεδο της σχετικής εξέτασης είναι το επίπεδο επάρκειας (Vantage).

²⁹ Για μία αναλυτική περιγραφή των τυπικών χαρακτηριστικών των τριών αυτών επιπέδων προκειμένου να γίνουν αντιληπτές οι διαφορές μεταξύ τους ως προς τις προβλεπόμενες γλωσσικές γνώσεις και ικανότητες των μαθητών βλ. Ιακώβου 2003: 167-171.

3.1.1.1.3. Εμπόδια στην κατάταξη των υποψηφίων στο κατάλληλο επίπεδο

Το επίπεδο γλωσσομάθειας κάθε πληροφορητή/-τριας είναι δύσκολο να προσδιοριστεί με ακρίβεια λόγω: α) των μεγάλων διαφορών μεταξύ των μαθητών/-τριών ως προς το αρχικό επίπεδο ελληνομάθειας, β) της απουσίας στοιχείων που να τεκμηριώνουν την ενδιάμεση πρόοδο τους και γ) της ανάγκης για περαιτέρω διερεύνηση της τελικής γραπτής δοκιμασίας για τη χορήγηση του πιστοποιητικού ελληνομάθειας για το κατά πόσο είναι σταθμισμένη στο επίπεδο B2. Λαμβάνοντας υπόψη τους παραπάνω σκοπέλους, γίνεται αντιληπτό ότι τα μόνα κοινά ποσοτικά στοιχεία που μοιράζονται όλοι/-ες οι πληροφορητές/-τριες είναι η υπέρβαση του κατώτατου βαθμολογικού ορίου του 60%, που αυτόματα χρήζει τον/την εξεταζόμενο/-η προβιβαστέο/-έα στο B2 επίπεδο και το σύνολο των ωρών διδασκαλίας που παρακολούθησαν έως τη στιγμή που συνεξετάστηκαν οι ΠΓΛ. Συγκεκριμένα, σύμφωνα με την αντιστοίχιση ωρών διδασκαλίας και επιπέδων ελληνομάθειας του επίσημου ιστοτόπου για την πιστοποίηση επάρκειας της ελληνομάθειας του ΚΕΓ, οι υποψήφιοι/-ες του επιπέδου B2 προβλέπεται να έχουν παρακολουθήσει από 350 έως 450 ώρες μαθημάτων.³⁰ Αυτό σημαίνει ότι οι πληροφορητές/-τριες του ΣΕΜΚ με τις 480 ώρες παρακολούθησης μαθημάτων στο ενεργητικό τους μπορούν να καταταχθούν συμβατικά στο επίπεδο B2. Με δεδομένη, όμως, τη διαφοροποίησή τους αφενός ως προς το επίπεδό τους κατά την έναρξη των μαθημάτων και αφετέρου ως προς τη βαθμολογία που έλαβαν οι ΠΓΛ, εκτιμάται ότι το επίπεδο B2 μπορεί να θεωρηθεί ως μέσος όρος των τελικών επιπέδων ελληνομάθειας όλων των μαθητών/-τριών του ΣΕΜΚ με το εύρος διακύμανσης του τελικού επιπέδου ελληνομάθειας των πληροφορητών/-τριών του ΣΕΜΚ να εντοπίζεται μεταξύ του B1 και του Γ1 και σε ελάχιστες περιπτώσεις έως και του Γ2. Για τους λόγους αυτούς μπορεί να γίνει μόνο μία - κατά προσέγγιση- εκτίμηση του εύρους διακύμανσης του τελικού επιπέδου ελληνομάθειας για το σύνολο των μαθητών/-τριών.

³⁰ Βλ. συγκεκριμένα στην ιστοσελίδα <http://www.greek-language.gr/certification/node/114.html> και ειδικότερα στην παράγραφο «Πόσες ώρες πρέπει να διδάξω;», όπου διευκρινίζεται, βέβαια, ότι οι αριθμοί ωρών διδασκαλίας για κάθε επίπεδο είναι ενδεικτικοί και δεν αποτελούν ασφαλές κριτήριο κατάταξης στο καθένα.

3.1.1.1.4. Διαφοροποιητικά στοιχεία

Όσον αφορά τις μεταβλητές που διαφέρουν μεταξύ των μαθητών/-τριών, αυτές είναι: α) η χώρα προέλευσης και β) το φύλο. Συγκεκριμένα, γίνεται λόγος για χώρες προέλευσης και όχι για μητρικές γλώσσες των μαθητών/-τριών, αφού μόνο τις πρώτες δήλωναν στην πρώτη σελίδα του εξεταστικού τετραδίου. Έτσι, από το ΣΕΜΚ πληροφορούμαστε τις εξής 39 διαφορετικές χώρες: 1. Αίγυπτος, 2. Αλβανία, 3. Αργεντινή, 4. Αρμενία, 5. Αυστραλία, 6. Αυστρία, 7. Βέλγιο, 8. Βενεζουέλα, 9. Βουλγαρία, 10. Μ. Βρετανία, 11. Γαλλία, 12. Γερμανία, 13. Γεωργία, 14. Ζιμπάμπουε, 15. Η.Π.Α., 16. Ιορδανία, 17. Ιράκ, 18. Ισπανία, 19. Καμερούν, 20. Καναδάς, 21. Κένυα, 22. Κονγκό, 23. Λευκορωσία, 24. Λίβανος, 25. Μεξικό, 26. Μολδαβία, 27. Μπουρουντί, 28. Νιγηρία, 29. Ουκρανία, 30. Παλαιστίνη, 31. Περού, 32. Πολωνία, 33. Ρουμανία, 34. Ρωσία, 35. Σερβία, 36. Συρία, 37. Τουρκία, 38. Τυνησία και 39. Χιλή. Επίσης, υπάρχουν τρεις ΠΓΛ στο ΣΕΜΚ (709 λέξεων συνολικά) των οποίων οι πληροφορητές/-τριες δεν είχαν δηλώσει συγκεκριμένη χώρα προέλευσης, αλλά είναι πιστοποιημένη η αραβική ως μητρική τους γλώσσα. Με βάση την κυρίαρχη γλώσσα στις προηγούμενες χώρες προέλευσης των πληροφορητών/-τριών, οι τελευταίοι κατανεμήθηκαν σε 16 γλωσσικές ομάδες, οι οποίες, ελλείψει της σχετικής τεκμηρίωσης, μπορεί να θεωρηθεί ότι αντιστοιχούν μόνο κατά προσέγγιση με τις πραγματικές μητρικές γλώσσες των μαθητών. Στον Πίνακα 2 παρατίθενται οι γλωσσικές ομάδες των πληροφορητών/-τριών του ΣΕΜΚ και οι αντίστοιχες σε καθεμιά χώρες προέλευσής τους, όπως αποτυπώνονται στον Τζιμώκα (2018).

Πίνακας 2: Γλωσσικές ομάδες των πληροφορητών/-τριών του ΣΕΜΚ και χώρα Προέλευσης

Γλωσσική ομάδα	Χώρες προέλευσης
Αραβικά	Αίγυπτος, Ιορδανία, Ιράκ, Λίβανος, Παλαιστίνη, Συρία, Τυνησία
αλβανικά	Αλβανία
ισπανικά	Αργεντινή, Βενεζουέλα, Ισπανία, Μεξικό, Περού, Χιλή
αρμενικά	Αρμενία
αγγλικά	Αυστραλία, Βρετανία, Η.Π.Α., Καναδάς
γερμανικά	Αυστρία, Γερμανία
γαλλικά	Βέλγιο, Γαλλία
βουλγαρικά	Βουλγαρία
γεωργιανά	Γεωργία
νιγηροκογκολεξικές γλώσσες	Ζιμπάμπουε, Καμερούν, Κένυα, Κονγκό, Μπουρουντί, Νιγηρία
ρωσικά	Λευκορωσία, Ρωσία
ρουμανικά	Μολδαβία, Ρουμανία
ουκρανικά	Ουκρανία
πολωνικά	Πολωνία
σερβικά	Σερβία
τουρκικά	Τουρκία

Από τις 317 ΠΓΛ των συμμετεχόντων/-ουσών, στην παρούσα έρευνα αξιοποιήθηκαν οι 224, καθώς το εύρος των λέξεων περιορίστηκε στις 180-300, αποκλείοντας όσες ΠΓΛ υπερέβαιναν το όριο των 300 ή δεν άγγιζαν τις 180 λέξεις. Στο συγκεκριμένο εύρος συναντούμε 33 χώρες: 1. Αλβανία (23 ΠΓΛ), 2. Αίγυπτος (4 ΠΓΛ), 3. Αρμενία (4 ΠΓΛ), 4. Αυστρία (1 ΠΓΛ), 5. Αργεντινή (1 ΠΓΛ), 6. Βουλγαρία (22 ΠΓΛ), 7. Βέλγιο (2 ΠΓΛ), 8. Βρετανία (3ΠΓΛ), 9. Γαλλία (5 ΠΓΛ), 10. Γερμανία (13 ΠΓΛ), 11. Γεωργία (11 ΠΓΛ), 12. Ζιμπάμπουε (1 ΠΓΛ), 13. Η.Π.Α. (4 ΠΓΛ), 14. Ιορδανία (2 ΠΓΛ), 15. Ιράκ (1 ΠΓΛ), 16. Ισπανία (3 ΠΓΛ), 17. Καμερούν (2 ΠΓΛ), 18. Καναδά (1 ΠΓΛ) 19. Κονγκό (3ΠΓΛ), 20. Λευκωσία (1 ΠΓΛ), 21. Λίβανος, 22. Μεξικό (2 ΠΓΛ), 23. Μολδαβία (6 ΠΓΛ), 24. Νιγηρία (1 ΠΓΛ), 25. Ουκρανία (24 ΠΓΛ), 26. Παλαιστίνη (3 ΠΓΛ), 27. Πολωνία (10 ΠΓΛ), 28. Ρουμανία (13 ΠΓΛ), 29. Ρωσία (23 ΠΓΛ), 30. Σερβία (14 ΠΓΛ), 31. Συρία (10 ΠΓΛ), 32. Τουρκία (6 ΠΓΛ), 33. Χιλή (1 ΠΓΛ). Άλλες τρεις (3) ΠΓΛ είναι γραμμένες στην αραβική γλώσσα χωρίς να δηλώνεται η χώρα προέλευσης τους, όπως προαναφέρθηκε. Οι χώρες προέλευσης των ερευνητικών υποκειμένων και το πλήθος των ΠΓΛ που αντιστοιχούν σε κάθε χώρα παρατίθενται επιγραμματικά και στον Πίνακα 3:

Πίνακας 3: Χώρες προέλευσης των ερευνητικών υποκειμένων και το πλήθος των 224 ΠΓΛ που αντιστοιχούν σε κάθε χώρα

Χώρες Προέλευσης	ΠΓΛ
Αλβανία	23
Αίγυπτος	4
Αραβικά*	3
Αρμενία	4
Αυστρία	1
Αργεντινή	1
Βουλγαρία	22
Βέλγιο	2
Βρετανία	3
Γαλλία	5
Γερμανία	13
Γεωργία	11
Ζιμπάμπουε	1
Η.Π.Α.	4
Ιορδανία	2
Ιρακ	1
Ισπανία	3

Καμερούν	2
Καναδάς	1
Κονγκό	3
Λευκωσία	1
Λίβανος	1
Μεξικό	2
Μολδαβία	6
Νιγηρία	1
Ουκρανία	24
Παλαιστίνη	3
Πολωνία	10
Ρουμανία	13
Ρωσία	23
Σερβία	14
Συρία	10
Τουρκία	6
Χιλή	1
Σύνολο	224

Τέλος, το φύλο δεν δηλώνονταν στο εξεταστικό τετράδιο, αλλά συναγόταν είτε από το ονοματεπώνυμο του/της εξεταζόμενου/-νης είτε από αναφορές μέσα στην ίδια την ΠΓΛ είτε και από τα δύο. Με βάση αυτά τα στοιχεία, τα δεδομένα του ΣΕΜΚ προέρχονται από 76 άνδρες και 148 γυναίκες.

3.1.1.2. Μεταβλητές των παραγωγών λόγου

Όσον αφορά τα γνωρίσματα που αφορούν τις ίδιες τις υπό εξέταση ΠΓΛ των μαθητών/-τριών, εντοπίζονται τόσο ενοποιητικά όσο και διαφοροποιητικά χαρακτηριστικά, τα οποία χρήζουν περαιτέρω διερεύνησης.

3.1.1.2.1. Κοινά γνωρίσματα των ΠΓΛ

Πράγματι, όπως επισημαίνει ο Τζιμώκας (2018:86-87), όσον αφορά τις ΠΓΛ των μαθητών/-τριών που περιέχονται στο ΣΕΜΚ, μοιράζονται κατ' αρχάς ορισμένα κοινά χαρακτηριστικά: 1. είναι αποκλειστικά γραπτές, 2.παρήχθησαν στο σύνολό τους κατά τη διάρκεια των τελικών (γραπτών και προφορικών) εξετάσεων για τη χορήγηση βεβαίωσης ελληνομάθειας του Διδασκαλείου της Νέας Ελληνικής Γλώσσας του

ΕΚΠΑ, 3. συντάχθηκαν στο πλαίσιο χρονικού περιορισμού, 4. κατά την παραγωγή τους οι μαθητές/-τριες δεν χρησιμοποίησαν υλικό αναφοράς, όπως λεξικά και γραμματικές, και 5. ως προς το γλωσσικό τους πεδίο συνιστούν εν γένει δείγματα νέας ελληνικής για γενικούς παρά για ειδικούς σκοπούς.

3.1.1.2.2. Διαφοροποιητικά στοιχεία των ΠΓΛ

Υπάρχουν, ωστόσο, και μεταβλητές μεταξύ των ΠΓΛ του ΣΕΜΚ που ποικίλλουν και ήταν διαθέσιμες κατά τη συλλογή και καταχώριση των δεδομένων. Αυτές είναι: 1. η εξεταστική περίοδος παραγωγής τους, 2. το μέγεθός τους σε αριθμό λέξεων, 3. το κειμενικό γένος τους, 4. το θέμα τους, 5. η βαθμολογική αξιολόγηση του συνόλου της γραπτής εξέτασης της οποίας τμήμα αποτελεί κάθε ΠΓΛ, 6. η βαθμολογική αξιολόγηση της ΠΓΛ (τόσο η ολιστική όσο και η αναλυτική για περιεχόμενο, γραμματική ακρίβεια και λεξιλόγιο ξεχωριστά). Ειδικότερα, συνελέγησαν ΠΓΛ από τις εξεταστικές περιόδους Ιουνίου και Σεπτεμβρίου 2001, Ιουνίου και Σεπτεμβρίου 2002, Ιουνίου και Σεπτεμβρίου 2004 και Ιουνίου 2005. Τέλος, στο ΣΕΜΚ εντάχθηκαν μόνο οι ΠΓΛ που συγκέντρωσαν ολιστική βαθμολογική αξιολόγηση ίση ή ανώτερη του 12 με άριστα το 20 (Τζιμώκας, 2018:87).

Στον Πίνακα 4 παρουσιάζονται αναλυτικά τόσο οι διαθέσιμες μεταβλητές (ΝΑΙ) όσο και οι μη διαθέσιμες προς καταχώριση μεταβλητές (ΟΧΙ) των πληροφορητών/-τριών και των ΠΓΛ του ΣΕΜΚ,³¹ όπως παρατίθεται από τον Τζιμώκα (2018:78). Κρίθηκε αναγκαίο να παρατεθούν και οι μη διαθέσιμες μεταβλητές προκειμένου το ΣΕΜΚ να καταστεί πιο εύκολα συγκρίσιμο με άλλα ομοειδή ΗΣΚ. Από τις διαθέσιμες μεταβλητές, όσες δεν έχουν αστερίσκο αντλήθηκαν από τα εξεταστικά τετράδια. Οι μεταβλητές εντός παρένθεσης, αν και ήταν διαθέσιμες στα εξεταστικά τετράδια, δεν κρίθηκε σκόπιμο να καταχωριστούν στα μεταδεδομένα του ΣΕΜΚ, ενώ τα στοιχεία των μεταβλητών με αστερίσκο συνήχθησαν λαμβάνοντας υπόψη τις γενικότερες συνθήκες παραγωγής των γραπτών κειμένων. Όπως γίνεται φανερό από τον παρακάτω Πίνακα, είναι διαθέσιμες όλες οι κοινές μεταβλητές τόσο των πληροφορητών/-τριών όσο και των ΠΓΛ και όλες οι διαφορετικές μεταβλητές των ΠΓΛ, ενώ αξιοσημείωτες ελλείψεις εντοπίζονται σε αρκετές από τις σημαντικότερες

³¹ Οι μεταβλητές σταχυολογήθηκαν από τη σχετική διεθνή βιβλιογραφία: Granger 2003b: 539-541, Ιακώβου κ.ά. 2004: 3, Ιακώβου κ.ά. 2008: 165, Prati Zagrebelsky 2004: 63-64, Φλώρου 2009:111 και είναι προσαρμοσμένες στις ανάγκες της έρευνας για την ελληνική ως δεύτερη/ξένη γλώσσα.

μεταβλητές των πληροφορητών/-τριών. Επιπλέον, εκτός από το μέσο παραγωγής λόγου και την παρουσία ή μη εξεταστικού πλαισίου, οι υπόλοιπες κοινές μεταβλητές τόσο των πληροφορητών/-τριών όσο και των ΠΓΛ συνάγονται από τις γενικότερες συνθήκες παραγωγής των ΠΓΛ.

Πίνακας 4: Διαθέσιμες και μη διαθέσιμες μεταβλητές των πληροφορητών/-τριών και των ΠΓΛ του ΣΕΜΚ

Μεταβλητές των πληροφορητών κοινές	Διαθεσιμότητα
Ηλικία (20 έως 30 ετών περίπου)	ΝΑΙ*
Χρονικό διάστημα φοίτησης σε σχολείο νέας ελληνικής γλώσσας για ξένους (8 μήνες, 540 ώρες περίπου)	ΝΑΙ*
Πλαίσιο εκμάθησης της ελληνικής (ως δεύτερη γλώσσα)	ΝΑΙ*
Κίνητρο/-α για την εκμάθηση της Ν.Ε. (φοίτηση σε ελληνικά Α.Ε.Ι./Τ.Ε.Ι.)	ΝΑΙ*
Επίπεδο ελληνομάθειας (B2)	ΝΑΙ*
Μεταβλητές των πληροφορητών διαφορετικές	Διαθεσιμότητα
(Επώνυμο μαθητή/-τριας)	(ΝΑΙ)
(Όνομα μαθητή/-τριας)	(ΝΑΙ)
Φύλο	ΝΑΙ*
Επάγγελμα	ΟΧΙ
Χώρα προέλευσης/τόπος καταγωγής	ΝΑΙ
Μητρική γλώσσα	ΟΧΙ
Μητρική γλώσσα πατέρα	ΟΧΙ
Μητρική γλώσσα μητέρας	ΟΧΙ
Ομιλούμενες γλώσσες στην οικογένεια	ΟΧΙ
Επίπεδο εκπαίδευσης	ΟΧΙ
Πρωτοβάθμια εκπαίδευση – γλώσσα διδασκαλίας	ΟΧΙ
Δευτεροβάθμια εκπαίδευση – γλώσσα διδασκαλίας	ΟΧΙ
Τριτοβάθμια εκπαίδευση – γλώσσα διδασκαλίας	ΟΧΙ
Χρονικό διάστημα σπουδών (έτη ή μήνες)	ΟΧΙ
Χρονικό διάστημα φοίτησης σε ελληνικό σχολείο (έτη ή μήνες)	ΟΧΙ
Χρονικό διάστημα παραμονής σε ελληνόφωνη χώρα (πού, για πόσο)	ΟΧΙ
Άλλες γνωστές ξένες γλώσσες και επίπεδό τους	ΟΧΙ
Μεταβλητές των ΠΓΛ κοινές	Διαθεσιμότητα
Μέσο παραγωγής λόγου (γραπτός λόγος)	ΝΑΙ
Εξετάσεις (ναι)	ΝΑΙ
Χρονικός περιορισμός (ναι)	ΝΑΙ*
Χρήση υλικού αναφοράς όπως λεξικά, γραμματικές κτλ. (όχι)	ΝΑΙ*
Γλωσσικό πεδίο (ελληνικά για γενικούς σκοπούς)	ΝΑΙ*
Μεταβλητές των ΠΓΛ διαφορετικές	Διαθεσιμότητα
Χρονολογία και μήνας παραγωγής	ΝΑΙ
Μέγεθος σε αριθμό λέξεων	ΝΑΙ
Κεμενικό γένος	ΝΑΙ
Θέμα	ΝΑΙ
Βαθμολογική αξιολόγηση του συνόλου της γραπτής εξέτασης	ΝΑΙ
Βαθμολογική αξιολόγηση της ΠΓΛ (ολιστική και αναλυτική)	ΝΑΙ

3.1.1.3. Μέγεθος

Αν και ο αρχικός ποσοτικός στόχος των σχεδιαστικών προδιαγραφών της ερευνητικής ομάδας του προγράμματος «Πυθαγόρας Ι» ήταν η συλλογή 2100 ΠΓΛ, συνολικής έκτασης περίπου 200.000 λέξεων που θα ισοκατανέμονταν στα τρία

επίπεδα αρχικής κατάταξης των μαθητών/-τριών του Διδασκαλείου της Νέας Ελληνικής ως Δεύτερης Γλώσσας του Πανεπιστημίου Αθηνών, εν τέλει, το υλικό που αποτέλεσε το ΣΕΜΚ ανήλθε στις 76.989 λέξεις. Παρ' ότι η εργασία των πέντε διαφορετικών ομάδων συλλογών-διορθωτών/-τριών που εργάστηκαν κατά το χρονικό διάστημα 2004-2010 απέδωσε επισημειωμένο ως προς τα λάθη γραπτό μαθητικό λόγο της τάξεως των 252.535 λέξεων, το μεγαλύτερο μέρος αυτού του υλικού κρίθηκε ερευνητικά μη αξιοποιήσιμο λόγω των σημαντικών ελλείψεων στη διαθέσιμη τεκμηρίωσή του (Τζιμώκας, 2018:89). Στον Πίνακα 5 παρουσιάζονται τα βασικά ποσοτικά δεδομένα του ΣΕΜΚ.

Πίνακας 5: Βασικά ποσοτικά δεδομένα του αρχικού ΣΕΜΚ

Συνολικός αριθμός λέξεων	76.989
Εύρος έκτασης ΠΓΛ	91-530
Πλήθος χωρών προέλευσης μαθητών	39 (+3 ΠΓΛ στην αραβική γλώσσα χωρίς να δηλώνεται η χώρα προέλευσης τους)
Εκτιμώμενο πλήθος μητρικών γλωσσών	16
Πλήθος ΠΓΛ	317

Όσον αφορά την έκταση των επιμέρους ΠΓΛ, η σχετική οδηγία που παρέχεται στους/στις εξεταζόμενους/-ες ήταν το γραπτό τους να κυμαίνεται μεταξύ 220 και 250 λέξεων.³² Ωστόσο, όπως φαίνεται και στον προηγούμενο Πίνακα, η οδηγία αυτή δεν τηρήθηκε από όλους/όλες με αποτέλεσμα το σύνολο των λέξεων σε κάθε ΠΓΛ να μην είναι σταθερή παρουσιάζοντας αξιοσημείωτες αποκλίσεις. Το γεγονός αυτό οφείλεται στην ανομοιογένεια του πραγματικού επιπέδου ελληνομάθειας μεταξύ των μαθητών/-τριών και στον τυχαίο χαρακτήρα της δειγματοληψίας των ΠΓΛ. Πάντως, η έκταση 165 ΠΓΛ, δηλαδή του 52% επί του συνόλου των ΠΓΛ, είναι σε γενικές γραμμές σύμφωνη με την προβλεπόμενη, καθώς κυμαίνεται μεταξύ 200 και 270 λέξεων, ενώ φτάνουν τις 233 (δηλαδή το 73,5% του συνόλου) οι ΠΓΛ με έκταση από 180 έως 300 λέξεις. Ως εκ τούτου, οι ΠΓΛ με μεγάλη απόκλιση ως προς την έκταση από την αρχική οδηγία αποτελούν τη μειοψηφία στο ΣΕΜΚ (Τζιμώκας, 2018:90).

³² Στα τετράδια των εξετάσεων για τη χορήγηση της βεβαίωσης ελληνομάθειας του Διδασκαλείου της Νέας Ελληνικής Γλώσσας στο μέρος «Παραγωγή γραπτού λόγου» αναφέρεται επί λέξει η εξής οδηγία: «Διαλέξτε ένα από τα παρακάτω θέματα (220-250 λέξεις)».

Στην παρούσα έρευνα, για την απόδοση των δεικτών λεξιλογικού «πλούτου», αξιοποιήθηκε το εργαλείο ποσοτικοποίησης γραπτού λόγου QUITA, το οποίο παρουσιάζει μεγάλη ευαισθησία σε μεγάλες αποκλίσεις ως προς το μέγεθος των κειμένων. Εφόσον ετίθετο ζήτημα αξιοπιστίας των αποτελεσμάτων, θεωρήθηκε αναγκαίος ο περιορισμός του εύρους των λέξεων των ΠΓΛ. Προκειμένου το αποτέλεσμα να καταστεί έγκριτο και αξιοποιήσιμο από τον ερευνητικό χώρο, αφαιρέθηκαν οι ΠΓΛ που απέκλειαν σημαντικά από την προβλεπόμενο όριο³³ και το εύρος των λέξεων των ΠΓΛ περιορίστηκε στις 180-300, απομακρύνθηκαν, δηλαδή, όσες ΠΓΛ υπερέβαιναν το όριο των 300 ή δεν άγγιζαν τις 180 λέξεις. Τελικά, αξιοποιήθηκαν 224 ΠΓΛ, αντί για τις 233, για τις οποίες έγινε ήδη λόγος, λόγω ελλιπών πληροφοριών αναφορικά με την βαθμολογία του λεξιλογίου, γεγονός που τις καθιστούσε μη αξιοποιήσιμες δεδομένου του στόχου της παρούσας έρευνας. Στον Πίνακα 6 παρουσιάζονται τα βασικά ποσοτικά δεδομένα του ΣΕΜΚ των 224 ΠΓΛ.

Πίνακας 6: Βασικά ποσοτικά δεδομένα του ΣΕΜΚ των 224 ΠΓΛ

Συνολικός αριθμός λέξεων	52.924
Εύρος έκτασης ΠΓΛ	180-300
Πλήθος χωρών προέλευσης μαθητών	33 (+3 ΠΓΛ στην αραβική γλώσσα χωρίς να δηλώνεται η χώρα προέλευσης τους)
Εκτιμώμενο πλήθος μητρικών γλωσσών	16
Πλήθος ΠΓΛ	224

Τα ποσοτικά δεδομένα βάσει των μεταβλητών των πληροφορητών/-τριών που παρουσιάζουν τις σημαντικότερες διαφοροποιήσεις διακρίνονται σε τρία κύρια υποσώματα, δηλαδή του αρχικού επιπέδου κατάταξης, της εκτιμώμενης μητρικής γλώσσας (γλωσσική ομάδα) και του φύλου. Για κάθε υποσώμα παρατίθεται η έκτασή του σε πλήθος ΠΓΛ, σε αριθμό λέξεων και το ποσοστό που καταλαμβάνουν οι τελευταίες επί του αριθμού των λέξεων ολόκληρου του ΣΕΜΚ (52.924 λέξεις). Στον Πίνακα 7, ακολουθεί μια επιγραμματική παρουσίαση των ποσοτικών δεδομένων του ΣΕΜΚ αναφορικά με το επίπεδο αρχικής κατάταξης, το φύλο και τη χώρα προέλευσης,

³³ Οι οδηγίες για τη Δομή Εξέτασης της Ελληνομάθειας είναι αναρτημένες σε σύνδεσμο στην κεντρική σελίδα του Διδασκαλίου της Νέας Ελληνικής Γλώσσας και συγκεκριμένα στο : http://www.greekcourses.uoa.gr/fileadmin/greekcourses.uoa.gr/uploads/DOCUMENTS_DIAFORA/TE_ST_ELLHNOMATHEIAS.pdf

Πίνακας 7: Κατανομή των δεδομένων του ΣΕΜΚ των 224 ΠΓΛ αναφορικά με το επίπεδο αρχικής κατάταξης, το φύλο και τη χώρα προέλευσης

Επίπεδο αρχικής κατάταξης	Πλήθος ΠΓΛ	Αριθμός Λέξεων	Ποσοστό επί του συνόλου των λέξεων
A (αρχάριοι)	72	17396	32,9%
B (μέσοι)	93	21507	40,6%
Γ (προχωρημένοι)	59	14021	26,5%
Φύλο	Πλήθος ΠΓΛ		
Ανδρες	76	17951	33,9%
Γυναίκες	148	34973	66,1%
Εκτιμώμενη μητρική γλώσσα	Πλήθος ΠΓΛ	Αριθμός Λέξεων	Ποσοστό επί του συνόλου των λέξεων
1.Αραβικά	24	5542	10,5%
2. Αλβανικά	23	5436	10,3%
3.Αρμενικά	4	1031	1,9%
4.Αγγλικά	8	1896	3,6%
5.Γερμανικά	14	3497	6,6%
6.Γαλλικά	7	1551	2,9%
7.Ισπανικά	7	1567	3,0%
8.Βουλγαρικά	22	5118	9,7%
9.Γεωργιανά	11	2745	5,2%
10. Νιγηροκογκολέζικες	7	1723	3,3%
11. Ουκρανικά	24	5493	10,4%
12. Πολωνικά	10	2435	4,6%
13. Ρουμανικά	19	4523	8,5%
14. Ρωσικά	24	5362	10,1%
15. Σερβικά	14	3487	6,6%
16. Τουρκικά	6	1518	2,9%

Όπως γίνεται φανερό, δεν υπάρχει απόλυτη ισοκατανομή του υλικού του ΣΕΜΚ ανά επίπεδο αρχικής κατάταξης, καθώς οι μέσοι αποτελούν την πλειοψηφία (40,6%) με τους αρχαρίους και τους προχωρημένους να ακολουθούν κατά σειρά αναφοράς, χωρίς όμως μεγάλη διαφορά (32,9% και 26,5% αντίστοιχα). Παρόμοια ανισοκατανομή παρατηρείται όσον αφορά το φύλο, καθώς οι γυναίκες αποτελούν τη μεγάλη πλειοψηφία (66,1%). Τέλος, όσον αφορά τις γλωσσικές ομάδες-υποσώματα, καταβλήθηκε προσπάθεια, ώστε όλες να έχουν έκταση τουλάχιστον διπλάσια των 1000 λέξεων, καθώς η τελευταία συνιστά σύμφωνα με τον Μικρό (2007:190-199) την

ελάχιστη απαραίτητη έκταση που εξασφαλίζει σε ένα ΗΣΚ το στοιχειώδες λειτουργικό εύρος. Ωστόσο, οι γλωσσικές ομάδες των Αρμενικών, Τουρκικών, Γαλλικών, Ισπανικών, Νιγηροκογκολέζικων και Αγγλικών δεν πληρούν την παραπάνω συνθήκη. Γι' αυτό, για πρακτικούς λόγους, οι παραπάνω 16 γλωσσικές ομάδες ομαδοποιήθηκαν περαιτέρω, ώστε να προκύψουν οι εξής οκτώ (8) ευρύτερες γλωσσικές ομάδες: α) σλαβικές (βουλγαρικά, ουκρανικά, πολωνικά, ρωσικά, σερβικά), β) ρομανικές (γαλλικά, ισπανικά, ρουμανικά), γ) αλβανικά, δ) γερμανικές (αγγλικά, γερμανικά), ε) αραβικά, στ) αρμενικά και γεωργιανά,³⁴ ζ) νιγηροκογκολέζικες και η) τουρκικά (Τζιμώκας, 2018:92). Τα βασικά ποσοτικά δεδομένα αυτών των ευρύτερων γλωσσικών ομάδων παρατίθενται στον ακόλουθο Πίνακα 8 κατά φθίνουσα σειρά μεγέθους:

Πίνακας 8: Ποσοτικά δεδομένα των ευρύτερων γλωσσικών ομάδων-υποσωμάτων του ΣΕΜΚ

	Γλωσσική Ομάδα	Πλήθος ΠΓΛ	Αριθμός Λέξεων	Ποσοστό επί του συνόλου των λέξεων του ΣΕΜΚ (%)
1.	Σλαβικές	94	21895	41,4 %
2.	Ρομανικές	33	7641	14,4 %
3.	Αλβανικά	23	5436	10,3 %
4.	Γερμανικές	22	5393	10,2 %
5.	Αραβικά	24	5542	10,5 %
6.	Αρμενικά-Γεωργιανά	15	3776	7,1 %
7.	Νιγηροκογκολέζικες	7	1723	3,3 %
8.	Τουρκικά	6	1518	2,9 %

Επομένως, σύμφωνα με τον Πίνακα 8, το υλικό του ΣΕΜΚ προέρχεται κατά συντριπτική πλειοψηφία (41,4%) από μαθητές/-τριες σλαβικής καταγωγής με 21.895 λέξεις και 94 ΠΓΛ, ενώ ακολουθούν με ποσοστό συμμετοχής 14,4%, που είναι μικρότερο του μισού των σλαβόφωνων, οι μαθητές/-τριες με μητρική κάποια ρομανική γλώσσα. Οι επόμενες τρεις γλωσσικές ομάδες, δηλαδή τα αλβανικά, οι γερμανικές γλώσσες και τα αραβικά αντιπροσωπεύονται περίπου ισόποσα στο ΣΕΜΚ τόσο ως προς τον αριθμό των λέξεων όσο και ως προς τον αριθμό των ΠΓΛ, ενώ τα

³⁴ Τα αρμενικά και τα γεωργιανά συνυπολογίστηκαν, αν και δεν είναι γλώσσες με άμεση τυπολογική συγγένεια, λόγω της μεγάλης πιθανότητας οι πληροφορητές από τις χώρες αυτές να ήταν εξοικειωμένοι με τη ρωσική ως Γ2 σε μεγάλο βαθμό.

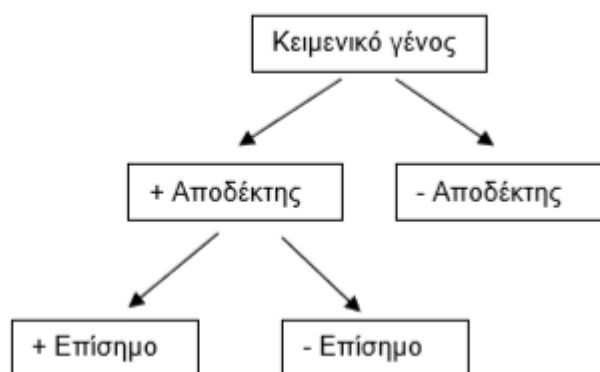
αρμενικά-γεωργιανά και κυρίως οι νιγηροκογκολεζικές γλώσσες και τα τουρκικά καταλαμβάνουν πολύ μικρό μερίδιο του ΣΕΜΚ (7,1%, 3,3% και 2,9 % αντίστοιχα).

Ο τυχαίος χαρακτήρας της δειγματοληψίας των δεδομένων, ο οποίος αντικατοπτρίζει αδρομερώς τα κυρίαρχα δημογραφικά χαρακτηριστικά του πληθυσμού των μαθητών/-τριών του Διδασκαλείου Νέας Ελληνικής Γλώσσας του ΕΚΠΑ κατά τη χρονική περίοδο 2001-2005, συνιστά την κύρια αιτία για τις προαναφερθείσες ανισοκατανομές του υλικού του ΣΕΜΚ στα διάφορα υποσώματά του.

3.1.1.4. Κειμενικά γένη, θέματα ΠΓΛ και αντιπροσωπευτικότητα

Λαμβάνοντας υπόψη τη θεματική των ΠΓΛ του ΣΕΜΚ και τους στόχους της διδασκαλίας της νέας ελληνικής για γενικούς σκοπούς, ήδη από την εποχή του ποιοτικού σχεδιασμού του ΗΣΚ εκμάθησης της Ν.Ε. ως ξένης γλώσσας, στο πλαίσιο του έργου «ΠΥΘΑΓΟΡΑΣ Ι» αποφασίστηκε να υιοθετηθεί μια ad hoc ιεραρχική τυπολογία κατηγοριοποίησης των ΠΓΛ βασισμένη στην έννοια του κειμενικού γένους (genre), η οποία καλύπτει τις βασικές λειτουργίες των κειμένων που θα συλλέγονταν από το αρχείο του Διδασκαλείου της Νέας Ελληνικής Γλώσσας (Μικρός, Μαρκόπουλος, & Ιακώβου, 2004). Στο σχήμα 1 αποδίδεται αυτή η τυπολογία:

Σχήμα 1: Τυπολογία των κειμενικών γενών των ΠΓΛ του ΣΕΜΚ



Η παραπάνω τυπολογία κατηγοριοποίησης επιτρέπει τη διαζευκτική κατηγοριοποίηση οποιουδήποτε κειμένου σε μια από τις ακόλουθες τρεις κατηγορίες: α) [+Αποδέκτης +Επίσημο], β) [+Αποδέκτης -Επίσημο] και γ) [-Αποδέκτης]. Η πρώτη κατηγορία περιλαμβάνει κείμενα που απευθύνονται ρητώς σε συγκεκριμένους/-ες αποδέκτες/-τριες και έχουν επίσημο ύφος, όπως επίσημες επιστολές, ομιλίες σε συνδιασκέψεις ή συνέδρια, ενώ η δεύτερη κατηγορία περιλαμβάνει επίσης κείμενα με συγκεκριμένους/-ες αποδέκτες/-ριες, αλλά χαρακτηριζόμενα από ανεπίσημο ύφος, όπως φιλικές επιστολές. Όσον αφορά την τελευταία κατηγορία, σχετίζεται με κείμενα που δεν απευθύνονται ρητώς σε έναν/μία αποδέκτη/-τρια και είναι κυρίως αφηγηματικής δομής ή κείμενα επιχειρηματολογίας. Το κριτήριο, λοιπόν, που χρησιμοποιεί η εν λόγω τυπολογία για τη διάκριση των κειμενικών γενών είναι τα εξωτερικά χαρακτηριστικά των ΠΓΛ, χαρακτηριστικά, δηλαδή, που αφορούν τη γλωσσική μορφή³⁵. Σύμφωνα με σχετικές μελέτες (McEneaney et al., 2006:21), η συγκεκριμένη σχεδιαστική επιλογή προσιδιάζει καλύτερα στα ΗΣΚ, και ειδικά σε αυτά που σχετίζονται με τη διδασκαλία της νέας ελληνικής ως δεύτερης γλώσσας και την πιστοποίηση ελληνομάθειας, όπως είναι το συγκεκριμένο.³⁶

Στον Πίνακα 9 παρατίθεται η κατανομή του γλωσσικού υλικού του ΣΕΜΚ ανά κειμενικό γένος. Και πάλι αποτυπώνεται η έκταση κάθε υποσώματος σε πλήθος ΠΓΛ και σε αριθμό λέξεων καθώς και το ποσοστό των τελευταίων επί του αριθμού των λέξεων του ΣΕΜΚ συνολικά.

Πίνακας 9: Κατανομή του γλωσσικού υλικού του ΣΕΜΚ ανά κειμενικό γένος

Κειμενικό γένος	Πλήθος ΠΓΛ	Αριθμός Λέξεων	Ποσοστό επί του συνόλου των λέξεων του ΣΕΜΚ
[+ Αποδέκτης + Επίσημο]	69	16290	30,78 %
[+ Αποδέκτης - Επίσημο]	118	27827	52,58 %
[- Αποδέκτης]	37	8807	16,64 %

³⁵ Η έννοια του κειμενικού γένους πρέπει να διακρίνεται από την έννοια του κειμενικού τύπου (text type), η οποία κατηγοριοποιεί τα κείμενα όχι με βάση τα χαρακτηριστικά της γλωσσικής μορφής, αλλά με βάση το θεματικό περιεχόμενό τους (εσωτερικό κριτήριο). 40 Βλ. Φλώρου (2009:155) για τα είδη των γενών που χρησιμοποιούνται συνήθως στη διδασκαλία της ελληνικής ως ξένης και στην πιστοποίηση ελληνομάθειας.

³⁶ Βλ. Φλώρου (2009: 155) για τα είδη των γενών που χρησιμοποιούνται συνήθως στη διδασκαλία της ελληνικής ως ξένης και στην πιστοποίηση ελληνομάθειας.

Σύμφωνα με τα παραπάνω στοιχεία, το κειμενικό γένος με τη μεγαλύτερη αντιπροσώπευση στο ΣΕΜΚ είναι το [+Αποδέκτης -Επίσημο], το γένος, δηλαδή, που περιλαμβάνει αποκλειστικά επιστολές σε ανεπίσημο ύφος που απευθύνονται σε φίλους/-ες, συνομήλικους/-ες ή συγγενικά πρόσωπα. Όπως επισημαίνει η Φλώρου (2009:155), η αυξημένη παρουσία του συγκεκριμένου γένους δικαιολογείται σε μεγάλο βαθμό από το γεγονός ότι ζητείται πολύ συχνά στις εξετάσεις οποιουδήποτε πιστοποιητικού γλωσσομάθειας και, κατ' επέκταση, και στις εξετάσεις για την απόκτηση της βεβαίωσης ελληνομάθειας του Διδασκαλείου Νέας Ελληνικής Γλώσσας, καθώς συνιστά μια πρόσφορη επικοινωνιακή δραστηριότητα με την οποία οι μαθητές/-τριες εξοικειώνονται από πολύ νωρίς. Συγκεκριμένα, από τις συνολικά επτά (7) εξεταστικές περιόδους που αντλήθηκαν δεδομένα για το ΣΕΜΚ στις έξι (6) εντοπίστηκαν θέματα που ανήκουν στο εν λόγω γένος, σε πέντε (5) εξεταστικές περιόδους εντοπίστηκαν τα θέματα που ανήκουν στο [-Αποδέκτης], ενώ τα θέματα που ανήκουν στο [+Αποδέκτης +Επίσημο] αντλήθηκαν από τέσσερις (4) περιόδους. Εκτός, όμως, από τη συχνή εμφάνιση εξεταστικών θεμάτων που ζητούν την παραγωγή φιλικών επιστολών, μία ακόμη αιτία που δικαιολογεί την επίδοση αυτού του κειμενικού γένους είναι η δυνατότητα επιλογής που διαθέτουν οι εξεταζόμενοι/-ες μεταξύ δύο διαφορετικών θεμάτων παραγωγής γραπτού λόγου.

Σύμφωνα με τους Γούτσος (2003:5-6) και McEnery et al. (2006:13-21), η ποικιλία των κειμενικών γενών συμβάλλει από κοινού με την αυθεντικότητα και το μέγεθος στην αντιπροσωπευτικότητα ενός ΗΣΚ. Έτσι, κάθε ΗΣΚ χρειάζεται να περιλαμβάνει όσα περισσότερα κειμενικά είδη ή γένη είναι δυνατόν, προκειμένου το εύρος του να καταστεί ουσιαστικά λειτουργικό, δηλαδή να συμπεριλαμβάνει τόσο τα συχνά όσο και τα σπανιότερα γλωσσικά χαρακτηριστικά. Όπως φαίνεται, λοιπόν, στον Πίνακα 10 που ακολουθεί, στο ΣΕΜΚ συμπεριλαμβάνονται και τα τρία βασικά κειμενικά γένη που χρησιμοποιούνται στη διδασκαλία και στην αξιολόγηση της ελληνομάθειας: [+Αποδέκτης +Επίσημο], [+Αποδέκτης -Επίσημο] και [-Αποδέκτης]. Το στοιχείο, όμως, αυτό δεν αρκεί από μόνο του για να αποδείξει το μέγεθος του λειτουργικού εύρους τόσο του συνόλου του ΣΕΜΚ όσο και των υποσωμάτων του. Με δεδομένο, βέβαια, το μικρό τρέχον μέγεθός του, το ΣΕΜΚ είναι πιο ενδεδειγμένο για την ανάλυση διαγλωσσικών χαρακτηριστικών υψηλής συχνότητας παρά όσων εμφανίζονται σπάνια (Granger, 2003b:541).

Όσον αφορά τα θέματα των ΠΓΛ, στο γένος [+Αποδέκτης +Επίσημο] αντιστοιχούν τέσσερα (3) διαφορετικά θέματα ΠΓΛ, στο [+Αποδέκτης -Επίσημο] οκτώ (8) θέματα και στο [-Αποδέκτης] επτά (6) θέματα. Στον Πίνακα 10 παρουσιάζεται η ποσοτική κατανομή των δεδομένων του ΣΕΜΚ ανά θέμα ΠΓΛ. Και εδώ καταγράφονται σε κάθε περίπτωση το σχετικό πλήθος των ΠΓΛ, ο αριθμός λέξεων και το ποσοστό επί του συνόλου των λέξεων του ΣΕΜΚ στο οποίο αντιστοιχούν οι τελευταίες.

Πίνακας 10: Ποσοτική κατανομή των δεδομένων του ΣΕΜΚ ανά θέμα ΠΓΛ

+Αποδέκτης +Επίσημο			
Θέμα ΠΓΛ	Πλήθος ΠΓΛ	Αριθμός Λέξεων	Ποσοστό επί του συνόλου των λέξεων του ΣΕΜΚ
Επιστολή διαμαρτυρίας για απόφαση δήμου	48	11546	21,82 %
Επιστολή προς τον δήμαρχο για αποτροπή απόλυσης ή εύρεση εναλλακτικής θέσης εργασίας	7	1526	2,9 %
Επιστολή σε εφημερίδα για επιπτώσεις της χρήσης κινητού τηλεφώνου	14	3218	6,1 %
+Αποδέκτης – Επίσημο			
Θέμα ΠΓΛ	Πλήθος ΠΓΛ	Αριθμός Λέξεων	Ποσοστό επί του συνόλου των λέξεων του ΣΕΜΚ
Επιστολή παρουσίασης του εαυτού σου σε ιστοσελίδα γνωριμιών	53	12230	23,1 %
Επιστολή σε φίλο - Περιγραφή μιας συναυλίας	4	1021	1,93 %
Επιστολή σε φίλο/φίλη για να του/της προτείνετε κέντρο για γαμήλια δεξίωση	24	5533	10,45 %
Επιστολή σε φίλο/φίλη με συμβουλές για την προετοιμασία του γιου/της κόρης του/της για σημαντικές εξετάσεις	1	205	0,39 %
Επιστολή σε φίλο/φίλη που σας ζητάει τη γνώμη σας για την αγορά μοτοσικλέτας	30	7133	13,48 %
Επιστολή στους γονείς με περιγραφή του ξεκινήματος της φοιτητικής ζωής	4	883	1,67 %
Επιστολή συμβουλών σε φίλο/-η για το πώς θα βελτιώσει ποιότητα ζωής	1	291	0,55 %
Συμβουλές σε φίλο/φίλη που αντιμετωπίζει προβλήματα υγείας	2	531	1 %

-Αποδέκτης			
Θέμα ΠΓΛ	Πλήθος ΠΓΛ	Αριθμός Λέξεων	Ποσοστό επί του συνόλου των λέξεων του ΣΕΜΚ
Γράψτε μια ιστορία που ξεκινάει ως εξής: "Δεν θα ξεχάσω ποτέ την ημέρα που..."	6	1492	2,82 %
Η αξία του ελεύθερου χρόνου στη ζωή μας	2	516	0,97 %
Η σημασία της μόδας στη ζωή μας	1	194	0,37 %
Ο ρόλος του αθλητισμού στη ζωή μας	12	2887	5,45 %
Περιγράψτε τον δικό σας ήρωα	14	3494	6,6 %
Πιστεύετε στην τύχη ή όχι;	1	224	0,42 %

3.1.1.5. Τύπος βάσης δεδομένων του ΣΕΜΚ

Το ΣΕΜΚ που αξιοποιείται στην παρούσα εργασία αποτελείται από 224 αρχεία «απλού κειμένου» τύπου .txt, που περιέχουν τις ηλεκτρολογημένες μαθητικές ΠΓΛ, οι οποίες αντλήθηκαν από τα εξεταστικά τετράδια με τις χειρόγραφες ΠΓΛ. Δεδομένου ότι η συγκεκριμένη εργασία στηρίχθηκε στην έρευνα που έγινε στο πλαίσιο του προγράμματος «Πυθαγόρας Ι» και παρουσιάζεται αναλυτικά στη διδακτορική διατριβή του Δημήτρη Τζιμόκα «Γραμματικά λάθη των μαθητών της νέας ελληνικής ως δεύτερης γλώσσας και η κωδικοποίησή τους βάσει ηλεκτρονικών σωμάτων κειμένων» (2018), αξίζει να αναφερθεί ότι κάθε ένα από αυτά τα αρχεία «απλού κειμένου» έχει ονομαστεί από τον/την εκάστοτε διορθωτή/-τρια με ενιαίο σε γενικές γραμμές τρόπο (συνδυασμός τριών γραμμάτων και τριών ψηφίων) και συνοδεύεται από άλλα τέσσερα συμπληρωματικά είδη αρχείων με μεταδεδομένα, τα οποία έχουν το ίδιο όνομα με το αρχικό αρχείο «απλού κειμένου», αλλά είναι διαφορετικού τύπου και περιεχομένου και εστιάζουν στη διερεύνηση των λαθών. Ωστόσο, στην παρούσα εργασία, δε συμπεριλαμβάνονται τα τέσσερα αυτά αρχεία, καθώς δε θα μας απασχολήσουν τα λάθη. Στη δεδομένη περίπτωση, στόχος είναι να διερευνηθούν έξι συγκεκριμένοι δείκτες λεξιλογικού «πλούτου». Γι' αυτόν τον λόγο, παρατίθενται, πέρα από τα αρχεία «απλού κειμένου», πέντε ακόμα αρχεία, τα οποία συνέβαλαν στην ταξινόμηση των αποτελεσμάτων που παρήχθησαν από το QUITA, ένα πρόγραμμα ποσοτικής ανάλυσης κειμένου, ιδιαίτερα χρήσιμο και δημοφιλές στις

ανθρωπιστικές επιστήμες. Συγκεκριμένα, παρατίθεται ένα αρχείο τύπου excel το οποίο αποτελείται από τρία φύλλα, ένα για κάθε επίπεδο γλωσσομάθειας σύμφωνα με το κατατακτήριο τεστ των μαθητών/-τριών. Το συγκεκριμένο αρχείο αποτυπώνει τους έξι παράγοντες λεξιλογικού «πλούτου» ανά επίπεδο, τις λέξεις και τους λεξικούς τύπους καθώς και τους μέσους όρους για κάθε παράγοντα, οι οποίοι κρίθηκαν αναγκαίοι για την εξαγωγή συμπερασμάτων αναφορικά με τους συγκεκριμένους δείκτες σε κάθε επίπεδο. Το δεύτερο αρχείο συνιστά ένα υπολογιστικό φύλλο του στατιστικού προγράμματος SPSS, στο οποίο αποτυπώνονται οι τιμές των έξι παραγόντων λεξιλογικού «πλούτου» των ΠΓΛ όλων των μαθητών/-τριών και τα αποτελέσματα που προκύπτουν από τη σύγκριση των δεικτών στα τρία επίπεδα μέσω του παραμετρικού στατιστικού κριτήριο one way ANOVA. Τα τελευταία τρία αρχεία συνιστούν, επίσης, υπολογιστικά φύλλα του προγράμματος SPSS. Καθένα από αυτά αντιστοιχεί σε ένα επίπεδο γλωσσομάθειας, ενώ και στα τρία καταγράφονται οι τιμές των έξι παραγόντων λεξιλογικού «πλούτου» για το εκάστοτε επίπεδο. Στα αρχεία αυτά γίνεται χρήση της προβλεπτικής εξίσωσης πολλαπλής παλινδρόμησης (Multiple Regression Analysis- MRA) προκειμένου να διαπιστωθεί η συσχέτιση μεταξύ των υπό εξέταση δεικτών με τη βαθμολόγηση των μαθητών/-τριών, να εξεταστεί, δηλαδή, αν οι δείκτες που κρίθηκαν στατιστικά σημαντικοί λήφθηκαν υπόψη κατά τη βαθμολόγηση του λεξιλογίου.

Επομένως, στο πλαίσιο της παρούσας έρευνας από το σύνολο των προαναφερθέντων 317 αρχείων αξιοποιήθηκαν τα 224 αρχεία τύπου .txt, ένα αρχείο excel και τα υπολογιστικά φύλλα του στατιστικού προγράμματος SPSS. Αξίζει να επισημανθεί ότι δεν χρησιμοποιήθηκε κάποιο λογισμικό αυτόματης δημιουργίας υποσωμάτων του ΣΕΜΚ με βάση τις διάφορες μεταβλητές των πληροφορητών και των ΠΓΛ (φύλο, χώρα προέλευσης, κειμενικό γένος, θέμα), αλλά όλες οι απαραίτητες ομαδοποιήσεις έγιναν χειροκίνητα προκειμένου να διενεργηθούν οι ανάλογες μετρήσεις.

3.1.1.6. Πλεονεκτήματα και μειονεκτήματα του ΣΕΜΚ

Σύμφωνα με τον Τζιμώκα (2018:99-100), είναι εμφανές ότι το ΣΕΜΚ χαρακτηρίζεται από συγκεκριμένα πλεονεκτήματα και μειονεκτήματα, όπως άλλωστε και κάθε ΗΣΚ, τα οποία είναι ανάγκη να λαμβάνονται σοβαρά υπόψη κατά την ερευνητική του χρήση. Αφενός, τα πλεονεκτήματά του είναι ως επί το πλείστον συμβατά με τις

ελάχιστες προδιαγραφές ενός ερευνητικά αξιοποιήσιμου ΗΣΚ (Leech, 1997a:3· Granger et al., 2002: 8-10· Cownan et al., 2003: 453· McEneary et al., 2006:253) και συνοψίζονται στα εξής: α) αυθεντικότητα του υλικού, υπό την έννοια ότι το σύνολο των ΠΓΛ του ΣΕΜΚ ήταν αποτέλεσμα αυθεντικής γλωσσικής δραστηριότητας στο πλαίσιο της τελικής γραπτής αξιολόγησης των μαθητών μετά από παρακολούθηση συγκεκριμένου προγράμματος γλωσσικών μαθημάτων, β) πιστή καταγραφή των ΠΓΛ των μαθητών με όλες τις ορθογραφικές ιδιαιτερότητές τους σε ηλεκτρονική μορφή, γ) αναλυτική καταγραφή των παραγόντων λεξιλογικού «πλούτου» όλων των γλωσσικών επιπέδων, άρα με δυνατότητα επαναχρησιμοποίησης των δεδομένων για μελλοντικές έρευνες, δ) στοχευόμενο επίπεδο ελληνομάθειας των ΠΓΛ το B2, το οποίο κατατάσσει τους πληροφορητές στους προχωρημένους μαθητές, ε) σταθερή έκταση των ΠΓΛ σε γενικές γραμμές και στ) κειμενικά γένη με τη συχνότερη χρήση στη διδασκαλία της νέας ελληνικής και στην πιστοποίηση ελληνομάθειας.

Όμως, σύμφωνα πάλι με τον Τζιμόκα (2018:100), υπάρχουν και μειονεκτήματα στο ΣΕΜΚ, τα κυριότερα από τα οποία είναι: α) η έμμεση συναγωγή της μητρικής γλώσσας κάθε πληροφορητή από τη δηλούμενη χώρα προέλευσης, β) το σχετικά μικρό μέγεθος, γ) ανομοιογένεια ως προς τη χώρα προέλευσης του 1 από τα 16 υποσώματα, συγκεκριμένα της γλωσσικής ομάδας «Νιγηροκογκολεζικές γλώσσες», η οποία περιλαμβάνει μαθητές/-τριες από έξι (6) διαφορετικές χώρες προέλευσης, δ) μη διαθεσιμότητα του ΣΕΜΚ, τουλάχιστον προς το παρόν, για ευρεία ερευνητική χρήση.

Όσον αφορά, βέβαια, το μικρό μέγεθος, η αρνητική επίδρασή του στα ερευνητικά δεδομένα μετριάζεται από τον γενικό χαρακτήρα των στόχων της παρούσας έρευνας, που αποβλέπουν σε μια σχετικά αδρομερή και όχι εις βάθος ανάλυση κάποιων δεικτών λεξιλογικού «πλούτου» που συνάδουν με τις βασικές προβλέψεις του ΑΠ-B2. Επιπλέον, όπως επισημαίνει η Prat-Zagrebelsky (2004:71), τα ΗΣΚ μικρού σχετικά μεγέθους ευνοούν τη συστηματική ανάγνωση και χειρωνακτική ανάλυση των δεδομένων τους, διαδικασίες δηλαδή που συνιστούν απαραίτητες προϋποθέσεις για τη γέννηση χρήσιμων ερευνητικών ερωτημάτων.

3.1.2. Υπολογιστικά εργαλεία επεξεργασίας του ΣΕΜΚ

Σύμφωνα με τους McEneary & Rayson (1997:194-195), υπάρχουν τρεις βασικές κατηγορίες υπολογιστικών εργαλείων για την ανάπτυξη και τον χαρακτηρισμό ΗΣΚ:

α) εργαλεία για την ανάπτυξη και την επισημείωση ΗΣΚ (κωδικοποίηση κειμένου, επισημείωση, κωδικοποίηση επισημείωσης), β) εργαλεία επεξεργασίας ΗΣΚ (διόρθωση, αποσαφήνιση των επισημειώσεων, μετατροπή των επισημειώσεων) και γ) εργαλεία για άντληση πληροφοριών (σχηματισμός συμπραστικών πινάκων, ανάλυση συχνοτήτων, εργαλεία εισαγωγής στοιχείων σε λεξικά και γραμματικές, εργαλεία ανάκλησης πληροφοριών, παραλλαγές των προηγούμενων εργαλείων σε δίγλωσση ή πολύγλωσση έκδοση).

Αναφορικά με την πρώτη κατηγορία εργαλείων, για την ανάπτυξη του ΣΕΜΚ χρησιμοποιήθηκε το λογισμικό «Σημειωματάριο» του λειτουργικού συστήματος Windows για την πληκτρολόγηση των ΠΓΛ των μαθητών/-τριών και την αποθήκευσή τους σε αρχεία τύπου «απλού κειμένου» (.txt). Τα αρχεία αυτά τα οποία αξιοποιήθηκαν για την εξαγωγή συμπερασμάτων σχετικών με τους δείκτες λεξιλογικού «πλούτου» δε συντάχθηκαν στο πλαίσιο της παρούσας εργασίας, αλλά από την ερευνητική ομάδα που συμμετείχε στο Προγράμματο «Πυθαγόρας Ι» και παρατίθενται από τον Τζιμόκα (2010). Από τη δεύτερη κατηγορία εργαλείων για την επεξεργασία και την ανάλυση του ΣΕΜΚ δεν χρειάστηκε να χρησιμοποιηθεί κάποιο εξειδικευμένο λογισμικό. Τέλος, από την τρίτη κατηγορία, για την ποσοτικοποίηση των ΠΓΛ και την απόδοση των δεικτών λεξιλογικού «πλούτου» σε αριθμούς με σκοπό την περαιτέρω επεξεργασία τους, αξιοποιήθηκε το πρόγραμμα QUITA³⁷ (Quantitative Indicator Text Analyzer). Πρόκειται για ένα εργαλείο ιδιαίτερα χρήσιμο για έρευνες στην ποσοτική γλωσσολογία με ενσωματωμένους κάποιους βασικούς δείκτες λεξιλογίου. Για την ποσοτική και ποιοτική ανάλυση των δεδομένων, όπως αυτά αποτυπώθηκαν στο QUITA, χρησιμοποιήθηκε το λογισμικό Microsoft Excel 2010. Μέσω αυτού δημιουργήθηκε ένα τρίπτυχο λογιστικό φύλλο με τους δείκτες λεξιλογικού «πλούτου» ανά επίπεδο και την εξαγωγή των μέσων όρων ανά δείκτη σε κάθε ένα από τα τρία επίπεδα. Πέρα από αυτά τα δύο λογισμικά, χρησιμοποιήθηκε το στατιστικό πρόγραμμα SPSS. Σε ένα από τα υπολογιστικά φύλλα του SPSS αποτυπώνονται οι τιμές για τους έξι παράγοντες λεξιλογικού «πλούτου» που προκύπτουν από τις ΠΓΛ όλων των μαθητών/-τριών. Για τη σύγκριση των τριών επιπέδων, προκειμένου να διαπιστωθεί ποιοι δείκτες σε ποια επίπεδα εμφανίζουν στατιστικά σημαντική διαφορά, αξιοποιήθηκε το παραμετρικό στατιστικό κριτήριο one way ANOVA (Analysis of Variance – ANOVA). Τέλος, αξιοποιήθηκαν και άλλα

³⁷ https://www.dropbox.com/s/i1rp5pga5fycqhx/QUITA_Setup_1_11_0.msi?dl=0

τρία υπολογιστικά φύλλα του SPSS, ένα για κάθε επίπεδο γλωσσομάθειας. Και στα τρία καταγράφονται οι τιμές των έξι παραγόντων λεξιλογικού «πλούτου» για το εκάστοτε επίπεδο. Τα αρχεία αυτά αποτελούν προέκταση του πρώτου αρχείου SPSS, αφού μέσω της προβλεπτικής εξίσωσης πολλαπλής παλινδρόμησης (Multiple Regression Analysis- MRA), επιδιώκεται να εξεταστεί αν οι δείκτες που κρίθηκαν στατιστικά σημαντικοί λήφθηκαν υπόψη κατά τη βαθμολόγηση του λεξιλογίου.

3.1.2.1. Πρόγραμμα ποσοτικής ανάλυσης κειμένου (QUITA)

Τα χαρακτηριστικά, τα οποία κατά καιρούς έχουν μετρηθεί ως υφομετρικοί δείκτες από το ξεκίνημα της υφομετρίας μέχρι σήμερα, είναι πολυάριθμα³⁸ και ανήκουν σε όλο το εύρος των γλωσσικών επιπέδων. Ανάλογα με το γλωσσικό επίπεδο στο οποίο ανήκει το κάθε γλωσσικό χαρακτηριστικό, κατατάσσεται στις εξής κατηγορίες: α) Φωνολογικό/Γραφηματικό, β) Μορφολογικό/Λεξιλογικό, γ) Συντακτικό, δ) Σημασιολογικό.³⁹ Οι έξι υπό εξέταση λεξιλογικοί δείκτες, δηλαδή το Harax Legomenon Percentage (HL) ή Ποσοστό Άπαξ Λεγομένων, το R1, το Lambda (Λ), το h-point (h), η Entropy (H) ή Εντροπία και το Average Tokens length (ATL) ή Μέσο Μήκος Λεξικού Τύπου κατατάσσονται στο λεξιλογικό επίπεδο.

Όπως προαναφέρθηκε, το πρόγραμμα που αξιοποιήθηκε για την ποσοτική ανάλυση των ΠΓΛ των μαθητών είναι το QUITA (Quantitative Indicator Text Analyzer). Πρόκειται για ένα εργαλείο ποσοτικοποίησης των ΠΓΛ, το οποίο αξιοποιείται για έρευνες στην ποσοτική γλωσσολογία. Από τους ενσωματωμένους δείκτες λεξιλογίου που εμπεριέχει επιλέχθηκαν έξι προαναφερθέντες δείκτες λεξιλογικού «πλούτου».

Όπως αναφέρεται στο εγχειρίδιο του προγράμματος, οι λεξιλογικοί δείκτες που χρησιμοποιούνται στο QUITA, συμπεριλαμβανομένων και των άνωθεν, επηρεάζονται από το μήκος του κειμένου, με εξαίρεση τον δείκτη LAMBDA (Λ), καθώς είναι ο μοναδικός που παραμένει ανεπηρέαστος από το κειμενικό μέγεθος (Popescu et al., 2011:1). Ως εκ τούτου, τα αποτελέσματα που προκύπτουν από την επεξεργασία των ΠΓΛ εκφράζουν το μέγεθος ενός δείκτη λαμβάνοντας υπόψη το μέγεθος του κειμένου. Αναπόφευκτα, λοιπόν, θεωρούνται συγκρίσιμα μόνο τα κείμενα με μικρές αποκλίσεις στο μέγεθος. Αυτός ήταν και ο λόγος που αποκλείστηκαν ορισμένες ΠΓΛ

³⁸ Ο Rudman (1997, p. 360)

³⁹ Ο Μικρός (2015) προτείνει μια ακόμα κατηγορία, την Εξωγλωσσική, η οποία συγκεντρώνει χαρακτηριστικά που δεν είναι γλωσσικά (π.χ. μεταδεδομένα κειμένων, μορφοποιήσεις και μη γλωσσική οργάνωση κειμενικών δομών κ.ά.).

από το αρχικό ΣΕΜΚ του Τζιμόκα (2010) και από τις 317 ΠΓΛ τελικά αξιοποιήθηκαν μόνο οι 224. Αφαιρέθηκαν κατά κύριο λόγο οι ΠΓΛ που απέκλειαν σημαντικά από την προβλεπόμενο όριο των 220 - 250 λέξεων και το εύρος τους περιορίστηκε στις 180-300. Τελικά, αξιοποιήθηκαν 224 ΠΓΛ αντί για τις 233 που εντάσσονταν στο συγκεκριμένο όριο λέξεων, λόγω ελλιπών πληροφοριών αναφορικά με την βαθμολογία του λεξιλογίου, γεγονός που τις καθιστούσε μη αξιοποιήσιμες. Παρά το γεγονός ότι στις ΠΓΛ των μαθητών/-τριών δεν τηρείται απόλυτα το όριο των λέξεων, η απόκλιση των ακραίων τιμών επιτρέπει την εξαγωγή αξιόπιστων συμπερασμάτων από τη μεταξύ τους σύγκριση. Οι δείκτες που επιλέγονται προς εξέταση είναι στην ευχέρεια του ερευνητή/-τριας και αφορούν το εκάστοτε ερευνητικό ερώτημα. Ως εκ τούτου κανένας δείκτης δεν είναι αξιολογικά ανώτερος από κανένα, αλλά εξαρτάται από τον στόχο που εξυπηρετεί.

3.1.2.2. Στατιστικό πρόγραμμα SPSS

Δεδομένου ότι κεντρικός στόχος της παρούσας εργασίας είναι να διαπιστωθούν τυχόν διαφορές στις ΠΓΛ των μαθητών/-τριών αναφορικά με τους δείκτες λεξιλογικού «πλούτου» ανά επίπεδο γλωσσομάθειας σύμφωνα με την αρχική κατάταξή τους, κρίθηκε απαραίτητη η σύγκριση των μέσων όρων των δεικτών αυτών ανά επίπεδο. Αυτό κατέστη δυνατό μέσω του παραμετρικού στατιστικού κριτηρίου One Way ANOVA ή αλλιώς Ανάλυση Διακύμανσης (Analysis of Variance – ANOVA). Εφόσον κάποιοι δείκτες κριθούν στατιστικά σημαντικοί, προβλέπεται να εξετασθεί αν λήφθηκαν υπόψη στη βαθμολόγηση του λεξιλογίου στις ΠΓΛ των μαθητών/-τριών. Για να καταστεί αυτό δυνατόν, προβλέπεται να αξιοποιηθεί η προβλεπτική εξίσωση πολλαπλής παλινδρόμησης (Multiple Regression Analysis-MRA).

3.1.2.2.1. Παραμετρικό στατιστικό κριτήριο One Way ANOVA (Analysis of Variance – ANOVA)

Η ανάλυση διακύμανσης μονής κατεύθυνσης (One-Way ANalysis Of VAriance) αναφέρεται σε διαφορές μεταξύ μέσων όρων και επιτρέπει στον/στην ερευνητή/-τρια να συγκρίνει μέσους όρους από περισσότερα από δύο δείγματα. Στη δεδομένη περίπτωση συνεξετάστηκε ο παράγοντας επίπεδο με τους δείκτες λεξιλογικού

«πλούτου». Η συγκεκριμένη στατιστική μέθοδος μπορεί να χρησιμοποιηθεί μόνο όταν τηρούνται ορισμένες προϋποθέσεις: α) το δείγμα να είναι αντιπροσωπευτικό και οι τιμές που το απαρτίζουν οφείλονται σε ανεξάρτητες παρατηρήσεις, β) τα δεδομένα να πληρούν τις προϋποθέσεις χρήσης παραμετρικών κριτηρίων. Στο πλαίσιο της συγκεκριμένης μεθόδου, πραγματοποιείται έλεγχος υποθέσεων με στόχο να ανιχνευθούν, εάν υπάρχουν, διαφορές στις μέσες τιμές περισσότερων από δύο πληθυσμών.

Για να δοθεί απάντηση στο συγκεκριμένο ερώτημα, κατασκευάζουμε έναν έλεγχο υποθέσεων με μηδενική υπόθεση H_0 , σύμφωνα με την οποία όλα τα δείγματα προέρχονται από πληθυσμούς με την ίδια μέση τιμή, έναντι μιας εναλλακτικής υπόθεσης (H_1), η οποία υπαγορεύει ότι τουλάχιστον δύο μέσες τιμές είναι διαφορετικές. Εάν συγκρίνουμε τρεις ή περισσότερες ομάδες ατόμων αξιοποιώντας το one way ANOVA και το αποτέλεσμα μας οδηγεί σε απόρριψη της H_0 , έχουμε καταλήξει ότι τουλάχιστον ένας μέσος όρος διαφέρει στατιστικά σημαντικά από τουλάχιστον έναν άλλο μέσο όρο, αλλά δεν έχουμε καμία πληροφορία για το πού ακριβώς εστιάζεται η επίδραση της ανεξάρτητης μεταβλητής. Για να μπορέσουμε να απαντήσουμε πού εστιάζεται η επίδραση της ανεξάρτητης μεταβλητής, είναι απαραίτητο να ελέγξουμε τη σημαντικότητα των διαφορών μεταξύ όλων των ζευγών μέσων όρων αξιοποιώντας άλλα κριτήρια, a priori (αποφασίζονται από τον ερευνητή πριν την πραγματοποίηση της έρευνας) ή σε post hoc (αποφασίζονται μετά την ολοκλήρωση της έρευνας και την πραγματοποίηση της ανάλυσης διακύμανσης). Στη δεδομένη περίπτωση, όπως θα φανεί στην Ενότητα 4, θα αξιοποιηθούν κριτήρια post hoc.

3.1.2.2.2. Προβλεπτική εξίσωση πολλαπλής παλινδρόμησης (Multiple Regression Analysis - MRA).

Στην απλή παλινδρόμηση εκτιμούμε τη σχέση μίας μεταβλητής (X) με μία άλλη (Y), εκφράζοντας αυτή τη σχέση ως γραμμικό αποτέλεσμα της πρώτης (X) επί της δεύτερης (Y). Στην παλινδρόμηση οι μεταβλητές πρέπει να είναι συνεχείς.⁴⁰ Η παλινδρόμηση εκφράζει την εξάρτηση της μεταβλητής Y από την ανεξάρτητη X με

⁴⁰ Τακτικές μεταβλητές θεωρούνται αποδεκτές και τις μεταχειριζόμαστε ως ποσοτικές αν και υπάρχει ισχυρή αμφισβήτηση ως προς την εγκυρότητα των αποτελεσμάτων.

μορφή ευθείας γραμμής με τη χρήση της μαθηματικής εξίσωσης, $Y = a + bX$. Η σχέση αυτή μεταξύ των δύο μεταβλητών είναι ουσιαστικά εξάρτηση της πρώτης από τη δεύτερη, δηλαδή το μέγεθος της μεταβολής της πρώτης προσδιορίζεται ως το αποτέλεσμα της δράσης της δεύτερης, χωρίς να ισχύει το αντίθετο. Η πρώτη μεταβλητή καλείται εξαρτημένη (dependent) ή μεταβλητή απόκρισης (response) ή προβλέψιμη (predicted) και συμβολίζεται με Y και η δεύτερη καλείται ανεξάρτητη (independent) ή προβλεπτική (predictive) ή επεξηγηματική (explanatory) και συμβολίζεται με X . Η μεταβλητή X εκλαμβάνεται ως ανεξάρτητη, καθόσον ελέγχεται με μετρήσεις που διεξάγει ο ερευνητής/-τρια, το αποτέλεσμα των οποίων αναμένεται να διαπιστωθεί επί της μεταβλητής Y , της οποίας οι τιμές εξαρτώνται άμεσα από τις τιμές της X . Τέτοια εξαρτημένη σχέση καλείται παλινδρόμηση και πιο συγκεκριμένα απλή παλινδρόμηση, όταν εμπλέκονται δύο μόνο μεταβλητές.

Η συμμετοχή περισσότερων των δύο μεταβλητών σε μία ανάλυση παλινδρόμησης ή και συσχέτισης συνιστά την πολλαπλή παλινδρόμηση (Multiple Regression Analysis-MRA) ή και πολλαπλή συσχέτιση (multiple correlation), η οποία αξιοποιείται και στην παρούσα εργασία. Αν πραγματοποιήσουμε μετρήσεις ταυτόχρονα για τρεις ή περισσότερες μεταβλητές, από τις οποίες η μία θεωρούμε ότι είναι εξαρτημένη (Y) από τη δράση των λοιπών (X_i), π.χ. τις X_1 , X_2 και X_3 , τότε αναφερόμαστε στην πολλαπλή παλινδρόμηση. Στη συγκεκριμένη έρευνα, η πολλαπλή παλινδρόμηση αξιοποιείται για να διαπιστωθεί αν στη βαθμολόγηση της ΠΓΛ των μαθητών/-τριών (εξαρτημένη μεταβλητή) λήφθηκαν υπόψη οι προς διερεύνηση λεξιλογικοί δείκτες (ανεξάρτητες μεταβλητές). Διενεργήθηκε τρεις φορές εξετάζοντας τους δείκτες λεξιλογικού πλούτου και τις βαθμολογίες για κάθε επίπεδο γλωσσομάθειας ξεχωριστά, προκειμένου να διαπιστωθεί αν υπάρχει και σε ποιο/ποια επίπεδα ισχυρή συσχέτιση μεταξύ εξαρτημένης και ανεξάρτητων μεταβλητών.

Η πολλαπλή παλινδρόμηση έχει ευρεία επιστημονική αποδοχή διότι θεωρείται ισχυρό και ευέλικτο στατιστικό εργαλείο με πλήθος εφαρμογές σε τελείως διαφορετικά ερευνητικά πεδία (Draper & Smith, 1989· Pedhazur, 1997 · Weisburg, 1985).

3.2. Μέθοδοι διερεύνησης των ερευνητικών ερωτημάτων και υποθέσεων

Στην ενότητα αυτή θα εκτεθούν οι βασικές μέθοδοι διερεύνησης των ερωτημάτων της παρούσας έρευνας, έτσι όπως αυτά παρουσιάστηκαν συνοπτικά στην ενότητα 1.2.

3.2.1. Ερώτημα 1^ο : Οι στατιστικά σημαντικοί δείκτες λεξιλογικού πλούτου ανά επίπεδο γλωσσομάθειας.

Προκειμένου να απαντηθεί το πρώτο ερώτημα σχετικά με την πιθανή συσχέτιση των έξι δεικτών λεξιλογικού «πλούτου» με τα επίπεδα γλωσσομάθειας στα οποία κατατάχτηκαν οι εξεταζόμενοι/-ες προτού παρακολουθήσουν μαθήματα ελληνικών, απαιτείται σύγκριση αυτών των δεικτών ανά επίπεδο, ώστε να διαπιστωθεί πιθανή απόκλιση μεταξύ τους. Η επιβεβαίωση της απόκλισης των/κάποιων δεικτών θα πιστοποιούσε ότι ο/οι συγκεκριμένος/-οι δείκτης/-ες σχετίζεται/-ονται άμεσα με τα επίπεδα γλωσσομάθειας από τα οποία προέρχονταν οι μαθητές/-τριες.

Για να διαπιστωθεί αυτό, οι ΠΓΛ ποσοτικοποιήθηκαν μέσω του προγράμματος QUITA, με αποτέλεσμα να υπάρχει μία αριθμητική τιμή για κάθε δείκτη αναφορικά με κάθε ΠΓΛ. Τα αριθμητικά δεδομένα αποτυπώθηκαν σε υπολογιστικό φύλλο του λογισμικού Microsoft Excel 2010, μέσω του οποίου υπολογίστηκε ο μέσος όρος κάθε δείκτη ανά επίπεδο. Για τη σύγκριση των μέσων όρων κάθε επιπέδου γλωσσομάθειας χρειάστηκε η μετάβαση στο στατιστικό πρόγραμμα SPSS και η χρήση του παραμετρικού στατιστικού κριτηρίου one way ANOVA. Τα στατιστικώς σημαντικά αποτελέσματα τόσο της εν λόγω δοκιμασίας όσο και των άλλων παρόμοιων στατιστικών δοκιμασιών που έγιναν στο πλαίσιο της παρούσας έρευνας (βλ. Κεφάλαιο 4) διαβαθμίζονται ανάλογα με το επίπεδο σημαντικότητάς τους (τιμή p) ως εξής (McEnery et al., 2006:55):

$p < 0,05$: «σημαντική διαφορά» ή «διαφέρει σε σημαντικό βαθμό»

$p < 0,01$: «πολύ σημαντική διαφορά» ή «διαφέρει σε πολύ σημαντικό βαθμό»

$p < 0,001$: «εξαιρετικά σημαντική διαφορά» ή «διαφέρει σε εξαιρετικά σημαντικό βαθμό»

Κατά τη συζήτηση των αποτελεσμάτων και την εξαγωγή των σχετικών συμπερασμάτων λήφθηκαν υπόψη τα αποτελέσματα με επίπεδο σημαντικότητας $p < 0,05$ (βλ. Κεφάλαιο 4).

Από τη στιγμή που κάποιοι δείκτες, μέσω του παραμετρικού τεστ, αποδειχθούν στατιστικά σημαντικοί, τίθεται ζήτημα επανεκτίμησης των προαπαιτούμενων για να καταταχθεί ένας μαθητής στο επίπεδο B2 σύμφωνα με το ΑΠ. Σημαντικό είναι, επίσης, στην περίπτωση αυτή, να διαπιστωθεί αν οι δείκτες που κρίθηκαν στατιστικά

σημαντικοί λήφθηκαν υπόψη στη βαθμολόγηση του λεξιλογίου από τους/τις βαθμολογητές/-τριες. Με αυτόν τον τρόπο, επιχειρήθηκε μία γενική χαρτογράφηση των υπό έρευνα δεικτών στη διαγλώσσας των πληροφορητών/-τριών του ΣΕΜΚ και δόθηκαν πολύ χρήσιμα στοιχεία για τη γλωσσική χρήση τους.

3.2.2. Ερώτημα 2^ο : Δείκτες λεξιλογικού «πλούτου» ισχύον πλαίσιο αξιολόγησης του μαθητικού κοινού

Τα κείμενα των μαθητών/-τριών που συμπεριλήφθηκαν στην έρευνα του Προγράμματος «Πυθαγόρας Ι» και έγιναν αντικείμενο εξέτασης στην παρούσα εργασία κατατάσσονται στο Β επίπεδο Γλωσσομάθειας, παρουσιάζοντας, εντούτοις, σημαντικές αποκλίσεις στο ποσοστό επιτυχίας. Εφόσον οι/κάποιοι δείκτες λεξιλογικού «πλούτου» εμφανίσουν στατιστικά σημαντική διαφορά ανά επίπεδο γλωσσομάθειας, γεγονός που σχετίζεται με την αρχική τους κατάταξη, θα ήταν χρήσιμο να εξεταστεί το ενδεχόμενο αναθεώρησης του ισχύοντος πλαισίου αξιολόγησης, προκειμένου η βαθμολόγηση του λεξιλογίου να είναι πιο έγκριτη και αντιπροσωπευτική των δυνατοτήτων των μαθητών/-τριών σε σχέση με το γνωστικό τους επίπεδο.

ΚΕΦΑΛΑΙΟ 4: ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΡΕΥΝΑΣ

Στο κεφάλαιο αυτό παρουσιάζονται τα ποσοτικά αποτελέσματα των μετρήσεων που διενεργήθηκαν για τον εντοπισμό των υπό εξέταση δεικτών σε κάθε επίπεδο γλωσσομάθειας. Πέρα από αυτό, παρατίθενται ποσοτικά αποτελέσματα που απαντούν στο κατά πόσο οι συγκεκριμένοι δείκτες λεξιλογικού «πλούτου» λήφθηκαν υπόψη στη βαθμολόγηση των ΠΓΛ των αλλόγλωσσων μαθητών/-τριών. Ύστερα από την παράθεση των πορισμάτων, ακολουθεί μία σύντομη ερμηνευτική ανάγνωσή τους προκειμένου να δοθεί απάντηση στο πρώτο ερευνητικό ερώτημα αναφορικά με τη σχέση των έξι δεικτών λεξιλογικού «πλούτου» με τα επίπεδα γλωσσομάθειας αλλά και στο δεύτερο αναφορικά με την επίδραση που ασκούν στη βαθμολόγηση των μαθητών/-τριών. Συγκεκριμένα, στην ενότητα 4.1. παρουσιάζονται τα ποσοτικά αποτελέσματα, όπως αυτά εξάγονται από τη χρήση του παραμετρικού στατιστικού κριτηρίου one way ANOVA (4.1.1.), ενώ επιχειρείται και μία ποιοτική ανάλυση των αποτελεσμάτων (4.1.2.), προκειμένου να εντοπιστούν οι στατιστικά σημαντικοί δείκτες λεξιλογικού «πλούτου» και τα επίπεδα γλωσσομάθειας στα οποία διαφοροποιούνται. Στην ενότητα 4.2. που ακολουθεί, παρουσιάζονται τα ποσοτικά αποτελέσματα από τη χρήση της προβλεπτικής εξίσωσης πολλαπλής παλινδρόμησης (Multiple Regression Analysis- MRA) (4.2.1.) αλλά και μία ερμηνευτική προσέγγιση αυτών (4.2.2.), ώστε να διαπιστωθεί η συσχέτιση μεταξύ των υπό εξέταση δεικτών με τη βαθμολόγηση του μαθητικού κοινού.

4.1. Παραμετρικό στατιστικό κριτήριο one way ANOVA

Η κατάταξη όσων μαθητών/-τριών υπερέβησαν την βάση του 60% στις εξετάσεις του B2 επιπέδου γλωσσομάθειας, πυροδότησε τη διερεύνηση του υπόλοιπου 40% και τη συσχέτιση των ΠΓΛ των μαθητών/-τριών με το πρότερο επίπεδο γλωσσομάθειας. Η συσχέτιση αυτή, όπως έχει επισημανθεί, επιχειρήθηκε, αρχικά, με τη μέτρηση κάποιων δεικτών λεξιλογικού «πλούτου». Συγκεκριμένα, επιλέχθηκαν οι έξι ακόλουθοι: Ποσοστό Άπαξ Λεγομένων ή Harax Legomenon Percentage (HL), R1, Lambda (Λ), h-point (h), Εντροπία ή Entropy (H), Μέσο Μήκος Λεξικού Τύπου ή

Average Tokens length (ATL). Οι δείκτες αυτοί εντοπίζονται στις 224 ΠΓΛ των μαθητών, των οποίων η έκταση κυμαίνεται μεταξύ 180-300 λέξεις. Από τις συγκεκριμένες ΠΓΛ, οι 72 ανήκουν σε μαθητές με αρχικό επίπεδο κατάταξης Α, οι 93 ήταν επιπέδου Β και οι 59 επιπέδου Γ. Οι μαθητικές ΠΓΛ έγιναν αντικείμενο επεξεργασίας από το ποσοτικό εργαλείο QUITA, ώστε να ποσοτικοποιηθούν οι προς εξέταση δείκτες. Τα ποσοστά των δεικτών που προέκυψαν για κάθε ΠΓΛ ανά επίπεδο αναλύθηκαν από το στατιστικό πρόγραμμα SPSS μέσω του παραμετρικού στατιστικού κριτηρίου one way ANOVA.

4.1.1. Ποσοτικά αποτελέσματα

Από την επεξεργασία των γραπτών των μαθητών/-τριών μέσω του εργαλείου QUITA, προκύπτει η αριθμητική απόδοσή των έξι δεικτών ανά επίπεδο, όπως αποτυπώνονται συνοπτικά κατά μέσο όρο στον πίνακα 11. Αναλυτικά, οι τιμές των δεικτών, όπως αποτυπώνονται και για τις 224 ΠΓΛ, παρατίθενται στο Παράστημα Α.

Πίνακας 11: Δεδομένα QUITA: Μέσος Όρος δεικτών ανά επίπεδο γλωσσομάθειας

Επίπεδο	ΠΓΛ	h-Point	Entropy	Average Tokens Length	R1	Λ (Lambda)	Harax Percentage
Α´	72	5,80	6,75	4,64	0,87	1,50	0,45
Β´	93	5,73	6,74	4,71	0,87	1,52	0,48
Γ´	59	5,81	6,79	4,80	0,86	1,54	0,48

Τα στοιχεία που προέκυψαν από το QUITA έγιναν αντικείμενο επεξεργασίας από το παραμετρικό στατιστικό κριτήριο one way ANOVA, για να διαπιστωθεί αν και πόσοι δείκτες είναι στατιστικά σημαντικοί. Στον Πίνακα ANOVA, που παρατίθεται στο Παράτημα Β, ως στατιστικά σημαντικοί δείκτες αναδείχθηκαν: 1) το Ποσοστό Άπαξ Λεγομένων (HL), το Lambda (Λ) και το Μέσο Μήκος Λέξης (ATL). Για τους τρεις αυτούς δείκτες διενεργήθηκε και Post Hoc Test για να διερευνηθούν περαιτέρω διαφορές μεταξύ των επιπέδων. Στον Πίνακα 12 αναγράφεται ο δείκτης σημαντικότητας (sig.) για κάθε παράγοντα λεξιλογικού «πλούτου».

Πίνακας 12: Ο δείκτης σημαντικότητας (sig.) για κάθε παράγοντα λεξιλογικού πλούτου

Δείκτες	h-Point	Entropy	Average Tokens Length	R1	Λ (Lambda)	Hapax Percentage
sig.	0,78	0,35	0,001	0,51	0,02	0,005

4.1.1.1. Δείκτης h-point

Πρώτος προς διερεύνηση δείκτης υπήρξε ο h-point. Όπως μαρτυρά και ο Πίνακας 12, η τιμή σημαντικότητας (sig.) για τα τρία επίπεδα γλωσσομάθειας αναφορικά με τον δείκτη h-point είναι 0,78. Εφόσον Sig. >0.05, τότε το δείγμα δεν παρουσιάζει στατιστικά σημαντική διαφοροποίηση στους μέσους όρους των επιπέδων γλωσσομάθειας.

4.1.1.2. Δείκτης Εντροπία (Entropy)

Όσον αφορά τον δεύτερο δείκτη λεξιλογικού πλούτου, την εντροπία, όπως φαίνεται και στον Πίνακα 12, η τιμή σημαντικότητας (sig.) μεταξύ των τριών επιπέδων γλωσσομάθειας είναι 0,35, δηλαδή Sig. >0.05. Κατ' επέκταση, το δείγμα δεν παρουσιάζει στατιστικά σημαντική διαφοροποίηση στους μέσους όρους των επιπέδων γλωσσομάθειας αναφορικά με την Εντροπία.

4.1.1.3. Δείκτης Average Token Length (ATL ή Μέσο Μήκος Λεξικού Τύπου)

Για τον δείκτη ATL, όπως προκύπτει από τον Πίνακα 12, η τιμή σημαντικότητας (sig.) για τα τρία επίπεδα γλωσσομάθειας είναι 0,001. Εφόσον Sig. < 0.05, τότε ο μέσος όρος του ATL ανά γλωσσικό επίπεδο φανερώνει μια συστηματική διαφοροποίηση που είναι στατιστικά σημαντική για το δείγμα αυτό. Για περαιτέρω διερεύνηση του αποτελέσματος έγινε post hoc test, μέσω του οποίου αναδείχθηκαν εναργέστερα οι διαφορές μεταξύ των τριών επιπέδων. Όπως αποτυπώνεται στον Πίνακα 13, μεγαλύτερη διαφορά εντοπίζεται μεταξύ των Επιπέδων Α' και Γ' όπου sig.= 0 και μικρότερη μεταξύ Α' και Β' (sig.=0,11). Συμπληρωματικά στοιχεία αναφορικά με το post hoc test εμπεριέχονται στο Παράρτημα Β.

Πίνακας 13: Post Hoc Tests: για τον δείκτη ATL

(I) ΕΠΙΠΕΔΟ	(J) ΕΠΙΠΕΔΟ	Sig.
ΕΠΙΠΕΔΟ Α	ΕΠΙΠΕΔΟ Β	,110
	ΕΠΙΠΕΔΟ Γ	,000
ΕΠΙΠΕΔΟ Β	ΕΠΙΠΕΔΟ Α	,110
	ΕΠΙΠΕΔΟ Γ	,085
ΕΠΙΠΕΔΟ Γ	ΕΠΙΠΕΔΟ Α	,000
	ΕΠΙΠΕΔΟ Β	,085

4.1.1.4. Δείκτης R1

Σε αντίθεση με τον δείκτη ATL, όπως φαίνεται στον Πίνακα 12, η τιμή σημαντικότητας (sig.) μεταξύ των τριών επιπέδων γλωσσομάθειας για τον δείκτη R1 είναι 0,51. Εφόσον Sig. >0.05, το δείγμα δεν παρουσιάζει στατιστικά σημαντική διαφοροποίηση στους μέσους όρους των επιπέδων γλωσσομάθειας.

4.1.1.5. Δείκτης LAMBDA (Λ)

Για τον δείκτη LAMBDA η τιμή σημαντικότητας (sig.) για τα τρία επίπεδα γλωσσομάθειας είναι 0,02. Εφόσον Sig. < 0.05, τότε ο μέσος όρος του Λ ανά γλωσσικό επίπεδο διαφέρει από τα άλλα γλωσσικά επίπεδα με τρόπο που να φανερώνει μια συστηματική διαφοροποίηση που είναι στατιστικά σημαντική για το δείγμα αυτό. Εφόσον παρατηρήθηκε στατιστικά σημαντική διαφοροποίηση, ακολούθησε post hoc test, μέσω του οποίου αναδείχθηκαν οι διαφορές μεταξύ των τριών επιπέδων. Όπως καταδεικνύει ο Πίνακας 14, προκύπτει ότι μεγαλύτερη απόκλιση υπάρχει μεταξύ των Επιπέδων Α' και Γ' όπου sig.= 0,022 και μικρότερη μεταξύ Β' και Γ' (sig.=0,626). Συμπληρωματικά στοιχεία αναφορικά με το post hoc test εμπεριέχονται στο Παράρτημα Β.

Πίνακας 14: Post Hoc Tests: για τον δείκτη Δ

(I) ΕΠΙΠΕΔΟ	(J) ΕΠΙΠΕΔΟ	Sig.
ΕΠΙΠΕΔΟ Α	ΕΠΙΠΕΔΟ Β	,112
	ΕΠΙΠΕΔΟ Γ	,022
ΕΠΙΠΕΔΟ Β	ΕΠΙΠΕΔΟ Α	,112
	ΕΠΙΠΕΔΟ Γ	,626
ΕΠΙΠΕΔΟ Γ	ΕΠΙΠΕΔΟ Α	,022
	ΕΠΙΠΕΔΟ Β	,626

4.1.1.6. Δείκτης Ποσοστού Άπαξ Λεγομένων (HL)

Η τιμή σημαντικότητας (sig.) για τα τρία επίπεδα αναφορικά με τον τελευταίο δείκτη, το ποσοστό των άπαξ λεγομένων, είναι 0,005. Δεδομένου ότι Sig. <0.05, τότε ο μέσος όρος του HL ανά γλωσσικό επίπεδο διαφέρει από τα άλλα γλωσσικά επίπεδα με τρόπο που να φανερώνει μια συστηματική διαφοροποίηση που είναι στατιστικά σημαντική για το δείγμα αυτό. Εφόσον παρατηρήθηκε στατιστικά σημαντική διαφοροποίηση, διενεργήθηκε post hoc test, μέσω του οποίου αναδείχθηκαν καθαρότερα οι διαφορές μεταξύ των επιπέδων γλωσσομάθειας και παρουσιάζονται στον Πίνακα 15. Είναι προφανές ότι μεγαλύτερη διαφορά εντοπίζεται μεταξύ των Επιπέδων Α' και Β' όπου sig.= 0,011, ενώ σημαντική είναι και η διαφορά μεταξύ Επιπέδων Α' και Γ'(sig.= 0,013). Μικρότερη διαφορά υπάρχει μεταξύ Β' και Γ' (sig.=0,958). Συμπληρωματικά στοιχεία αναφορικά με το post hoc test εμπεριέχονται στο Παράρτημα.

Πίνακας 15: Post Hoc Tests: για τον δείκτη HL

(I) ΕΠΙΠΕΔΟ	(J) ΕΠΙΠΕΔΟ	Sig.
ΕΠΙΠΕΔΟ Α	ΕΠΙΠΕΔΟ Β	,011
	ΕΠΙΠΕΔΟ Γ	,013
ΕΠΙΠΕΔΟ Β	ΕΠΙΠΕΔΟ Α	,011
	ΕΠΙΠΕΔΟ Γ	,958
ΕΠΙΠΕΔΟ Γ	ΕΠΙΠΕΔΟ Α	,013
	ΕΠΙΠΕΔΟ Β	,958

Λαμβάνοντας υπόψη τα παραπάνω δεδομένα, τα στατιστικά σημαντικά δείγματα σχετίζονται με τους δείκτες ATL, Lambda και HL. Ο λεξιλογικός δείκτης ATL εμφανίζει μεγαλύτερη διαφορά μεταξύ των Επιπέδων Α' και Γ' (sig.= 0), όπως και ο δείκτης LAMBDA (sig.= 0,022). Μεγαλύτερη απόκλιση εντοπίζεται μεταξύ των Επιπέδων Α' και Γ' (sig.= 0,013) και στον δείκτη HL. Ο τελευταίος δείκτης εμφανίζει στατιστικά σημαντικό αποτέλεσμα και μεταξύ των Επιπέδων Α' και Β' (sig.= 0,011), σε αντίθεση με τους δύο άλλους δείκτες.

4.1.2. Ποιοτική επεξεργασία δεδομένων

Οι δείκτες λεξιλογικού «πλούτου» που συνδέονται με τα στατιστικά σημαντικά δείγματα πιστοποιούν την απουσία ομοιογένειας στις ΠΓΛ των μαθητών/-τριών που προβιβάστηκαν στο Β2 Επίπεδο Γλωσσομάθειας. Τα αποτελέσματα αποδεικνύουν ότι η ανομοιογένεια σχετίζεται άμεσα με την προηγούμενη γνώση που είχαν γύρω από την ελληνική γλώσσα. Σύμφωνα με τις κατατακτήριες εξετάσεις που έδωσαν προκειμένου να στελεχωθούν ομοιογενώς τα τμήματα, οι μαθητές/-τριες προέρχονταν και από τα τρία επίπεδα ελληνομάθειας. Το επίπεδο της αρχικής κατάταξής τους σχετίζεται άμεσα με το ποσοστό εμφάνισης των λεξιλογικών δεικτών στο λόγο τους και συγκεκριμένα με τρεις από τους υπό εξέταση δείκτες, το μέσο μήκος του λεξικού τύπου (ATL), το Lambda (Λ) και το Ποσοστό των Άπαξ Λεγομένων (HL). Οι άλλοι τρεις δείκτες, το h-point, η εντροπία και ο R1, δε φάνηκε να επηρεάζονται σημαντικά από την αρχική γνώση των εκπαιδευομένων.

Όσον αφορά τον δείκτη ATL, αποδείχθηκε πως το μέσο μέγεθος των λεξικών τύπων διαφέρει σημαντικά μεταξύ του Επιπέδου Α' και Γ', ενώ δεν υπάρχει σημαντική διαφοροποίηση μεταξύ του Α' και Β' Επιπέδου. Ο δείκτης Lambda, επίσης, διαφοροποιείται σημαντικά μεταξύ Α' και Γ' επιπέδου, ενώ μικρή είναι η διαφορά ανάμεσα στα γραπτά μαθητών/-τριών προερχόμενων από το Επίπεδο Β' και Γ'. Τρίτος και τελευταίος στατιστικά σημαντικός δείκτης λεξιλογικού «πλούτου» αποδείχθηκε αυτός που σχετίζεται με την εμφάνιση των άπαξ λεγομένων, λέξεων, δηλαδή, που εμφανίζονται μόνο μία φορά στην ΠΓΛ. Σύμφωνα με τα αποτελέσματα της έρευνας, όσοι/-ες μαθητές/-τριες προέρχονταν από το Επίπεδο Β' και Γ' παρουσίασαν σημαντική διαφορά σε σχέση με όσους/-ες προέρχονταν από το Επίπεδο Α' στο ποσοστό χρήσης λέξεων άπαξ (HL) στην ΠΓΛ.

4.2. Προβλεπτική εξίσωση πολλαπλής παλινδρόμησης

Όπως διαπιστώθηκε στην ενότητα 4.1., μέσω του παραμετρικού στατιστικού κριτηρίου one way ANOVA, τρεις από τους έξι δείκτες λεξιλογικού «πλούτου» αποδείχθηκαν στατιστικά σημαντικοί, δηλαδή φάνηκε πως υπάρχει διαφοροποίηση στις τιμές τους σε συγκεκριμένα επίπεδα γλωσσομάθειας. Βάσει αυτών των αποτελεσμάτων, θα ήταν χρήσιμο να εξετασθεί αν οι προς διερεύνηση δείκτες λήφθηκαν υπόψη στη βαθμολόγηση των γραπτών των μαθητών/-τριών. Για να διαπιστωθεί η συσχέτιση των δεικτών με τη βαθμολογία ανά επίπεδο, αξιοποιήθηκε η προβλεπτική εξίσωση της πολλαπλής παλινδρόμησης (Multiple Regression Analysis- MRA). Μέσω αυτής μελετάται η απόκριση μίας εξαρτημένης μεταβλητής ως γραμμικό αποτέλεσμα της δράσης πολλών ανεξάρτητων μεταβλητών. Στη δεδομένη περίπτωση, ως εξαρτημένη μεταβλητή ορίζεται ο βαθμός λεξιλογίου των μαθητών/-τριών, ο οποίος καταλαμβάνει το ανώτερο 6 μονάδες στις 20 του συνολικού βαθμού. Ως ανεξάρτητες μεταβλητές ορίζονται οι έξι δείκτες λεξιλογικού «πλούτου», Harax Legomenon Percentage (HL) ή ποσοστό άπαξ λεγομένων, R1, Lambda (Λ), h-point (h), Entropy (H) ή εντροπία, Average Tokens length (ATL) ή μέσο μήκος λεξικού τύπου.

4.2.1. Ποσοτικά αποτελέσματα

Για να διαπιστωθεί αν στη βαθμολόγηση της ΠΓΛ των του μαθητικού κοινού (εξαρτημένη μεταβλητή) λήφθηκαν υπόψη οι προς διερεύνηση λεξιλογικοί δείκτες (ανεξάρτητες μεταβλητές) εφαρμόστηκε η προβλεπτική εξίσωση της πολλαπλής παλινδρόμησης. Διενεργήθηκε τρεις φορές εξετάζοντας τους δείκτες λεξιλογικού «πλούτου» και τις βαθμολογίες για κάθε επίπεδο γλωσσομάθειας ξεχωριστά, προκειμένου να διαπιστωθεί αν υπάρχει και σε ποιο/ποια επίπεδα ισχυρή συσχέτιση μεταξύ εξαρτημένης και ανεξάρτητων μεταβλητών.

Από την εφαρμογή της εξίσωσης πολλαπλής παλινδρόμησης προκύπτουν τρεις πίνακες, Model Summary, ANOVA και Coefficients, οι οποίοι παρατίθενται στο παράρτημα. Οι τιμές που αποδεικνύουν τη συσχέτιση μεταξύ εξαρτημένης μεταβλητής (βαθμολογία γραπτών ανά επίπεδο γλωσσομάθειας) και ανεξάρτητων μεταβλητών (Ποσοστό Άπαξ Λεγομένων, R1, Lambda, h-point, εντροπία, μέσο μήκος

λεξικού τύπου) είναι το R^2 , το F και το significance (sig.) από τον πίνακα ANOVA και t και significance (sig.) από τον πίνακα Coefficients.

4.2.1.1. Πίνακας Model Summary: τιμή R^2

Το R Square είναι το τετράγωνο του συντελεστή γραμμικής συσχέτισης και ονομάζεται συντελεστής προσδιορισμού. Ο συντελεστής προσδιορισμού φανερώνει το ποσοστό της μεταβλητότητας των δεδομένων που εξηγείται από το γραμμικό μοντέλο που προσαρμόσαμε. Όπως αποτυπώνεται στον Πίνακα 16, το R^2 (R Square) για το Επίπεδο Α' είναι 0,79 που σημαίνει ότι η μεταβλητή X ερμηνεύει την Y κατά 79% ή η διασπορά της Y ερμηνεύεται κατά 79% από τη διασπορά των ανεξάρτητων μεταβλητών του μοντέλου. Για το Επίπεδο Β' το R^2 είναι 0,78 και για το Επίπεδο Γ' 0,165, δηλαδή η μεταβλητή X ερμηνεύει την Y κατά 78% ή η διασπορά της Y ερμηνεύεται κατά 78% από τη διασπορά του μοντέλου στο Επίπεδο Β' και η μεταβλητή X ερμηνεύει την Y κατά 16,5% ή η διασπορά της Y ερμηνεύεται κατά 16,5% από τη διασπορά του μοντέλου στο Επίπεδο Γ'. Οι μετρήσεις για το R^2 αποτυπώνονται αναλυτικότερα στο Παράρτημα.

Ο συντελεστής προσδιορισμού R^2 εκφράζει το ποσοστό της ολικής μεταβλητότητας της Y (βαθμολογία λεξιλογίου). Εξηγείται από τη συνδυασμένη επίδραση όλων των μεταβλητών X (ποσοστό άπαξ λεγομένων, R1, Lambda, H-point, Εντροπία, μέσο μήκος λεξικού τύπου) που συμμετέχουν στην περιγραφή της εξίσωσης της πολλαπλής γραμμικής παλινδρόμησης επί της εξαρτημένης Y (επίπεδο γλωσσομάθειας). Ο συντελεστής R^2 λαμβάνει τιμές από μηδέν (κανένα ποσοστό ερμηνείας της εξαρτημένης μεταβλητής) μέχρι 1 (άριστη ερμηνεία της εξαρτημένης μεταβλητής). Εφόσον το R^2 είναι υψηλό για τα Επίπεδα Α' και Γ', αφού λαμβάνει τιμές 0,79 και 0,78 αντίστοιχα, το μοντέλο έχει καλή προσαρμογή σ' αυτά. Δεν ισχύει το ίδιο για το Επίπεδο Γ', στο οποίο το R^2 είναι 0,17. Επομένως οι ανεξάρτητες μεταβλητές που χρησιμοποιήσαμε δεν ερμηνεύουν ικανοποιητικά την εξαρτημένη μας μεταβλητή.

Πίνακας 16: Η τιμή R^2 στα τρία επίπεδα γλωσσομάθειας

Επίπεδο	Α'	Β'	Γ'
R^2	0,79	0,78	0,165

4.2.1.2. Πίνακας ANOVA: τιμές F και Significance

Η στατιστική σημαντικότητα της πολλαπλής παλινδρόμησης μπορεί να εξεταστεί από την τιμή F και το significance (sig.). Αν το Sig. < 0,05 τότε το F είναι μεγάλο, άρα η γραμμική παλινδρόμηση είναι στατιστικά σημαντική.

Στους πίνακες ANOVA για τα τρία επίπεδα γλωσσομάθειας, η εστίαση γίνεται στο F και το significance. Όπως φαίνεται στον Πίνακα 17, στο Επίπεδο Α' το sig. είναι 0,482, άρα >0,05, και το F δεν είναι μεγάλο (0,926). Επομένως, η γραμμική παλινδρόμηση δεν είναι στατιστικά σημαντική. Όσον αφορά το Επίπεδο Β' το sig. είναι 0,310, άρα > 0,05, και το F 1,208. Τέλος, στο Επίπεδο Γ' το sig. είναι 0,137, άρα > 0,05, και το F 1,709. Άρα, και στα Επίπεδα Β' και Γ' η γραμμική παλινδρόμηση δεν εμφανίζεται στατιστικά σημαντική. Εφόσον το sig. και στα τρία επίπεδα ελληνομάθειας είναι >0,05, το μοντέλο μας δεν έχει καλή προσαρμογή.

Πίνακας 17: Οι τιμές F και Significance για τα τρία επίπεδα γλωσσομάθειας

Επίπεδο	Α'	Β'	Γ'
F	0,926	1,208	1,709
sig.	0,482	0,310	0,137

4.2.1.3. Πίνακας Coefficients : τιμές B, Beta, t και significance

Από τον τελευταίο Πίνακα που εξάγεται στο Output, τον πίνακα Coefficients, ο οποίος παρατίθεται στο παράρτημα και για τα τρία επίπεδα γλωσσομάθειας, αντλούμε πληροφορίες μέσω της τιμής B για την εξίσωση της παλινδρόμησης : $y = b_0 + b_1x$ | $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6$, όπου y είναι η εξαρτημένη μεταβλητή (βαθμολογία λεξιλογίου) και x_{1-n} οι ανεξάρτητες μεταβλητές (δείκτες λεξιλογικού πλούτου). Επίσης, μέσω των Beta, μπορούμε να ταξινομήσουμε τις ανεξάρτητες μεταβλητές ως προς την ερμηνευτική τους ικανότητα. Δεν λαμβάνουμε υπόψη τα πρόσημα, αλλά τις απόλυτες τιμές. Ενδεικτικά, όσον αφορά το Επίπεδο Α', πρώτος σε ερμηνευτική ισχύ είναι ο λεξιλογικός δείκτης Lambda (Beta = 0,38), δεύτερη η Εντροπία (Beta = 0,284), τρίτος ο δείκτης μέσο μήκος λεξικού τύπου (Beta = 0,245), τέταρτος ο δείκτης R1 (Beta = 0,203), πέμπτος ο δείκτης h-point (Beta = 0,143) και τελευταίος ο δείκτης που αφορά το ποσοστό των άπαξ λεγομένων (Beta = 0,009).

Πίνακας 18: Η τιμή Beta για τους έξι δείκτες από τον πίνακα Coefficients του Output

Δείκτες	Beta
H-point	0,143
Εντροπία	-0,284
Μέσο Μήκος	-0,245
R1	0,203
LAMBDA	0,380
Άπαξ Λεγόμενα	0,009

Ιδιαίτερη έμφαση αξίζει να δοθεί στους δείκτες t και significance που συνιστούν τον έλεγχο σημαντικότητας των ανεξάρτητων μεταβλητών. Αν το $p = \text{Sig.} < 0,05$, οι μεταβλητές x είναι στατιστικά σημαντικοί. Συγκεκριμένα, όσον αφορά το Επίπεδο Α', $t = 0,279$ και $\text{sig.} = 0,781 > 0,05$ άρα οι συντελεστές της x (δείκτες λεξιλογικού πλούτου) δεν είναι στατιστικά σημαντικοί. Στο Επίπεδο Β', $t = -0,278$ και $\text{sig.} = 0,781 > 0,05$ αλλά και στο Επίπεδο Γ', $t = -0,922$ και $\text{sig.} = 0,361 > 0,05$. Άρα, οι δείκτες λεξιλογικού πλούτου και στα δύο επόμενα επίπεδα γλωσσομάθειας δεν είναι στατιστικά σημαντικοί.

Πίνακας 19: Οι τιμές t & sig. ανά επίπεδο από τον πίνακα Coefficients του Output

Επίπεδο	t	sig.
Α'	0,279	0,781
Β'	-0,278	0,781
Γ'	-0,922	0,361

4.2.2. Ποιοτική ανάλυση αποτελεσμάτων

Μέσω της εξέτασης του συντελεστή προσδιορισμού R^2 , της ειδικής τιμής του F και του δείκτη σημαντικότητας (sig.) από τον πίνακα ANOVA, του t και του δείκτη σημαντικότητας (sig.) από τον πίνακα Coefficients, αποδείχθηκε ότι το μοντέλο μας δεν έχει καλή προσαρμογή και πως οι δείκτες λεξιλογικού πλούτου και στα τρία επίπεδα γλωσσομάθειας δεν είναι στατιστικά σημαντικοί. Με άλλα λόγια, είναι προφανές πως η βαθμολογία του λεξιλογίου στις ΠΓΛ των μαθητών/-τριών δεν επηρεάστηκε από τους υπό εξέταση λεξιλογικούς δείκτες.

Όσον αφορά τον συντελεστή R^2 που κυμαίνεται από 0 (μηδενική ερμηνευτική ισχύς) μέχρι 1 (ερμηνευτική ισχύς), λαμβάνει τις τιμές 0,79, 0,78, για τα Επίπεδα Α' και Β' αντίστοιχα. Οι τιμές αυτές πλησιάζουν το 1, γεγονός που προμηνύει καλή προσαρμογή. Ο συντελεστής R^2 για το Επίπεδο Γ είναι 0,165, δηλαδή στο επίπεδο αυτό η προσαρμογή δεν είναι καλή.

Πέραν του R^2 , εξετάστηκαν η ειδική τιμή F και ο δείκτης σημαντικότητας (sig.) όπως αποτυπώθηκαν στον Πίνακα ANOVA του Output και κατέδειξαν την απουσία καλής προσαρμογής. Συγκεκριμένα, στο Επίπεδο Α' το sig. είναι 0,482, άρα $>0,05$, και το F δεν είναι μεγάλο (0,926). Στο Επίπεδο Β' το sig. είναι 0,310, άρα $>0,05$, και το F 1,208. Τέλος, στο Επίπεδο Γ' το sig. είναι 0,137, άρα $>0,05$, και το F 1,709. Άρα, και στα τρία επίπεδα γλωσσομάθειας το sig. είναι $>0,05$ και το F έχει μικρή τιμή. Βάσει αυτών των τιμών η γραμμική παλινδρόμηση δεν εμφανίζεται στατιστικά σημαντική και το μοντέλο δεν έχει καλή προσαρμογή.

Τέλος, η συνεξέταση του δείκτη t και sig. όπως προέκυψαν από τον Πίνακα Coefficients επιβεβαιώνουν ότι οι έξι δείκτες λεξιλογικού πλούτου που εξετάστηκαν δεν είναι στατιστικά σημαντικοί. Όσον αφορά το Επίπεδο Α', $t = 0,279$ και sig. = 0,781 $>0,05$. Στο Επίπεδο Β', $t = -0,278$ και sig. = 0,781 $>0,05$ αλλά και στο Επίπεδο Γ', $t = -0,922$ και sig. = 0,361 $>0,05$. Άρα, οι δείκτες t και sig. επιβεβαιώνουν ότι η βαθμολογία των γραπτών των μαθητών δε σχετίζεται με τους υπό εξέταση δείκτες.

Παρά το γεγονός ότι τρεις από τους έξι λεξιλογικούς δείκτες που εξετάζονται στην παρούσα εργασία έχει αποδειχθεί στην ενότητα 4.1. ότι είναι στατιστικά σημαντικοί και διαφοροποιούνται ανά επίπεδο γλωσσομάθειας (το μέσο μήκος του λεξικού τύπου, το Lambda, το Ποσοστό των Άπαξ Λεγομένων), φάνηκε από την εξίσωση πολλαπλής παλινδρόμησης ότι δεν επηρέασαν το βαθμολογικό κριτήριο των διορθωτών του λεξιλογίου των ΠΓΛ των μαθητών. Τα αποτελέσματα της ενότητας 4.2. αποτελούν εφαλτήριο για περαιτέρω διερεύνηση και αναθεώρηση των κριτηρίων των βαθμολογητών του λεξιλογίου στα μαθητικά γραπτά.

ΚΕΦΑΛΑΙΟ 5: ΔΙΔΑΚΤΙΚΕΣ ΠΡΟΕΚΤΑΣΕΙΣ

Στο κεφάλαιο αυτό επιχειρείται η σύνδεση της θεωρίας με την πράξη, δηλαδή η αξιοποίηση των αποτελεσμάτων της παρούσας έρευνας στη διδακτική πραγματικότητα και συγκεκριμένα στη βαθμολόγηση. Τα πορίσματα της έρευνας, όπως παρουσιάστηκαν στο Κεφάλαιο 4, απέδειξαν ότι οι υπό εξέταση δείκτες λεξιλογικού πλούτου (Ποσοστό Άπαξ Λεγομένων, R1, Lambda, h-point, Εντροπία, Μέσο Μήκος Λεξικού Τύπου) δε λαμβάνονται υπόψη από τους βαθμολογητές, παρά το γεγονός ότι οι τρεις από αυτούς κρίθηκαν στατιστικά σημαντικοί (Μέσο Μήκος Λεξικού Τύπου, Lambda, Ποσοστό Άπαξ Λεγομένων). Ως εκ τούτου, καταδεικνύεται η ανάγκη αναθεώρησης του ισχύοντος πλαισίου, προκειμένου η βαθμολόγηση να είναι πιο αντικειμενική και αντιπροσωπευτική των δυνατοτήτων των μαθητών/-τριών σε λεξιλογικό επίπεδο. Το κεφάλαιο αυτό διακρίνεται ουσιαστικά σε δύο ενότητες, την ενότητα 5.1. η οποία αφορά τον ισχύοντα τρόπο βαθμολόγησης και την ενότητα 5.2. στην οποία εμπεριέχονται διαρθρωτικές αλλαγές στο ισχύον καθεστώς βαθμολόγησης βάσει των ερευνητικών αποτελεσμάτων. Ακολουθεί ένα τρίτο κεφάλαιο (5.3.) που περιλαμβάνει την αποτίμηση των προτάσεων.

5.1. Ισχύον πλαίσιο βαθμολόγησης του ΣΕΜΚ

Για τη βαθμολόγηση του ΣΕΜΚ υιοθετήθηκε η αναλυτική αξιολόγηση, δηλαδή η μέθοδος σύμφωνα με την οποία για κάθε ξεχωριστό κριτήριο στο πλάνο αξιολόγησης αποδίδεται ένας διαφορετικός βαθμός και ο τελικός βαθμός είναι μια σύνθεση αυτών των ξεχωριστών εκτιμήσεων" (Weir, 1990:63). Οι διαφορετικές εκτιμήσεις, στην περίπτωση του ΣΕΜΚ, σχετίζονται με διαφορετικές δεξιότητες στο πλαίσιο της Παραγωγή Γραπτού Λόγου (ΠΓΛ) του αλλόγλωσσου μαθητικού κοινού και αφορούν το περιεχόμενο, τη γραμματική – σύνταξη και το λεξιλόγιο. Οι συγκεκριμένες ΠΓΛ εντάσσονται στο B2 Επίπεδο Ελληνομάθειας, όπου ο/η υποψήφιος/-α καλείται να παραγάγει γραπτό κείμενο εκτάσεως 200 περίπου λέξεων και βαθμολογείται με είκοσι (20) μονάδες συνολικά. Οι μονάδες αυτές καταμερίζονται ως εξής: επτά μονάδες (7) αντιστοιχούν στο περιεχόμενο, επτά μονάδες στη γραμματική-σύνταξη

(7) και έξι μονάδες (6) στο λεξιλόγιο. Ο συγκεκριμένος καταμερισμός των ομάδων ισχύει όχι μόνο για το ΣΕΜΚ, αλλά για όσες ΠΓΛ εντάσσονται σε εξετάσεις οι οποίες διενεργούνται υπό την αιγίδα του Διδασκαλείου Νέας Ελληνικής Γλώσσας, όπως πιστοποιεί η Βεβαίωση Ελληνομάθειας για το Β2 Επίπεδο.⁴¹ Δεδομένης της διερεύνησης των δεικτών λεξιλογικού πλούτου σε σχέση με τα επίπεδα γλωσσομάθειας στην παρούσα εργασία, η εστίαση έγινε στο λεξιλόγιο.

5.1.1. Αξιολόγηση του ισχύοντος πλαισίου βαθμολόγησης του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας

Η βαθμολόγηση των παρουσών ΠΓΛ των μαθητών/-τριών αλλά και εν γένει των ΠΓΛ στο πλαίσιο των εξετάσεων του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας γίνεται βάσει της αναλυτικής αξιολόγησης. Η στροφή στη συγκεκριμένη μέθοδο προέκυψε από την ανάγκη για μεγαλύτερη αξιοπιστία και αντικειμενικότητα στα αποτελέσματα των γραπτών δοκιμασιών αλλά και για πιο λεπτομερή ανάλυση των ικανοτήτων των εξεταζομένων, με στόχο την πληρέστερη ενημέρωση αυτών που θα κάνουν χρήση των αποτελεσμάτων (Ηλιοπούλου, 2009). Οι Cohen (1994:318) και Weigle (2002:120) υποστηρίζουν, μάλιστα, ότι τα αναλυτικά σχέδια αξιολόγησης καθιστούν την αναλυτική βαθμολογία ως ένα βαθμό ιδιαίτερα χρήσιμο εργαλείο για την εκπαίδευση των αξιολογητών και την τυποποίηση και σταθεροποίηση των αποτελεσμάτων τους.

Η επιλογή του συγκεκριμένου τύπου αξιολόγησης δε θα μπορούσε να τεθεί υπό αίρεση, ειδικά αν αναλογιστεί κανείς τα μειονεκτήματα της ολιστικής αξιολόγησης σύμφωνα με την οποία ο γραπτός λόγος αντιμετωπίζεται ως ένα αδιάσπαστο σύνολο (O'Malley-Pierce, 1996:142). Τον περιορισμένο βαθμό αξιοπιστίας της συγκεκριμένης μεθόδου αποδεικνύουν σε έρευνά τους οι Spandel & Stiggins (1980) στον Lee

⁴¹ Η Βεβαίωση Ελληνομάθειας είναι το επίσημο πιστοποιητικό γνώσης της Ελληνικής ως ξένης γλώσσας που χορηγείται από το Πανεπιστήμιο Αθηνών σε αλλοδαπούς. Η εξέταση πραγματοποιείται στον χώρο του Πανεπιστημίου Αθηνών δύο φορές τον χρόνο (Μάιο και Σεπτέμβριο) και στοχεύει στην αξιολόγηση της γλωσσικής και επικοινωνιακής ικανότητας των υποψηφίων βάσει των κριτηρίων που έχουν καθοριστεί από το Συμβούλιο της Ευρώπης για το Επίπεδο Επάρκειας (Vantage Level). Ανακτήθηκε από:

http://www.greekcourses.uoa.gr/fileadmin/greekcourses.uoa.gr/uploads/DOCUMENTS_DIAFORA/TEST_ELLHNOMATHEIAS.pdf

(2004:11) εξαιτίας της ταχύτητας ανάγνωσης των γραπτών σε αντιδιαστολή με την επικεντρωμένη σε συγκεκριμένα χαρακτηριστικά και για αυτό μεγαλύτερης διάρκειας και άρα πιο αξιόπιστης ανάγνωσης των γραπτών της αναλυτικής προσέγγισης. Επίσης, εμφανίζει μεγαλύτερο περιθώριο υποκειμενικότητας σε σχέση με την αναλυτική αξιολόγηση αφού δε δίνουν όλοι/-ες οι αξιολογητές/-τριες την ίδια βαρύτητα στα ίδια σημεία (Ηλιοπούλου, 2009). Τέλος, για τους Hamp-Lyons & Henning (1991:340), η αναλυτική αξιολόγηση συνιστά τη δυνατότητα να παρέχονται ξεχωριστές βαθμολογίες στις διαφορετικές δεξιότητες ως μια εναλλακτική μέθοδο απέναντι στην απόδοση πολλαπλών βαθμολογιών σε μια δοκιμασία, κάτι που συμβαίνει κατά κόρον στην ολιστική. Έτσι, επιτρέπονται αξιόπιστες σύνθετες βαθμολογίες.

5.1.2. Αξιολόγηση του ισχύοντος πλαισίου βαθμολόγησης του Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς για τη γλώσσα (ΚΕΠΑ)

Στο Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς για τη γλώσσα (ΚΕΠΑ), ορίζονται με περιεκτικό τρόπο οι κατευθυντήριες γραμμές για την εκμάθηση, τη διδασκαλία και την αξιολόγηση μιας Γ2 με απώτερο σκοπό την ενίσχυση της διαφάνειας των μαθημάτων, των αναλυτικών προγραμμάτων διδασκαλίας και των τίτλων γλωσσομάθειας, προωθώντας τη διεθνή συνεργασία στον τομέα των σύγχρονων γλωσσών (ΚΕΠΑ, 2008:1). Κατ' αυτόν τον τρόπο, θα θεωρούνταν φυσικό επακόλουθο ο τύπος βαθμολόγησης να ήταν προκαθορισμένος και τα κριτήρια ενιαία για την εκμάθηση οποιασδήποτε γλώσσας συμμορφωνόταν με το εν λόγω πλαίσιο.

Ωστόσο, κάτι τέτοιο δεν ισχύει. Εστιάζοντας στη λεξιλογική ικανότητα των μαθητών⁴², δεδομένου ότι αντικείμενο διερεύνησης είναι οι έξι δείκτες λεξιλογικού πλούτου, διαπιστώνουμε μια σχετικιστική προσέγγιση αναφορικά με το λεξιλογικό φάσμα. Σύμφωνα με το ΚΕΠΑ, ο/η μαθητής/-τρια που εντάσσεται στο Επίπεδο Επάρκειας (B2) διαθέτει ικανοποιητικό φάσμα λεξιλογίου για θέματα που συνδέονται με τον τομέα του και τα περισσότερα γενικά θέματα. Μπορεί να τροποποιεί τη

⁴² Η λεξιλογική ικανότητα, η γνώση και η ικανότητα χρήσης του λεξιλογίου της γλώσσας, αποτελείται από λεξικά στοιχεία και γραμματικά στοιχεία και είναι μία τις γλωσσικές δεξιότητες. Άλλες γλωσσικές δεξιότητες θεωρούνται η γραμματική ικανότητα, η σημασιολογική ικανότητα, η φωνολογική ικανότητα, η ορθογραφική ικανότητα και η ορθοφωνική ικανότητα (ΚΕΠΑ, 2008:127-128)

διατύπωση για να αποφεύγει τις συχνές επαναλήψεις, αλλά τα λεξιλογικά κενά μπορεί ακόμα να προκαλούν διστακτικότητα και προσφυγή σε περιφράσεις (ΚΕΠΑ, 130). Τα γνωρίσματα που φέρει ο/η κάθε μαθητής/-τρια με βάση το γνωστικό του/της επίπεδο επισημαίνονται με σαφήνεια και στον Πίνακα Αυτοαξιολόγησης που παρατίθεται στο πλαίσιο του Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς.⁴³

Πρόκειται, λοιπόν, για μία πιο ολιστική, γενικευτική προσέγγιση στην περιγραφή των γλωσσών Αυτό έχει ως αποτέλεσμα να γίνει το ισχύον πλαίσιο ιδιαίτερα ευέλικτο σε σχέση με την έκδοση του ΚΕΠΑ του 2001 που αφαιρέθηκαν οι λίστες λεξιλογίου που ανταποκρίνονταν σε κάθε γλωσσικό στάδιο. Το επιχείρημα για αυτή την αλλαγή ήταν ότι οι περιγραφητές πρέπει να παραμένουν ολιστικοί (ΚΕΠΑ, 2008:205). Η υπερβολική αυτή, όμως, ευελιξία οδήγησε σε πολλές παρανοήσεις.

Τελικά, το ΚΕΠΑ παρέχει πληροφορίες για το θεωρητικό πλαίσιο της αξιολόγησης και τους τύπους αξιολόγησης, χωρίς να αναφέρεται στις μονάδες που αντιστοιχούν στις εξετάσεις για κάθε δραστηριότητα. Όπως αναφέρουν χαρακτηριστικά οι εισηγητές του «Δε θεωρούμε ότι, σε μια πρακτική προσέγγιση αξιολόγησης, πρέπει να χρησιμοποιηθούν όλες οι κλίμακες σε όλα τα επίπεδα. Έτσι, το σύνολο των κλιμάκων προορίζεται για να χρησιμοποιηθεί μάλλον ως εργαλείο αναφοράς» (ΚΕΠΑ, 2008:223).

5.2. Συμπερίληψη στη βαθμολόγηση των στατιστικά σημαντικών δεικτών λεξιλογικού πλούτου

Παρ' ότι σκοπός του ΚΕΠΑ (2008:1) είναι η ενίσχυση της διαφάνειας των μαθημάτων, των αναλυτικών προγραμμάτων διδασκαλίας και των τίτλων γλωσσομάθειας, οι κατευθυντήριες γραμμές για την εκμάθηση, τη διδασκαλία και την αξιολόγηση της Γ2 αποδείχθηκαν - εν πολλοίς- γενικευτικές και επιδεκτικές ποικίλων ερμηνειών και εφαρμογών κατά το δοκούν.

Η υπερβολική του ευελιξία σχετίζεται και με τη λεξιλογική γνώση που προβλέπεται για κάθε επίπεδο γλωσσομάθειας και κυρίως με την βαθμολόγηση. Στο μεν ΚΕΠΑ

⁴³http://www.greekcourses.uoa.gr/fileadmin/greekcourses.uoa.gr/uploads/DOCUMENTS_DIAFORA/LEVELS_GR.pdf

(2008) δε γίνεται καμία αναφορά στην βαθμολόγηση των γραπτών των μαθητών/-τριών στις προαγωγικές εξετάσεις, στο δε ισχύον πλαίσιο βαθμολόγησης του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας, σύμφωνα με το οποίο βαθμολογήθηκαν οι υπό εξέταση ΠΓΛ των μαθητών/-τριών, η βαθμολογία που αντιστοιχεί στο λεξιλόγιο ανέρχεται στις έξι (6) μονάδες, δίχως, όμως, να υπάρχουν σαφή κριτήρια για την κατανομή των μονάδων αυτών. Το συγκεκριμένο έλλειμμα αισθητοποιείται στο ΣΕΜΚ από τον αποκλεισμό από την βαθμολογία των τριών λεξιλογικών δεικτών που κρίθηκαν στατιστικά σημαντικοί στην παρούσα εργασία, όπως αποδείχθηκε μέσω της προβλεπτικής εξίσωσης της πολλαπλής παλινδρόμησης στο κεφάλαιο 4.2.1.. Βάσει του συγκεκριμένου πορίσματος, κρίνεται αναγκαία η συγκεκριμενοποίηση των βαθμολογικών κριτηρίων που αντιστοιχούν στο λεξιλόγιο. Καθώς η συμπερίληψη στη βαθμολογία των λεξιλογικών κριτηρίων εν γένει ξεπερνά κατά πολύ τον στόχο της συγκεκριμένης εργασίας, η εστίαση, στην παρούσα φάση, θα γίνει στους τρεις λεξιλογικούς δείκτες που αξιολογήθηκαν ως στατιστικά σημαντικοί και είναι αναγκαίο να μη διαφεύγουν την προσοχή των βαθμολογητών. Πρόκειται για το μέσο μήκος του λεξικού τύπου (ATL), το Lambda και το Ποσοστό των Άπαξ Λεγομένων. Οι άλλοι τρεις υπό εξέταση δείκτες, το h-point, η εντροπία και ο R1, δε φάνηκε να επηρεάζονται από το αρχικό επίπεδο γλωσσομάθειας, άρα, δε θα ληφθούν υπόψη.

5.2.1. Η φύση των στατιστικά σημαντικών λεξιλογικών δεικτών και ο τρόπος συμπερίληψής τους στη βαθμολόγηση του λεξιλογίου

Προκειμένου οι βαθμολογητές/-τριες να συμπεριλάβουν στην αξιολόγησή τους δείκτες που έχουν επισημανθεί ως στατιστικά σημαντικοί, είναι απαραίτητο σε πρώτη φάση να κατανοήσουν τη φύση των συγκεκριμένων δεικτών και σε δεύτερη φάση να θεσπιστεί το ακριβές ποσοστό που θα καταλαμβάνει ο κάθε δείκτης σε σχέση με τη συνολική βαθμολογία που θα αντιστοιχεί στη λεξιλογική γνώση των μαθητών/-τριών στο πλαίσιο της εξέτασης μιας Γ2. Ο λόγος γίνεται, φυσικά, για το μέσο μήκος του λεξικού τύπου (ATL), το Lambda (Λ) και το Ποσοστό των Άπαξ Λεγομένων (HL).

5.2.1.1. Μέσο Μήκος του Λεξικού Τύπου (ATL)

Το Μέσο Μήκος του Λεξικού Τύπου σχετίζεται με τον αριθμητικό μέσο όρο των λεξικών τύπων. Η μέτρηση του λεξικού μήκους μπορεί να γίνει τόσο σε χαρακτήρες όσο και σε συλλαβές. Προκρίνεται η μέτρηση του μήκους σε συλλαβές ως πιο αξιόπιστη, γιατί δεν επηρεάζονται από την πολυπλοκότητα του γραφηματικού συστήματος (Grzybek, 2007:18). Βέβαια, για γλώσσες όπως είναι η ελληνική, ο συγκεκριμένος τύπος μέτρησης θεωρείται ενίοτε προβληματικός⁴⁴. Ο μέσος όρος, που ονομάζεται και πρώτη κεντρική ροπή (first central moment), υπολογίζεται αθροίζοντας το μήκος κάθε λέξης του κειμένου (είτε σε χαρακτήρες είτε σε συλλαβές) και διαιρώντας με το κειμενικό μέγεθος (το συνολικό αριθμό, δηλαδή, των λέξεων που απαρτίζουν το κείμενο). Εφόσον αυτός ο δείκτης λεξιλογικού πλούτου κρίθηκε στατιστικά σημαντικός και κατ' επέκταση διαφοροποιείται ανά επίπεδο γλωσσομάθειας, είναι σημαντικό να οριστεί το προβλεπόμενο μέσο μήκος λεξικού τύπου που θα αντιστοιχεί σε κάθε επίπεδο γλωσσομάθειας και το ποσοστό των μονάδων που θα καταλαμβάνει σε κάθε περίπτωση. Εναλλακτικά, προτείνεται, αφού καθοριστεί το ποσοστό των μονάδων που αντιστοιχούν στον συγκεκριμένο δείκτη από το σύνολο της βαθμολογίας του λεξιλογίου, να αντιστοιχούν συγκεκριμένες μονάδες σε συγκεκριμένα μέσα μήκη, με το ανώτερο ποσοστό να αντιστοιχεί σε κείμενα με μεγάλο μέσο μήκος λεξικού τύπου. Η αυτόματη απόδοση του ATL μέσω κάποιου εργαλείου ποσοτικοποίησης του γραπτού λόγου, όπως είναι το QUITA, που αξιοποιήθηκε στη συγκεκριμένη εργασία, θα διευκόλυνε πολύ τον υπολογισμό του.

5.2.1.2. Lambda (Λ)

Το Lambda (Λ) είναι ένας δείκτης ο οποίος σχετίζεται άμεσα με τον «πλούτο» του λεξιλογίου αλλά και με την επίδραση μεταξύ των γειτονικών συχνοτήτων. Για να καταστεί σαφής η έννοια της επίδρασης γειτονικών συχνοτήτων θα ήταν χρήσιμο να

⁴⁴ Σύμφωνα με τον Μικρό (2015), σε ορισμένες γλώσσες όπως τα Νέα Ελληνικά η αυτόματη αναγνώριση και μέτρηση των συλλαβών στις λέξεις είναι από μόνο του ξεχωριστό πρόβλημα Επεξεργασίας Φυσικής Γλώσσας. Στη Νέα Ελληνική υπάρχει το πρόβλημα της συνίζησης, της επιλεκτικής, δηλαδή, ένωσης του άτονου ι με το προηγούμενο γειτονικό του φωνήεν. Έτσι λέξεις όπως «λόγια» ή «χρόνια» μπορούν να έχουν δύο δυνατές συλλαβικές τεμαχίσεις π.χ.: α) «χρό-νια» (έχω 20 χρόνια να τον δω) β) «χρό-νι-α» (χρόνια ασθένεια). Το φαινόμενο αυτό είναι αρκετά συχνό και παράγει πολλά λάθη στον αυτόματο συλλαβισμό της Νέας Ελληνικής. Για το λόγο αυτό στα Νέα Ελληνικά και την έρευνα που θα παρουσιάσουμε εδώ, θα χρησιμοποιήσουμε ως βάση μέτρησης το χαρακτήρα.

επισημανθεί ότι κείμενα με τον ίδιο αριθμό λέξεων και λεξικών τύπων δεν παρουσιάζουν τις ίδιες συχνότητες εμφάνισης των λεξικών τους τύπων. Στην ουσία, πρόκειται για έναν δείκτη ο οποίος καταδεικνύει τον ιδιαίτερο τρόπο γραφής του κάθε συγγραφέα. Ένα από τα μεγαλύτερα πλεονεκτήματα του δείκτη Λ είναι η ανεξαρτησία του από το μέγεθος του κειμένου, σε αντίθεση με τους υπόλοιπους λεξιλογικούς δείκτες οι οποίοι δεν δίνουν αξιόπιστα αποτελέσματα αν τα υπό εξέταση κείμενα είναι διαφορετικού μεγέθους. (Popescu et al., 2011:10-12). Και ο συγκεκριμένος δείκτης λεξιλογικού «πλούτου» αναμένεται να ληφθεί υπόψη στη βαθμολόγηση του λεξιλογίου και να οριστεί συγκεκριμένο ποσοστό επί του συνολικού βαθμού. Επίσης, θα ήταν χρήσιμο συγκεκριμένες ποσοστιαίες μονάδες να αντιστοιχούν σε συγκεκριμένες τιμές του δείκτη Λ . Βέβαια, είναι αλήθεια ότι χωρίς το κατάλληλο εργαλείο κωδικοποίησης του κειμένου προκειμένου να εξαχθεί αυτόματα ο δείκτης Λ , όπως συνέβη εν προκειμένω με το εργαλείο QUITA, η διαδικασία υπολογισμού του προβλέπεται εξαιρετικά χρονοβόρα⁴⁵.

5.2.1.3. Ποσοστό των Άπαξ Λεγομένων (HL)

Πρόκειται για μια απλή αναλογία μεταξύ του αριθμού των λέξεων (N) και του αριθμού των άπαξ λεγομένων (*hapax legomena*) σε ένα κείμενο, δηλαδή των λέξεων που εμφανίζονται μόνο μία φορά σε αυτό. Ο υπολογισμός του συγκεκριμένου δείκτης, παρ' ότι φαίνεται πολύ απλός, δεν είναι, λόγω της επενέργειας του παράγοντα «μήκος κειμένου». Για να κριθεί ο δείκτης HL αξιόπιστος, είναι αναγκαίο να επενεργήσει κάποιος άλλος παράγοντας, όπως είναι ο Λ ⁴⁶, που θα εξαλείψει την επιβλαβή επίδραση του κειμενικού μεγέθους (Popescu, et al., 2009:31-34· Tuzzi, Popescu & Altmann, 2010· Popescu, Mačutek & Altmann, 2010). Και πάλι είναι αναγκαίο να καθοριστούν με ακρίβεια οι μονάδες που θα αντιστοιχούν σε συγκεκριμένα ποσοστά άπαξ λεγομένων στις ΠΓΛ των μαθητών/-τριών. Προκειμένου ο υπολογισμός του συγκεκριμένου δείκτη να καταστεί λιγότερο χρονοβόρος και περισσότερο αξιόπιστος προκρίνεται η αξιοποίηση ενός εργαλείου ποσοτικοποίησης του κειμένου, όπως είναι το QUITA.

⁴⁵ Βλ. τον τρόπο υπολογισμού του Λ στο Popescu et al., 2011:1-6.

⁴⁶ Βλ. τον τρόπο υπολογισμού του Λ στο Popescu et al., 2011:1.

5.3. Αποτίμηση των προτάσεων

Το γεγονός ότι το μέσο μήκος του λεξικού τύπου (ATL), το Lambda (Λ) και το Ποσοστό των Άπαξ Λεγομένων (HL), τρεις παράγοντες λεξιλογικού πλούτου στατιστικά σημαντικοί, δε λήφθηκαν υπόψη στην αξιολόγηση της λεξιλογικής ικανότητας των μαθητών/-τριών χρήζει προβληματισμού και περαιτέρω διερεύνησης. Η απεμπόληση των δεικτών αυτών πυροδοτεί μία συζήτηση γύρω από την αξιοπιστία των βαθμολογικών κριτηρίων των αξιολογητών και την εγκυρότητα των τελικών αποτελεσμάτων των προαγωγικών εξετάσεων για τους μαθητές της Γ2.

Ως φυσικό επακόλουθο, η απουσία των τριών δεικτών από την βαθμολόγηση πλήττει την αξιοπιστία του ισχύοντος πλαισίου αξιολόγησης του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας και καταδεικνύει τις αδυναμίες του ισχύοντος πλαισίου βαθμολόγησης του Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς για τη γλώσσα (ΚΕΠΑ). Δεν τίθεται αμφιβολία ότι το γενικευτικό πλαίσιο του ΚΕΠΑ δε δύναται να εκπληρώσει τις δεσμεύσεις του σχετικά με την ενίσχυση της διαφάνειας των μαθημάτων, των αναλυτικών προγραμμάτων διδασκαλίας και των τίτλων γλωσσομάθειας, με απώτερο σκοπό την προώθηση της διεθνούς συνεργασίας στον τομέα των σύγχρονων γλωσσών (ΚΕΠΑ, 2008:1).

Πέρα από την ανάγκη επανεξέτασης του τρόπου αξιολόγησης του λεξιλογίου από το Διδασκαλείο της Ελληνικής αλλά και από το ΚΕΠΑ, αναδεικνύεται η ανάγκη καθορισμού σαφέστερων αξιολογικών κριτηρίων για το λεξιλόγιο ανά Επίπεδο Γλωσσομάθειας, ανεξάρτητα από τους συγκεκριμένους φορείς. Ο σαφής καθορισμός των απαιτήσεων σε κάθε επίπεδο και η αντιστοίχιση των εξεταστικών κριτηρίων με ακριβείς μονάδες κρίνεται αναγκαία ώστε τα αποτελέσματα να μην είναι επισφαλής. Στη διαμόρφωση ενός αντικειμενικού και έγκριτου πλαισίου βαθμολόγησης των ΠΓΛ των μαθητών/-τριών θα συνέβαλλε καθοριστικά η δημιουργία ενός λογισμικού αυτόματης διόρθωσης του λεξιλογίου. Η αυτόματη απόδοση των τριών στατιστικά σημαντικών δεικτών μέσω του εργαλείου ποσοτικοποίησης QUITA ήδη ανέδειξε τη συμβολή των εργαλείων κωδικοποίησης του γραπτού λόγου στην απόδοση ακριβών αποτελεσμάτων. Βέβαια, η δημιουργία ενός τέτοιου εργαλείου που θα αφορά αποκλειστικά τη βαθμολόγηση του λεξιλογίου υπερβαίνει κατά πολύ τις δυνατότητες της συγκεκριμένης εργασίας.

ΚΕΦΑΛΑΙΟ 6: ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα έρευνα τέθηκε ως βασικός στόχος η διερεύνηση της πιθανής συσχέτισης έξι συγκεκριμένων δεικτών λεξιλογικού «πλούτου» με τα επίπεδα γλωσσομάθειας, όπως αυτά ορίζονται στο Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς (ΚΕΠΑ:2008). Πρόκειται για το Ποσοστό Άπαξ Λεγομένων ή Harax Legomenon Percentage (HL), το R1, το Lambda (Λ), το h-point (h), την Εντροπία ή Entropy (H) και το Μέσο Μήκος Λεξικού Τύπου ή Average Tokens length (ATL). Οι δείκτες λεξιλογικού «πλούτου» (lexical richness indices) ή εναλλακτικά δείκτες λεξιλογικής διαφοροποίησης (lexical diversity indices) αξιοποιούνται για να περιγραφεί πόσο διαφορετικό λεξιλόγιο επιστρατεύει το συγγραφικό υποκείμενο στα έργα του και, κατ' επέκταση, για να προσεγγισθεί ο βαθμός του λεξιλογικού «πλούτου» που το χαρακτηρίζει (Μικρός, 2015:78). Παρά το γεγονός ότι οι δείκτες αυτοί συνιστούν υφομετρικά χαρακτηριστικά και μέσω αυτών επιδιώκεται κατά βάση η απόδοση συγγραφικής πατρότητας σε έργα που είναι αγνώστου ή αμφισβητούμενου πνευματικού δημιουργού, η αξιοπιστία της μεθόδου είναι πλέον υψηλή και ήδη έχει αρχίσει να χρησιμοποιείται σε άλλους επιστημονικούς κλάδους όπως είναι η εγκληματολογική γλωσσολογία. Στη δεδομένη περίπτωση, στόχος δεν ήταν η απόδοση της συγγραφικής πατρότητας για ένα έργο, αλλά ο εντοπισμός των δεικτών λεξιλογικού «πλούτου» στα γραπτά αλλόγλωσσων μαθητών/-τριών της νέας ελληνικής, οι οποίοι/-ες επιχειρούν καθοδηγούμενη εκμάθηση, δηλαδή παρακολουθούν γλωσσικά μαθήματα ελληνικών στην Ελλάδα, και η συσχέτιση των συγκεκριμένων δεικτών με το επίπεδο γλωσσομάθειας στο οποίο βρίσκονται. Οι ΠΓΛ των μαθητών/-τριών τυπικά μπορούν να θεωρηθούν επιπέδου B2, αφού αποτελούν απαντήσεις σε εξεταστικά θέματα αυτού του επιπέδου και επιλέχθηκαν μόνο όσες είχαν συγκεντρώσει το απαιτούμενο ποσοστό (τουλάχιστον 60%) που καθιστά κάποιον/-α προβιβαστέο/-α. Παρά την τυπική κατάταξη των μαθητών/-τριών στο B2 επίπεδο, το πραγματικό επίπεδο ελληνομάθειας αυτών ποικίλλει και αυτό οφείλεται στη γλωσσική τους αφετηρία, αφού προέρχονται και από τα τρία επίπεδα γλωσσομάθειας: A (αρχάριο), B (μέσο) και Γ (προχωρημένο). Οι δείκτες λεξιλογικού «πλούτου» επιδιώχθηκε να συσχετιστούν με τα αρχικά επίπεδα κατάταξης και όχι με το κατ' επίφαση B2 επίπεδο στο οποίο επίσημα εντάσσονταν.

Χάρη στο εργαλείο ποσοτικοποίησης γραπτού λόγου QUITA και την επεξεργασία των δεδομένων που προέκυψαν από το λογισμικό Microsoft Excel 2010 και το στατιστικό πρόγραμμα SPSS, έγινε εμφανές ότι τα στατιστικά σημαντικά δείγματα σχετίζονται με τρεις από τους έξι δείκτες: το Μέσο Μήκος Λεξικού Τύπου (ATL), το Lambda (Λ) και το Ποσοστό Άπαξ Λεγομένων (HL). Ο λεξιλογικός δείκτης ATL εμφανίζει μεγαλύτερη διαφορά μεταξύ των Επιπέδων Α' και Γ' (sig.= 0), όπως και ο δείκτης Λ (sig.= 0,022). Μεγαλύτερη απόκλιση εντοπίζεται μεταξύ των Επιπέδων Α' και Γ' (sig.= 0,013) και στον δείκτη HL. Ο τελευταίος εμφανίζει στατιστικά σημαντικό αποτέλεσμα και μεταξύ των Επιπέδων Α' και Β' (sig.= 0,011), σε αντίθεση με τους δύο άλλους δείκτες. Οι τρεις συγκεκριμένοι δείκτες, κατ' αυτόν τον τρόπο, πιστοποιούν την απουσία ομοιογένειας στις ΠΓΛ των μαθητών/-τριών που προβιβάστηκαν στο Β2 Επίπεδο Γλωσσομάθειας, γεγονός που σχετίζεται άμεσα με την προηγούμενη γνώση που είχαν γύρω από την ελληνική γλώσσα, δηλαδή με το αρχικό επίπεδο ελληνομάθειας. Οι άλλοι τρεις δείκτες, το h-point, η εντροπία (H) και ο R1, δε φάνηκε να επηρεάζονται σημαντικά από την αρχική γνώση των μαθητών/-τριών.

Αφού αποδείχθηκε η συσχέτιση ορισμένων δεικτών λεξιλογικού «πλούτου» με τα επίπεδα γλωσσομάθειας, επιχειρήθηκε να διερευνηθεί αν οι υπό εξέταση δείκτες λήφθηκαν υπόψη στη βαθμολόγηση του λεξιλογίου στις ΠΓΛ των εξεταζομένων. Αξιοποιώντας την προβλεπτική εξίσωση πολλαπλής παλινδρόμησης (Multiple Regression Analysis- MRA), φάνηκε ότι το βαθμολογικό κριτήριο των διορθωτών/-τριών του λεξιλογίου των ΠΓΛ των μαθητών/-τριών δεν επηρεάστηκε από τους συγκεκριμένους δείκτες. Τα συγκεκριμένα ευρήματα αποτέλεσαν το έναυσμα για περαιτέρω διερεύνηση των βαθμολογικών κριτηρίων των ιθυνόντων και αναθεώρηση του υπάρχοντος θεωρητικού πλαισίου.

Όσον αφορά τον τρόπο βαθμολόγησης, λήφθηκε υπόψη το ισχύον πλαίσιο βαθμολόγησης του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας, εφόσον οι ΠΓΛ των μαθητών/-τριών συνιστούν γραπτές δοκιμασίες οι οποίες διενεργούνται υπό την αιγίδα του Διδασκαλείου αλλά και το πλαίσιο βαθμολόγησης του Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς για τη δεύτερη γλώσσα (ΚΕΠΑ), αφού οι μαθητές/-τριες είναι αλλόγλωσσοι/-ες και επιχειρούν καθοδηγούμενη εκμάθηση της ελληνικής. Και στις δύο περιπτώσεις υιοθετείται η αναλυτική αξιολόγηση, δηλαδή ο τελικός βαθμός είναι μια σύνθεση των ξεχωριστών εκτιμήσεων (Weir, 1990:63). Οι

διαφορετικές εκτιμήσεις, στην περίπτωση του ΣΕΜΚ, σχετίζονται με διαφορετικές δεξιότητες στο πλαίσιο της Παραγωγή Γραπτού Λόγου (ΠΓΛ) του αλλόγλωσσου μαθητικού κοινού και αφορούν το περιεχόμενο, τη γραμματική και το λεξιλόγιο. Από τον συνολικό βαθμό των είκοσι μονάδων, στο λεξιλόγιο αντιστοιχούν οι έξι μονάδες. Η επιλογή του συγκεκριμένου τύπου αξιολόγησης δε θα μπορούσε να αμφισβητηθεί, ειδικά αν αναλογιστεί κανείς τα μειονεκτήματα της ολιστικής αξιολόγησης σύμφωνα με την οποία ο γραπτός λόγος αντιμετωπίζεται ως ένα αδιάσπαστο σύνολο (O'Malley-Pierce, 1996:142).

Πράγματι, μέσω της αναλυτικής αξιολόγησης επιτρέπονται αξιόπιστες σύνθετες βαθμολογίες. Ωστόσο, τα στατιστικά αποτελέσματα που προέκυψαν από την παρούσα έρευνα καθιστούν αναγκαία την επανεξέταση των συγκεκριμένων βαθμολογικών πλαισίων. Στο μεν πλαίσιο του ΚΕΠΑ, διαπιστώνουμε μια σχετικιστική προσέγγιση αναφορικά με το λεξιλογικό φάσμα που καλείται να γνωρίζει ο/η εκπαιδευόμενος/-η. Μάλιστα, ορίζει ότι ο/η μαθητής/-τρια που εντάσσεται στο Επίπεδο Επάρκειας (B2) διαθέτει ικανοποιητικό φάσμα λεξιλογίου για θέματα που συνδέονται με τον τομέα του και τα περισσότερα γενικά θέματα, μπορεί να τροποποιεί τη διατύπωση για να αποφεύγει τις συχνές επαναλήψεις, αλλά τα λεξιλογικά κενά μπορεί ακόμα να προκαλούν διστακτικότητα και προσφυγή σε περιφράσεις (ΚΕΠΑ, 2008:130). Η αφαίρεση, το 2001, των ενδεικτικών λιστών λεξιλογίου που αντιστοιχούσαν σε κάθε επίπεδο γλωσσομάθειας προσέδωσε στο ΚΕΠΑ μία άμετρη ευελιξία, ευνοώντας φαινόμενα διδασκαλικής αυθαιρεσίας και ενισχύοντας την ανασφάλεια τόσο των μαθητών/-τριών όσο και των εκπαιδευτικών. Στο δε ισχύον πλαίσιο βαθμολόγησης του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας, παρ' ότι η βαθμολογία που αντιστοιχεί στο λεξιλόγιο ανέρχεται στις έξι (6) μονάδες, δεν υπάρχουν σαφή κριτήρια για την κατανομή των μονάδων αυτών.

Δεδομένων των ελλειμμάτων των ισχυόντων πλαισίων βαθμολόγησης αλλά και της απουσίας των τριών δεικτών λεξιλογικού «πλούτου» από την βαθμολόγηση, πλήττεται η αξιοπιστία του πλαισίου αξιολόγησης του Διδασκαλείου της Ελληνικής ως Δεύτερης Γλώσσας και του πλαισίου βαθμολόγησης του Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς για τη γλώσσα (ΚΕΠΑ). Παράλληλα, κρίνεται αναγκαία η συγκεκριμενοποίηση των βαθμολογικών κριτηρίων που αντιστοιχούν στο λεξιλόγιο ανά Επίπεδο Γλωσσομάθειας, ανεξάρτητα από τους συγκεκριμένους φορείς. Ο σαφής καθορισμός των λεξιλογικών απαιτήσεων σε κάθε επίπεδο και η αντιστοίχιση των

εξεταστικών κριτηρίων με ακριβείς μονάδες είναι σημαντική ώστε τα αποτελέσματα να μην είναι επισφαλής.

Οι τρεις στατιστικά σημαντικοί δείκτες κατέδειξαν με ενάργεια καίρια ελλείμματα στον ισχύοντα τρόπο βαθμολόγησης. Στόχος της εργασίας αυτής δεν είναι απλώς να συμπεριληφθούν οι συγκεκριμένοι δείκτες στη βαθμολόγηση των ΠΓΛ. Οι δείκτες αυτοί, άλλωστε, είναι απλώς ενδεικτικοί, αν αναλογιστεί κανείς τους πολυάριθμους παράγοντες λεξιλογικού «πλούτου» που έχουν κατά καιρούς προταθεί (Rudman, 1997:360). Για να καταστεί, βέβαια, δυνατός ο υπολογισμός τους και η συμπερίληψή τους στα βαθμολογικά κριτήρια, κρίνεται απαραίτητη η χρήση ενός εργαλείου ποσοτικοποίησης του κειμένου, όπως είναι το QUITA. Απώτερος στόχος είναι η συγκεκριμένη εργασία να αποτελέσει το εφαλτήριο για τη διαμόρφωση ενός αντικειμενικού και έγκριτου πλαισίου βαθμολόγησης των ΠΓΛ των μαθητών/-τριών. Για να καταστεί αυτό δυνατό, καθοριστική κρίνεται η δημιουργία ενός λογισμικού αυτόματης διόρθωσης του λεξιλογίου. Η αυτόματη απόδοση των τριών στατιστικά σημαντικών δεικτών μέσω του εργαλείου ποσοτικοποίησης QUITA, ήδη ανέδειξε τη συμβολή των εργαλείων κωδικοποίησης του γραπτού λόγου στην απόδοση ακριβών αποτελεσμάτων. Το βέβαιο είναι ότι η δημιουργία ενός τέτοιου εργαλείου που θα αφορά αποκλειστικά τη βαθμολόγηση του λεξιλογίου υπερβαίνει κατά πολύ τις δυνατότητες της συγκεκριμένης εργασίας. Θα μπορούσε, εντούτοις, να αποτελέσει πηγή έμπνευσης για ένα τέτοιο επίτευγμα το οποίο θα άλλαζε άρδην τις βαθμολογικές πρακτικές.

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΞΕΝΟΓΛΩΣΣΗ

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Bailey, R. W. (1979). Authorship attribution in a forensic context. In D. E. Alger, F. E. Knowles, & J. Smith (Eds.), *Advances in Computer-aided Literary and Linguistic Research. Proceedings of the Fifth International Symposium on Computers in Literary and Linguistic Research, 3-7 April 1978, Birmingham, UK* (pp. 1-15). Birmingham: AMLC University of Aston.

Beacco, J. C., & Byram, M. (2007). From linguistic diversity to plurilingual education. *Guide for the development of Language Education Policies in Europe Executive Version/J. Beacco, M. Byram, Council of Europe, 51.*

Bentz, C., Alikaniotis, D., Cysouw, M., & Ferrer-i-Cancho, R. (2017). The entropy of words—Learnability and expressivity across more than 1000 languages. *Entropy, 19*(6), 275.

Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL, 17*(1), 47-64.

Čech, R. (2011). Frequency structure of New Year's presidential speeches in Czech. The author shipanalysis. *Issues in Quantitative Linguistics, 2*, 82-94.

- Cohen, A. D. (1994). *Assessing Language ability in the classroom* (2nd ed.). Boston, MA: Heinle and Heinle.
- Cowan, R., Choi, H. E., & Kim, D. H. (2003). Four questions for error diagnosis and correction in CALL. *CalicoJournal*, 451-463.
- Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System*, 26(2), 163-174.
- Davies, A., & Elder, C. (Eds.) (2004). *The Handbook of Applied Linguistics*. Oxford: Blackwell.
- Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2nd ed). New York: John Wiley and Sons.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Granger, S., & Tribble, C. (1997). Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning. In S. Granger (Ed.), *Learner English on Computer* (pp. 199-209). London & New York: Addison Wesley Longman.
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3-18). London: Longman.
- Granger, S. (1999). Use of Tenses by Advanced EFL Learners: Evidence from an Error-tagged Computer Corpus. In H. Hasselgård, & S. Oksefjell (Eds), *Out of*

- Corpora - Studies in Honour of Stig Johansson* (pp 191-202). Amsterdam & Atlanta: Rodopi.
- Granger, S. (2002). A Bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3-33). Amsterdam/Philadelphia: John Benjamins.
- Granger, S. (2003a). Error-tagged Learner Corpus and CALL: A Promising Synergy. *CALICO Journal*, 20(3), 465-480.
- Granger, S. (2003b). The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL Quarterly: Special issue on Corpus Linguistics*, 37(3), 538-546.
- Granger, S. (2008). Learner corpora. In A. Lüdeling, & M. Kytö (Eds), *Corpus Linguistics: An International Handbook* (vol. 1, pp. 259-275). Walter de Gruyter.
- Granger, S., Gilquin, G., & Meunier, F. (Eds.). (2013). *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead: Proceedings of the First Learner Corpus Research Conference (LCR 2011)* (Vol. 1). Presses universitaires de Louvain.
- Grzybek, P. (2007). History and methodology of word length studies: The state of the art. In P. Grzybek (Ed.), *Contributions to the science of text and language: Word length studies and related issues* (pp. 15-90). Dordrecht: Springer.

- Hamp-Lyons, L., & Henning, G. (1991). Communicative writing profiles: An investigation of the transferability of a multiple-trait scoring instrument across ESL writing assessment contexts. *Language Learning*, 41(3), 337-373.
- Heyworth, F. (2006). The common European framework. *ELT Journal*, 60(2), 181-183.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, 102(46), 16569-16572.
- Hunston, S. (2002). Evaluation and organization in a sample of written academic discourse. In *Advances in written text analysis* (pp. 205-232). Routledge.
- James, C. (1998). *Errors in language learning and use. Exploring error analysis*. London & New York: Longman.
- Johns, T. (2002). Data-driven Learning: The Perpetual Challenge. In B. Kettemann, & G. Marko (Eds), *Teaching and Learning by Doing Corpus Analysis* (pp. 107-117). Amsterdam: Rodopi.
- Kaltenböck, G., & Mehlmauer-Larcher, B. (2005). Computer corpora and the language classroom: On the potential and limitations of computer corpora in language teaching. *ReCALL*, 17(1), 65-84.
- Kilgarriff, A., Charalabopoulou, F., Gavrilidou, M., Johannessen, J. B., Khalil, S., Kokkinakis, S. J., & Volodina, E. (2014). Corpus-based vocabulary lists for language learners for nine languages. *Language resources and evaluation*, 48(1), 121-163.

- Lee, H. K. (2004). A comparative study of ESL writers' performance in a paper-based and a computer-delivered writing test. *Assessing Writing*, 9(1), 4-26.
- Leech, G. (1997a). Introducing Corpus Annotation. In R. Garside, G. Leech, & T. McEnery (Eds), *Corpus Annotation: linguistic information from computer text corpora* (pp. 1-18). London and New York: Longman.
- Leech, G. (1997b). Grammatical tagging. In R. Garside, G. Leech, & T. McEnery (Eds), *Corpus Annotation: linguistic information from computer text corpora* (pp. 19-33). London & New York: Longman.
- Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on Computer* pp. 14-20). London & New York: Addison Wesley Longman.
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T., & Rayson, P.(1997). A Corpus/ Annotation Toolbox. In R. Garside, G. Leech, & T. McEnery (Eds), *Corpus Annotation: linguistic information from computer text corpora* (pp. 194-208). London & New York: Longman.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London & New York: Routledge.
- Meara, P. (1996). The dimensions of lexical competence. *Performance and competence in second language acquisition*, 35.
- Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In S. Granger, J. Hung, & S. Petch-Tyson (Eds), *Computer*

Learner Corpora, Second Language Acquisition and Foreign Language Teaching (pp. 119-141). Amsterdam/Philadelphia: John Benjamins.

Mikros, G., Tsakona, V., Drakopoulou, M., Koutra, A., Triantafylli, E., & Trypanagnostopoulou, S. (2009). Developing an EnglishGreek comparable corpus using web texts. In L. Krommyda, & I. Anagnostopoulos (Eds), *Second Meeting on Lexicographic Theory and Practice, 16 May 2008, Athens*. Athens: National and Kapodistrian University of Athens, Faculty of English Studies, Faculty of German Studies, 1-11. Retrieved from <http://users.uoa.gr/~gmikros/Publications/default.html> Προσπελάστηκε στις 25/07/2019

Milton, J. (1998). Exploiting L1 and interlanguage in the design of an electronic language learning and production environment. In S. Granger (Ed.), *Learner English on Computer* (pp. 186-198). London: Longman.

Milton, J., & Alexiou, T. (2009). Vocabulary size and the common European framework of reference for languages. In (Ed.), *Vocabulary studies in first and second language acquisition* (pp. 194-211). Palgrave Macmillan UK.

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, 211-232.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 Vocabulary*

Acquisition, Knowledge and Use: New Perspectives on Assessment and Corpus Analysis. European Second Language Association.

Morrow, K. (2004). *Insights from the Common European Framework*. Oxford: Oxford University Press.

Mukherjee, J. (2004). Bridging the gap between Applied Corpus Linguistics and the reality of English language teaching in Germany. In U. Connor, & T. Upton (Eds), *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 239-250). Amsterdam: Rodopi.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

O’Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly* 37(3), 389-418.

O’ Malley, J.M., & Pierce, L.V. (1996). *Authentic assessment for English language learners*. New-York: Addison-Wesley Publishing Company.

Pedhazur E.J. (1997). *Multiple regression in behavioral research. Explanation and prediction* (3rd ed.). Orlando FL: Harcourt Brace.

Popescu, I. I., Altmann, G., Grzybek, P., Jayaram, B. D., Köhler, R., Krupa, V., Mačutek, J., Pustet, R., Uhlířová, L., & Vidya, M. N. (2009a). *Word frequency studies*. Berlin-New York: Mouton de Gruyter.

Popescu, I. I., Mačutek, J., & Altmann, G. (2009b). *Aspects of word frequencies*. Lüdenscheid: RAM.

- Popescu, I. I., Mačutek, J., Altmann, G. (2010). Word forms, style and typology. *Glottotheory* 3(1), 89-96.
- Popescu, I. I., Čech, R., & Altmann, G. (2011). *The lambda-structure of texts*. Lüdenscheid: Ram-Verlag.
- Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal*, 26, 81-114.
- Renouf, A. (2007). Corpus development 25 years on: from super-corpus to cybercorpus. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years on* (pp. 27-49). Amsterdam: Rodopi.
- Römer, U. (2008). Corpora and language teaching. In A. Lüdeling, & M. Kytö (Eds), *Corpus Linguistics: An International Handbook* (Vol. 1, pp. 112-131). Berlin: Walter de Gruyter.
- Rudman, J. (1997). The state of authorship attribution studies: some problems and solutions. *Computers and the Humanities*, 31(4), 351-365.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing and use* (Vol. 9). John Benjamins Publishing.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language teaching research*, 12(3), 329-363.

- Sinclair, J. (1997). Corpus Evidence in Language Description. In A. Wichmann et al. (Eds.), *Teaching and Language Corpora* (pp. 27-39). London: Longman.
- Solan, L. M., & Tiersma, P. M. (2004). Author identification in american courts, *Applied Linguistics*, 25(4), 448-465.
- Spandel, V., & Stiggins, R.J. (1980). *Direct measures of writing skill: Issues and applications*. Portland, OR: Northwest Regional Educational Development Laboratory.
- Sylvie, D.C., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on Computer* (pp. 67-79). London & New York: Addison Wesley Longman.
- Tuzzi, A., Popescu, I.I., & Altmann, G. (2010). *Quantitative analysis of Italian texts*. Lüdenscheid: RAM.
- Weigle, S.C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- Weir, C. J. (1990). *Communicative language testing*. NJ: Prentice Hall Regents.
- Weisberg, S. (1985). *Applied Linear Regression* (2nd ed.). New York: John Wiley and Sons.
- Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21(1), 3-25.

Zagrebelsky, P., & Teresa, M. (2004). *Computer learner corpora: theoretical issues and empirical case studies of Italian advanced EFL learners' interlanguage*.
Alessandria: Edizioni dell' Orso.

ΕΛΛΗΝΟΓΛΩΣΣΗ

Αντωνοπούλου, Ν., Βαλετόπουλος, Φ., Καρακύργιου, Μ., Μουμτζή, Μ., & Παναγιωτίδου, Β. (2006). *Ανάλυση λαθών*. Θεσσαλονίκη: Κέντρο Ελληνικής Γλώσσας.

Βαρλοκώστα, Σ., & Τριανταφυλλίδου, Λ. (2011). *Επίπεδα γλωσσομάθειας στην Ελληνική ως δεύτερη γλώσσα*. Αθήνα.

Γούτσος, Δ. (2002). Η χρήση των ηλεκτρονικών σωμάτων κειμένων στην ανάλυση λόγου. Στο C. Clairis (Επιμ.), *Γλωσσολογικές έρευνες για την ελληνική Ι. Πρακτικά του 5ου Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας, Σορβόνη, Παρίσι, 13-15 Σεπτεμβρίου 2001* (σσ. 219-222). Paris: L' Harmattan.

Γούτσος, Δ. (2003). Σώμα Ελληνικών Κειμένων: Σχεδιασμός και υλοποίηση. Στα *Πρακτικά του 6ου Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας, Πανεπιστήμιο Κρήτης, 18-21 Σεπτεμβρίου 2003*. Ρέθυμνο: Πανεπιστήμιο Κρήτης.

Γούτσος, Δ., Σηφianού, Μ., & Γεωργακοπούλου, Α. (2005). *Η Ελληνική ως Ξένη Γλώσσα: από τις Λέξεις στα Κείμενα*. Αθήνα: Πατάκης.

Γούτσος, Δ. (2007). Βασικά τοπικά επιρρήματα σε ηλεκτρονικά σώματα κειμένων: προκαταρκτικές παρατηρήσεις. Στο Τομέας Γλωσσολογίας Πανεπιστήμιο Αθηνών (Επιμ.), *Γλωσσικός Περίπλους* (σσ. 36-46). Αθήνα: Ινστιτούτο του βιβλίου - Α. Καρδαμίτσα.

Γούτσος, Δ., & Κουτσουλέλου-Μίχου, Σ. (2009). Η διδασκαλία του ακαδημαϊκού λεξιλογίου στα Ελληνικά με τη χρήση σωμάτων κειμένων. *Πρακτικά του 3ου*

- Διεθνούς Συνεδρίου για τη Διδασκαλία της Νέας Ελληνικής ως Ξένης Γλώσσας, Πανεπιστήμιο Αθηνών, 22-23 Οκτωβρίου 2004 (σσ. 213-223). Αθήνα: Πανεπιστήμιο Αθηνών.
- Γούτσος, Δ. & Φραγκάκη, Γ. (2015). *Εισαγωγή στη γλωσσολογία σωμάτων κειμένων*. Αθήνα: Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα.
- Ηλιοπούλου, Κ. (2009). Διαπολιτισμική εκπαίδευση και γλωσσικά λάθη αλλόγλωσσων μαθητών: Ενδείκτες σχολικής αποτυχίας ή εργαλεία μάθησης και διδασκαλίας; Στο Ν. Παλαιολογού (Επιμ.), *Intercultural Education: Paideia, Polity, Demoi. Cd-rom Proceedings of the International Conference coorganized by the International Association for Intercultural Education (IAIE) and the Hellenic Migration Policy Institute (IMEPO), under the aegis of Unesco*. Athens.
- Θώμου, Π. (2001). Εκμάθηση Λεξιλογίου στην Ελληνική ως Ξένη Γλώσσα: Ανάλυση των μη αποδεκτών Λεξιλογικών Συνάψεων (Lexical Collocations) στο Γραπτό Λόγο Ξένων Φοιτητών. Στο Μ. Βάμβουκας, & Α. Χατζηδάκη (Επιμ.), *Πρακτικά Διεθνούς Συνεδρίου: "Μάθηση και Διδασκαλία της Ελληνικής ως μητρικής και ως δεύτερης γλώσσας"* (σσ. 99-109). Αθήνα: Ατραπός.
- Ιακώβου Μ. (2003). Επίπεδα Γλωσσομάθειας στη Διδασκαλία της Νέας Ελληνικής ως Ξένης Γλώσσας και Πιστοποίηση. Στο "Γλώσσα και Πολιτισμός": *Πρακτικά Συνεδρίου, Αθήνα 24-26 Μαΐου 2001* (σσ. 167-172). Αθήνα: Πανεπιστήμιο Αθηνών,.
- Ιακώβου, Μ. (2009). Εκμάθηση και Διδασκαλία του Λεξιλογίου στην Ελληνική ως Δεύτερη Γλώσσα. *Παρουσία*, 76, 1-210.

Ιακώβου, Μ., Μαρκόπουλος, Γ., & Μικρός, Γ. (2008). Καθορισμός βασικού λεξιλογίου μέσω Ηλεκτρονικών Σωμάτων Κειμένων για τη διδασκαλία της Νέας Ελληνικής ως Ξένης Γλώσσας. Στο Θ. Αγάθος, Β. Γιαννακού, Ζ. Δετσούδη, Σ. Δημητράκου, Α. Θώδα, Ν. Λαβίδας, Α. Λιβιέρη, Γ. Μωραΐτου, Θ. Νάτσης, Ε. Νουχουτίδου, Ν. Τασσοπούλου, & Α. Τσοτσορού (Επιμ.), *Πρακτικά του 2ου Διεθνούς Συνεδρίου για τη διδασκαλία της Ελληνικής ως Ξένης Γλώσσας, 27-28 Σεπτεμβρίου 2002 Αθήνα* (σσ. 161-171). Αθήνα: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.

Ιακώβου, Μ., Βασιλειάδη-Λιναρδάκη, Η., Βλάχου, Φ., Δήμα, Ο., Καββαδία, Μ., Κατσίνα, Τ., Κουτσομπού, Μ., Κύτρου, Σ.Ν., Κωστάκου, Χ., Παππά Φ. και Περρέα, Σ. (2017). ΣΕΠΑΜΕ2: Μια καινούρια πηγή αναφοράς για την ελληνική ως Γ2. Στο Th. Georgakopoulos, Th. Pavlidou, M. Pechlivanos, A. Alexiadou, J. Androutsopoulos, A. Kalokairinos, S. Skopeteas and K. Stathi (Eds) *Proceedings of the 12th International Conference on Greek Linguistics, 16-19 September 2015 Freie Universität Berlin, CEMOG. Vol. 1*. Berlin: Edition Romiosini, 419-432.

Κέντρο Ελληνικής Γλώσσας (2008). *Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς για τις γλώσσες. Εκμάθηση, διδασκαλία, αξιολόγηση*. Συμβούλιο της Ευρώπης. Θεσσαλονίκη: Ζήτη.

Κουλιάρη, Α., Αρχάκης, Α., & Τσάκωνα, Β. (2017). Ηλεκτρονικά Σώματα Μαθητικών Κειμένων: Προτάσεις Αξιοποίησής τους. *Δελτίο Επιστημονικής Ορολογίας και Νεολογισμών*, 14, 247-269.

Κόντα, Ε. (2012). Η απόδοση γένους στην ελληνική ως δεύτερη γλώσσα: ενδείξεις από παιδιά με μητρική γλώσσα την τουρκική. Στο *Μελέτες για την Ελληνική*

Γλώσσα. Πρακτικά της 32ης Ετήσιας Συνάντησης του Τομέα Γλωσσολογίας του Τμήματος Φιλολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, 6-8 Μαΐου 2011 (σσ. 387-399). Θεσσαλονίκη: Ινστιτούτο Νεοελληνικών Σπουδών.

Κοντός, Π., Ιακώβου, Μ., Μπέλλα, Σ., Μόζερ, Α., Χειλά-Μαρκοπούλου, Δ. (2002). *Αναλυτικό Πρόγραμμα Διδασκαλίας σε Ενηλίκους: Επίπεδο Επάρκειας*. Αθήνα: Πανεπιστήμιο Αθηνών.

Κουτσουλέλου, Σ., & Μικρός, Γ. (2005). Το “γεγονός” ως ουσιαστικό κέλυφος. Χρήση και λειτουργία σε ηλεκτρονικά σώματα κειμένων της ελληνικής. *Γλωσσολογία*, 16, 65-95. Διαθέσιμο στο

<http://glossologia.phil.uoa.gr/sites/default/files/Koutsoulelou%20&%20Mikros.pdf>

Προσπελάστηκε στις 27/07/2019

Μαγγανά, Α. (1999). Η κατάκτηση του πρωτικού χαρακτηριστικού της ΝΕ ως ξένης γλώσσας. Στο *Μελέτες για την Ελληνική Γλώσσα. Πρακτικά της 19ης Ετήσιας Συνάντησης του Τομέα Γλωσσολογίας του Τμήματος Φιλολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, 23-25 Μαΐου 1998* (σσ.189-198). Θεσσαλονίκη: Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.

Μικρός, Γ. (2007). Η σχέση του κειμενικού μεγέθους και του εύρους αντιπροσώπευσης των γλωσσικών χαρακτηριστικών. Τομέας Γλωσσολογίας Πανεπιστημίου Αθηνών (Επιμ.), Στο *Γλωσσικός Περίπλους: Μελέτες αφιερωμένες στη Δήμητρα Θεοφανοπούλου-Κοντού* (σσ. 190-199). Αθήνα: Α. Καρδαμίτσα.

Μικρός, Γ. (2015). *Υπολογιστική υφολογία*. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Ανακτήθηκε από: <http://hdl.handle.net/11419/4860> Προσπελάστηκε στις 20/09/2019

Μικρός, Γ., Μαρκόπουλος, Γ., και Ιακώβου, Μ. (2004). *Προδιαγραφές σχεδιασμού του ΗΣΚ γλωσσικής εκμάθησης. Αδημοσίευτο κείμενο Εργασίας*. Αθήνα: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.

Μικρός, Γ., & Φουντοπούλου, Μ. Ζ. (2002). Η αξιοποίηση των Ηλεκτρονικών Σωμάτων Κειμένων στη διδασκαλία της Ισπανικής ως ξένης γλώσσας. Στο Διδασκαλείο Ξένων Γλωσσών του Πανεπιστημίου Αθηνών (Επιμ.), *Πρακτικά: Η διδασκαλία της ξένης γλώσσας σε ενηλίκους: Εμπειρίες και προοπτικές* (σσ. 49-75). Αθήνα.

Μπέλλα, Σ. (2011). *Η Δεύτερη Γλώσσα. Κατάκτηση και διδασκαλία*. Αθήνα: Ελληνικά Γράμματα.

Παπαδοπούλου, Δ., & Τάντος, Α. (2014). Η διδακτική αξιοποίηση των σωμάτων κειμένων μαθητών στην ελληνική ως δεύτερη/ξένη γλώσσα. Στο Ζ. Γαβριηλίδου, & Α. Ρεβυθιάδου (Επιμ.), *Μελέτες Αφιερωμένες στην Ομότιμη Καθηγήτρια Α.Π.Θ. Άννα Αναστασιάδη-Συμεωνίδη* (σσ. 246-257). Καβάλα: Σαΐτα. Διαθέσιμο στο <http://www.diapolis.auth.gr/index.php/> Προσπελάστηκε στις 25/09/2019

Σαββίδου, Π. (2012). Το επίθημα -ιστικ(ος) στη νέα ελληνική: ανάλυση με τη χρήση ηλεκτρονικών σωμάτων κειμένων. Στο *Μελέτες για την Ελληνική Γλώσσα. Πρακτικά της 32ης Ετήσιας Συνάντησης του Τομέα Γλωσσολογίας του Τμήματος Φιλολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου*

Θεσσαλονίκης, 6-8 Μαΐου 2011 (σσ.416-427). Θεσσαλονίκη: Ινστιτούτο Νεοελληνικών Σπουδών.

Τάντος, Α., Αλεξανδρή, Κ., Δόση, Ι., Πούλιου, Κ., Σαββίδου, Π., & Φωτιάδου, Γ. (2015). Ανάλυση λαθών στο ελληνικό σώμα κειμένων μαθητών (ΕΣΚΕΙΜΑΘ): πρώτα ευρήματα. Στις Μελέτες για την Ελληνική Γλώσσα. *Πρακτικά της 35ης Ετήσιας Συνάντησης του Τομέα Γλωσσολογίας του Τμήματος Φιλολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, 9-10 Μαΐου 2014* (σσ. 719-732). Θεσσαλονίκη: Ινστιτούτο Νεοελληνικών Σπουδών.

Τζεβελέκου, Μ., Σταμούλη, Σ., Κάντζου, Β., Παπαγεωργακόπουλος, Γ., Χονδρογιάννη, Β., Βαρλοκόστα, Σ., Ιακώβου, Μ., & Λύτρα, Β. (2008). Η ελληνική ως δεύτερη γλώσσα: Ευρωπαϊκή κλίμακα αξιολόγησης των γλωσσών και κατάταξη σε επίπεδα γλωσσομάθειας. Στο Θ. Δραγώνα, & Α. Φραγκουδάκη (Επιμ.), *Πρόσθεση, όχι αφαίρεση, Πολλαπλασιασμός, όχι διαίρεση: Η μεταρρυθμιστική παρέμβαση στην εκπαίδευση της μειονότητας της Θράκης* (σσ. 155-174). Αθήνα: Μεταίχμιο.

Τζιμώκας, Δ. (2010). Ηλεκτρονικό σώμα κειμένων (ΗΣΚ) εκμάθησης της νέας ελληνικής ως δεύτερης γλώσσας: προς ένα ερευνητικό και διδακτικό εργαλείο. Στο *Μελέτες για την Ελληνική Γλώσσα. Πρακτικά της 30ης Ετήσιας Συνάντησης του Τομέα Γλωσσολογίας του Τμήματος Φιλολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, 2-3 Μαΐου 2009* (σσ. 602-616). Θεσσαλονίκη: Ινστιτούτο Νεοελληνικών Σπουδών,.

Τζιμώκας, Δ.(2014). Ηλεκτρονικό σώμα κειμένων (ΗΣΚ) εκμάθησης της ελληνικής ως δεύτερης γλώσσας: προκαταρκτικές παρατηρήσεις. Στο G. Kotzoglou, K.

Nikolou, E. Karantzola, K. Frantzi, I. Galantomos, M. Georgalidou, V. Kourti-Kazoullis, Ch. Papadopoulou, E. Vlachou (eds) *Selected Papers of the 11th International Conference on Greek Linguistics, 26-29 September 2013* (pp. 1679-1694). Rhodes: University of the Aegean.

Τζιμόκας, Δ. (2018). *Γραμματικά λάθη των μαθητών της νέας ελληνικής ως δεύτερης γλώσσας και η κωδικοποίησή τους βάσει ηλεκτρονικών σωμάτων κειμένων* (Αδημοσίευτη Διδακτορική Διατριβή). Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών/ Σχολή Φιλοσοφική/ Τμήμα Φιλολογίας, Αθήνα.

Τοκατλίδου, Β. (1986). *Εισαγωγή στη διδακτική των ζωντανών γλωσσών*. Αθήνα: Οδυσσέας.

Φλώρου, Α. (2009). *Τα ηλεκτρονικά σώματα κειμένων και η συμβολή τους στη μελέτη της ιταλικής γλώσσας ως ξένης* (Αδημοσίευτη Διδακτορική Διατριβή). Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Αθήνα.

Χαλισιάνη, Ι. (2008). Ηλεκτρονικά σώματα κειμένων και γλωσσική διδασκαλία: Διεθνείς αναζητήσεις και διαφαινόμενες προοπτικές για την ελληνική γλώσσα. Στο Συνέδριο «2008, Ευρωπαϊκό Έτος Διαπολιτισμικού Διαλόγου: συνομιλώντας με τις γλώσσες-πολιτισμούς», 12-14 Δεκεμβρίου 2008 (σσ. 818-830). Θεσσαλονίκη: Τμήμα Γαλλικής Γλώσσας και Φιλολογίας του Αριστετέλειου Πανεπιστημίου Θεσσαλονίκης.

ΙΣΤΟΣΕΛΙΔΕΣ

Σώμα Εξελικτικών Παραγωγών Μαθητών της Ελληνικής ως Γ2 (ΣΕΠΑΜΕ2). Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Φιλοσοφική Σχολή.
<http://sepame2.phil.uoa.gr/> Προσπελάστηκε στις 17/8/2019

Εργαλείο αναγνωσιμότητας για την ελληνική από το Κέντρο Ελληνικής Γλώσσας (ΚΕΓ):
<http://www.greek-language.gr/certification/readability/index.html>
Προσπελάστηκε στις 29.10.2019

Για τη ερμηνεία των όρων “synsemantic” και “autosemantic”:
<https://www.thefreedictionary.com> Προσπελάστηκε στις 17/10/2019

Η δομή εξέτασης για το Επίπεδο Επάρκειας (Vantage Level) αναρτημένη στην επίσημη ιστοσελίδα του Διδασκαλείου της Νέας Ελληνικής ως Ξένης Γλώσσας:
http://www.greekcourses.uoa.gr/fileadmin/greekcourses.uoa.gr/uploads/DOCUMENTS_DIAFORA/TEST_ELLHNOMATHEIAS.pdf Προσπελάστηκε στις 12/11/2019

Το εγχειρίδιο του εργαλείου ποσοτικοποίησης γραπτού λόγου QUITA:
https://www.dropbox.com/s/i1rp5pga5fycqhx/QUITA_Setup_1_11_0.msi?dl=0
Προσπελάστηκε στις 30.10.2019

Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς:
http://www.greekcourses.uoa.gr/fileadmin/greekcourses.uoa.gr/uploads/DOCUMENTS_DIAFORA/LEVELS_GR.pdf Προσπελάστηκε στις 11/09/2019

Τα ΗΣΜΚ που υπάρχουν παγκοσμίως καταγεγραμμένα στον διαδικτυακό κατάλογο του πανεπιστημίου της Louvain: <https://uclouvain.be/en-cecl-lcworld.html>

Προσπελάστηκε στις 10/07/2019

ΠΑΡΑΡΤΗΜΑ Α

Στους παρακάτω πίνακες αποτυπώνονται τα δεδομένα που εξήχθησαν από το εργαλείο ποσοτικοποίησης γραπτού λόγου QUITA. Ακολουθεί ένας πίνακας για κάθε επίπεδο γλωσσομάθειας.

Δεδομένα QUITA: Επίπεδο Α΄

Κωδικός ΠΓΛ	Types	Tokens	h-Point	Entropy	Average Tokens Length	R1	Λ (Lambda)	Hapax Percentage
29	153	271	5,5	6,747265	4,265683	0,85286	1,453793	0,405904
31	159	291	6,5	6,827377	4,725086	0,883591	1,425637	0,38488
ADA002	141	222	5	6,706525	4,5	0,862613	1,561811	0,5
ADA016	138	252	6,5	6,645288	4,75	0,873512	1,362915	0,380952
ARJ003	151	243	5,666667	6,783265	4,670782	0,864426	1,538426	0,477366
ARJ005	122	214	6	6,487953	4,705607	0,85514	1,404589	0,406542
ARJ007	131	227	6,5	6,587396	5,004405	0,859581	1,387958	0,400881
ARJ012	166	299	6,5	6,799145	4,70903	0,829849	1,433078	0,404682
ARJ013	112	200	5,5	6,414974	4,965	0,885625	1,341597	0,38
ARJ021	151	246	5,5	6,786641	4,861789	0,866362	1,532398	0,463415
CHA013	132	228	6	6,666628	4,557018	0,877193	1,393264	0,390351
CHA030	143	236	6	6,79303	4,724576	0,894068	1,469351	0,436441
CHA033	133	223	5,5	6,680704	4,838565	0,888453	1,429285	0,41704
CHA036	129	200	5	6,647296	4,855	0,8725	1,535586	0,48
CHA042	166	228	5	7,114624	5,166667	0,927632	1,732103	0,600877
CHA054	146	222	5	6,850285	4,86036	0,903153	1,580546	0,518018
CHA059	125	238	6	6,587209	4,659664	0,894958	1,275377	0,323529
CHA077	144	238	6	6,729015	4,806723	0,865546	1,492219	0,457983
CHA084	163	248	5	6,988417	4,923387	0,893145	1,62486	0,512097
CHA090	121	215	6	6,523768	4,251163	0,860465	1,355607	0,362791
CHA092	146	247	6	6,696899	4,935223	0,838057	1,460667	0,441296
CHA098	141	249	7	6,600628	4,678715	0,829317	1,406862	0,433735
chi063	184	285	6	7,148582	5,059649	0,898246	1,615458	0,508772
chi081	146	271	7	6,592065	4,734317	0,813653	1,378037	0,391144
COR005	155	245	5,5	6,895403	4,767347	0,902551	1,550105	0,485714
COR025	135	229	6	6,644248	4,283843	0,860262	1,445763	0,427948
COR044	176	289	6	6,849797	4,588235	0,816609	1,634285	0,50173
COR046	170	284	5	7,002534	4,894366	0,885563	1,523875	0,426056
COR047	155	255	6	6,819581	4,690196	0,866667	1,4963	0,47451
CRP009	129	212	5	6,626627	4,570755	0,879717	1,454434	0,448113
CRP016	161	265	7	6,918295	4,437736	0,888679	1,493393	0,460377
CRP020	138	221	5	6,760387	4,59276	0,893665	1,508411	0,457014
CRP024	155	258	6	6,841321	4,596899	0,879845	1,512129	0,445736
CRP027	171	296	6,5	6,992843	4,554054	0,888936	1,450447	0,412162
CRP033	134	282	7	6,514839	4,056738	0,828014	1,235767	0,294326

DIA006	166	282	6	6,910209	4,202128	0,87234	1,526482	0,439716	
DIA020	123	217	5	6,472917	4,423963	0,845622	1,385784	0,40553	
DRA001	156	258	5,5	6,8495	4,682171	0,872578	1,521477	0,44186	
DRA002	136	241	6	6,634878	4,477178	0,858921	1,391447	0,39834	
DRA003	144	218	5,5	6,785822	4,995413	0,881307	1,588168	0,53211	
EXJ007	140	200	5	6,718727	4,68	0,8525	1,707113	0,59	
EXJ023	144	233	5	6,808917	4,60515	0,899142	1,499087	0,459227	
EXJ026	125	186	5	6,559223	4,650538	0,862903	1,629505	0,553763	
EXJ043	151	229	5	6,852289	4,960699	0,884279	1,644131	0,519651	
EXJ047	130	213	5	6,575105	4,497653	0,856808	1,482121	0,464789	
EXJ048	168	247	5	7,046472	4,94332	0,892713	1,680024	0,538462	
EXJ051	126	206	6	6,505164	4,718447	0,849515	1,465554	0,485437	
FOT001	166	276	6	6,944721	4,48913	0,880435	1,519771	0,449275	
FOT012	178	284	6	7,060279	4,669014	0,883803	1,576309	0,478873	
FOT020	135	220	5	6,71799	4,531818	0,893182	1,484983	0,445455	
FOT059	148	260	6	6,766861	4,715385	0,884615	1,430393	0,415385	
NIK011	143	271	6	6,685173	4,276753	0,856089	1,346503	0,346863	
NIK012	175	274	6	7,014761	5,029197	0,864964	1,61264	0,507299	
NIK027	143	270	6,5	6,595748	4,362963	0,826389	1,364941	0,37037	
NIK048	145	270	6	6,67005	4,892593	0,851852	1,461865	0,355556	
PAP004	163	272	7	6,790358	4,547794	0,818015	1,562167	0,466912	
PAP012	164	272	6,5	6,90251	4,529412	0,886489	1,495967	0,459559	
PAP016	132	209	6	6,619154	4,555024	0,856459	1,509288	0,502392	
PAP017	169	300	7	6,886261	4,123333	0,858333	1,458602	0,423333	
PAR015	155	234	5,5	6,855062	4,837607	0,872329	1,638408	0,534188	
PAR023	140	224	5,5	6,750741	4,433036	0,897879	1,497458	0,473214	
PAR024	131	221	6	6,589229	4,108597	0,855204	1,457689	0,438914	
PAT001	129	190	5	6,665548	4,631579	0,886842	1,602424	0,547368	
PAT004	133	211	6	6,668335	4,464455	0,872038	1,490556	0,483412	
SOF030	156	260	6	6,84691	4,553846	0,884615	1,5047	0,461538	
SOF068	147	275	6	6,620326	4,374545	0,818182	1,369182	0,389091	
TXT002	113	219	6	6,339965	4,515982	0,844749	1,301591	0,3379	
TXT012	177	298	6	7,080062	4,801325	0,897351	1,510852	0,410596	
TXT020	155	240	6	6,76638	4,7	0,829167	1,62067	0,529167	
TXT021	162	264	6	6,907211	4,541667	0,863636	1,557042	0,450758	
TXT038	132	209	5,5	6,683573	4,382775	0,89055	1,527348	0,45933	
KPEM17	126	210	5,5	6,56406	4,871429	0,867262	1,441674	0,428571	
M.O.	10548	17592		5,802928	6,750538	4,643306	0,868933	1,49179	0,448976

Δεδομένα QUITA: Επίπεδο Β´

Κωδικός ΠΓΑ	Types	Tokens	h-Point	Entropy	Average Tokens Length	R1	Λ (Lambda)	Hapax Percentage
22	160	239	9	6,936644	4,785714	0,888655	1,671491	0,55042
35	176	292	9	7,008864	4,70068	0,87774	1,570989	0,442177
100	150	254	7,5	6,673455	4,70428	0,824903	1,528979	0,439689
ADA023	181	298	7,333333	6,993868	4,404682	0,853029	1,576877	0,461538
AMP003	126	187	7	6,675656	4,708995	0,896825	1,54052	0,507937
AMP004	124	205	7	6,532163	4,563107	0,874393	1,466588	0,446602
AMP015	119	180	7	6,589382	4,187845	0,900552	1,518193	0,497238
CHA001	176	236	7	7,122619	5	0,890947	1,781157	0,621399
CHA004	143	230	7	6,748167	4,428571	0,87395	1,478097	0,432773
CHA005	122	200	7	6,482002	4,872549	0,857843	1,41206	0,455882
CHA015	157	257	7	6,773542	4,784615	0,832692	1,514145	0,469231
CHA025	150	238	7	6,806612	4,619835	0,888946	1,525678	0,487603
CHA035	132	196	6,5	6,692269	4,582915	0,886935	1,562219	0,522613
CHA051	153	217	6,5	6,871919	5,213636	0,877841	1,691277	0,577273
CHA063	175	279	6,5	7,107801	4,629893	0,911032	1,54871	0,455516
CHA067	130	218	6,5	6,598322	4,446429	0,870536	1,388259	0,424107
CHA069	135	186	6,5	6,749836	4,836842	0,892105	1,656642	0,6
CHA074	134	215	6,5	6,730478	4,729358	0,904689	1,462148	0,444954
CHA076	150	221	6,5	6,841252	4,675556	0,877778	1,605163	0,542222
CHA085	116	196	6,333333	6,467129	5,132653	0,893495	1,378856	0,438776
CHA091	152	230	6,333333	6,883359	4,772532	0,89485	1,576228	0,519313
CHA097	137	183	6	6,827357	4,657895	0,907895	1,660915	0,594737
chi042	149	228	6	6,760683	5,017467	0,860808	1,613679	0,519651
chi074	123	191	6	6,547758	5,041885	0,875	1,552245	0,492147
COR006	129	202	6	6,713849	4,671642	0,912935	1,504607	0,467662
COR009	151	252	6	6,770391	4,570281	0,859438	1,548916	0,465863
COR017	173	248	6	7,146875	5,072874	0,917004	1,702218	0,562753
COR018	129	243	6	6,483557	4,523207	0,839662	1,366697	0,396624
COR019	113	188	6	6,439445	4,592593	0,87037	1,409982	0,42328
COR024	172	271	6	6,935092	4,567766	0,860806	1,64516	0,501832
COR026	105	190	6	6,228637	4,625668	0,820856	1,344948	0,390374
COR032	130	193	6	6,736474	4,864583	0,922526	1,573422	0,526042
COR039	128	231	6	6,563813	4,714286	0,865801	1,361099	0,380952
COR040	184	291	6	7,028512	4,578397	0,857578	1,666054	0,522648
COR041	157	268	6	6,869044	4,754717	0,875472	1,497572	0,422642
COR045	140	257	6	6,639989	4,410853	0,868702	1,353366	0,375969
CRP002	189	299	6	7,150269	5,050167	0,893395	1,637175	0,488294
CRP003	128	211	6	6,534396	4,848039	0,839325	1,519572	0,490196
CRP011	151	235	6	6,885804	4,592437	0,903887	1,539139	0,470588
CRP019	183	285	6	7,068016	4,77193	0,866667	1,639922	0,508772

CRP023	137	231	6	6,681074	4,782609	0,871739	1,456915	0,434783
CRP026	137	222	6	6,671387	4,797297	0,858108	1,515157	0,463964
CRP029	143	232	6	6,816515	4,534483	0,914332	1,485785	0,452586
CRP030	164	288	6	6,943618	5,090592	0,900697	1,494728	0,407666
CRP031	168	259	6	6,94962	4,582031	0,867676	1,613739	0,535156
CRP032	111	189	6	6,407248	4,507937	0,880952	1,364838	0,428571
CRP034	120	184	5,75	6,566072	4,798913	0,908288	1,505453	0,516304
CRP046	141	239	5,666667	6,696271	4,520833	0,875	1,446808	0,441667
DIA022	133	194	5,666667	6,684428	4,752577	0,882088	1,63084	0,56701
DRA021	161	265	5,666667	6,896542	4,74717	0,875472	1,500909	0,460377
DRA027	117	190	5,666667	6,436778	4,795812	0,84555	1,483603	0,450262
eag81	166	249	5,5	6,963342	4,88755	0,875725	1,672567	0,53012
eag93	158	260	5,5	6,829801	4,492308	0,857692	1,5158	0,469231
EXJ008	113	203	5,5	6,405505	4,348039	0,887255	1,30086	0,397059
EXJ020	130	208	5,5	6,59081	4,485577	0,85203	1,502052	0,475962
EXJ028	161	258	5,5	6,980284	4,786822	0,908915	1,566319	0,457364
EXJ030	132	217	5,5	6,613967	4,603687	0,866359	1,469379	0,460829
EXJ033	152	233	5,5	6,805306	5,11588	0,856223	1,624655	0,519313
EXJ038	151	256	5,5	6,776163	4,640625	0,88303	1,466426	0,457031
EXJ041	150	238	5,5	6,797803	4,844538	0,865546	1,550745	0,487395
EXJ042	133	193	5,5	6,656476	4,623037	0,880236	1,6434	0,612565
EXJ049	147	248	5,333333	6,672357	4,506073	0,838282	1,509166	0,445344
EXJ050	142	260	5,333333	6,443488	4,69112	0,778113	1,466414	0,413127
FOT078	168	249	5,333333	6,827219	4,996377	0,82971	1,608939	0,478261
FOT081	177	267	5	6,962943	4,945946	0,868666	1,533188	0,466216
FOT083	131	209	5	6,581123	4,64574	0,861236	1,435217	0,426009
FOT088	157	235	5	6,800682	4,408759	0,85219	1,440262	0,427007
GKA002	144	251	5	6,68116	4,645418	0,85259	1,441911	0,422311
GKA017	123	205	5	6,53465	4,642157	0,862745	1,437376	0,436275
ILM093	133	205	5	6,590183	4,834951	0,840885	1,547587	0,529126
ILM094	139	196	5	6,760331	4,828283	0,874369	1,668876	0,575758
KAV003	150	240	5	6,721279	4,761317	0,843621	1,551201	0,489712
NIK023	149	271	5	6,729846	4,340659	0,849817	1,402965	0,369963
NIK031	146	214	5	6,846074	4,700461	0,896313	1,596908	0,539171
NIK033	126	233	5	6,36122	4,776824	0,811695	1,417219	0,412017
NIK040	146	263	5	6,5724	4,92803	0,808712	1,394398	0,42803
PAP001	161	220	5	6,997357	4,612069	0,894397	1,669314	0,564655
PAP003	142	259	5	6,609742	4,669202	0,834601	1,37322	0,391635
PAP008	126	198	5	6,622827	4,407035	0,900126	1,486992	0,482412
PAR012	145	235	5	6,776092	4,597458	0,89036	1,542174	0,457627
PAR026	125	205	5	6,619977	5,07767	0,90534	1,434613	0,451456
pkr023.txt	174	254	5	7,033321	4,679688	0,873047	1,697115	0,5625
pkr059.txt	152	251	5	6,823087	4,829365	0,881173	1,486176	0,456349
TXT004	156	229	4,5	6,875701	4,549356	0,879828	1,616802	0,553648

TXT009	145	247	4,5	6,7354	4,550607	0,8583	1,459889	0,42915
TXT014	147	229	4,5	6,867878	5,017241	0,922953	1,51819	0,482759
TXT029	166	255	4	7,001948	4,823529	0,894608	1,600111	0,501961
TXT033	122	201	4	6,55562	4,5	0,883663	1,423399	0,445545
TXT039	131	215	4	6,552907	4,582222	0,846667	1,465672	0,422222
TXT041	144	223	4	6,696721	5,022422	0,848094	1,567821	0,524664
TXT046	120	203	4	6,553002	4,699507	0,889163	1,427514	0,403941
TXT048	163	262	4	6,897496	4,851145	0,870706	1,558522	0,477099
TXT049	145	227	4	6,815478	4,808696	0,896196	1,529552	0,469565
M.O.	13452	21507	5,727837	6,740931	4,710553	0,872229	1,522448	0,475545

Δεδομένα QUITA: Επίπεδο Γ'

Κωδικός ΠΓΑ	Types	Tokens	h-Point	Entropy	Average Tokens Length	R1	Λ (Lambda)	Hapax Percentage
14	163	273	8,5	6,952128	4,824176	0,901099	1,475205	0,446886
69	124	199	8	6,545927	4,537688	0,869975	1,488867	0,467337
070	171	298	6,5	6,970309	4,711921	0,271847	1,473098	0,39404
ADA022	177	288	7,5	6,985853	4,600694	0,871528	1,561411	0,493056
CHA012	161	261	7	6,980444	4,528736	0,913793	1,542431	0,452107
CHA041	119	217	7	6,341422	4,645161	0,827189	1,375953	0,414747
CHA056	162	258	7	6,94426	4,895349	0,883721	1,568903	0,472868
CHA099	162	292	7	6,774177	4,684932	0,833904	1,437999	0,417808
CHA100	164	284	7	6,901896	5,390845	0,883803	1,46247	0,429577
chi050	151	262	6,666667	6,766395	4,446565	0,866412	1,451522	0,423664
chi072	124	190	6,5	6,613505	4,7	0,900658	1,509968	0,505263
COR013	122	215	6,5	6,344812	4,967442	0,797674	1,437872	0,423256
COR023	159	263	6,5	6,865005	4,844106	0,874525	1,50324	0,467681
COR034	111	204	6	6,320991	4,357843	0,838235	1,333366	0,372549
COR035	130	197	6	6,707266	4,812183	0,909264	1,531414	0,497462
COR036	143	212	6	6,751002	4,70283	0,865566	1,648494	0,556604
COR037	186	273	6	7,107172	5,208791	0,864469	1,729444	0,56044
COR038	168	288	6	6,897583	4,618056	0,852431	1,483424	0,430556
COR043	172	276	6	7,062063	4,811594	0,905797	1,541608	0,471014
CRP006	145	212	6	6,849238	5,245283	0,889151	1,616443	0,54717
CRP014	142	246	6	6,597538	4,723577	0,833333	1,467291	0,455285
CRP015	172	250	6	7,074033	5,392	0,886	1,687741	0,556
CRP021	142	230	6	6,649272	4,552174	0,840353	1,58004	0,46087
DIA017	141	237	6	6,686291	4,417722	0,860759	1,444243	0,451477
DIA024	106	205	6	6,125052	4,439024	0,815244	1,397527	0,35122
eag59	154	244	6	6,835978	4,659836	0,877561	1,571059	0,487705
eag65	128	232	6	6,564432	4,491379	0,866379	1,337067	0,37931

EXJ031	189	265	6	7,187527	5,098113	0,887264	1,78892	0,6
EXJ035	139	200	6	6,782582	4,695	0,8825	1,635291	0,565
EXJ039	139	216	6	6,762336	5,203704	0,881944	1,605305	0,467593
EXJ044	145	221	5,666667	6,774286	4,936652	0,875566	1,600175	0,529412
EXJ053	174	263	5,666667	7,008648	4,920152	0,882605	1,678126	0,539924
FOT071	152	201	5,666667	7,033862	5,427861	0,930348	1,75874	0,626866
FOT072	154	252	5,666667	6,854716	4,761905	0,893353	1,50135	0,464286
FOT075	166	272	5,5	6,868157	4,654412	0,838235	1,547934	0,474265
GKA010	161	242	5,5	6,962945	4,942149	0,90932	1,612733	0,541322
KAV004	152	248	5,5	6,894752	4,689516	0,907258	1,485907	0,439516
KAV006	135	231	5,5	6,580304	4,402597	0,835498	1,465422	0,441558
PAR006	140	241	5,5	6,71159	4,522822	0,871369	1,414736	0,406639
PAR013	133	255	5	6,516404	4,266667	0,847549	1,327361	0,364706
PAT024	139	207	5	6,77677	4,845411	0,886473	1,606728	0,531401
pk005.txt	154	226	5	6,968221	4,641593	0,909292	1,632474	0,530973
pk036.txt	168	278	5	6,908141	4,640288	0,867106	1,560753	0,460432
SOF019	174	267	5	7,063691	4,494382	0,897004	1,657464	0,505618
SOF045	176	251	5	7,131507	5,175299	0,912849	1,736028	0,577689
SOF054	151	243	5	6,83091	4,765432	0,876543	1,522289	0,473251
SOF055	141	240	5	6,674416	4,775	0,858854	1,465522	0,433333
TXT007	135	211	5	6,54862	4,791469	0,812796	1,591523	0,511848
TXT013	145	200	5	6,857065	5,35	0,8875	1,736613	0,605
TXT016	167	253	5	6,978187	4,996047	0,885595	1,645635	0,529644
TXT017	146	239	5	6,810596	4,58159	0,887029	1,484116	0,447699
TXT022	129	183	5	6,746742	4,819672	0,920765	1,618245	0,562842
TXT024	105	206	5	6,233926	4,723301	0,839806	1,225022	0,330097
TXT032	129	229	5	6,612041	4,724891	0,884279	1,366928	0,393013
TXT044	140	199	4,5	6,799843	5,070352	0,88191	1,653506	0,577889
TXT047	162	252	4,5	6,8595	5,206349	0,857363	1,608751	0,52381
TXT050	180	300	6,666667	6,996486	4,910596	0,861663	1,53601	0,44702
KPEM05	145	217	4	6,83827	4,958525	0,90553	1,59053	0,534562
KPEM21	140	233	4	6,662802	4,832618	0,849785	1,52119	0,446352
M.O.	8804	14147	5,805085	6,787286	4,797191	0,863146	1,539651	0,479144

ΠΑΡΑΡΤΗΜΑ Β

Στον Πίνακα ANOVA καταδεικνύονται οι στατιστικά σημαντικούς δείκτες (sig.<0,05). Ως στατιστικά σημαντικοί δείκτες αναδείχθηκαν το Ποσοστό Άπαξ Λεγομένων (HL), το Lambda (Λ) και το Μέσο Μήκος Λέξης (ATL). Για τους τρεις αυτούς δείκτες διενεργήθηκε και Post Hoc Test για να διερευνηθούν περαιτέρω διαφορές μεταξύ των επιπέδων, οι πίνακες των οποίων παρατίθενται κάτωθι.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
ΗΡΟΙΝΤ	Between Groups	,352	2	,176	,252	,778
	Within Groups	154,809	221	,700		
	Total	155,161	223			
ΕΝΤΡΟΠΙΑ	Between Groups	,082	2	,041	1,071	,345
	Within Groups	8,441	221	,038		
	Total	8,522	223			
ΜΕΣΟΜΗΚΟΣ	Between Groups	,811	2	,405	7,334	,001
	Within Groups	12,216	221	,055		
	Total	13,026	223			
R1	Between Groups	,003	2	,002	,678	,509
	Within Groups	,510	221	,002		
	Total	,514	223			
LAMBDA	Between Groups	,083	2	,041	3,872	,022
	Within Groups	2,361	221	,011		
	Total	2,444	223			
ΑΠΑΞ	Between Groups	,041	2	,020	5,532	,005
	Within Groups	,811	221	,004		
	Total	,851	223			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: ΜΕΣΟΜΗΚΟΣ

Tukey HSD

(I) ΕΠΙΠΕΔΟ	(J) ΕΠΙΠΕΔΟ	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
ΕΠΙΠΕΔΟ Α	ΕΠΙΠΕΔΟ Β	-,07458	,03691	,110	-,1617	,0125
	ΕΠΙΠΕΔΟ Γ	-,15806*	,04129	,000	-,2555	-,0606
ΕΠΙΠΕΔΟ Β	ΕΠΙΠΕΔΟ Α	,07458	,03691	,110	-,0125	,1617
	ΕΠΙΠΕΔΟ Γ	-,08348	,03913	,085	-,1758	,0089
ΕΠΙΠΕΔΟ Γ	ΕΠΙΠΕΔΟ Α	,15806*	,04129	,000	,0606	,2555
	ΕΠΙΠΕΔΟ Β	,08348	,03913	,085	-,0089	,1758

*. The mean difference is significant at the 0.05 level.

Multiple Comparisons

Dependent Variable: LAMBDA

Tukey HSD

(I) ΕΠΙΠΕΔΟ	(J) ΕΠΙΠΕΔΟ	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
ΕΠΙΠΕΔΟ Α	ΕΠΙΠΕΔΟ Β	-,03266	,01623	,112	-,0709	,0056
	ΕΠΙΠΕΔΟ Γ	-,04854*	,01815	,022	-,0914	-,0057
ΕΠΙΠΕΔΟ Β	ΕΠΙΠΕΔΟ Α	,03266	,01623	,112	-,0056	,0709
	ΕΠΙΠΕΔΟ Γ	-,01588	,01720	,626	-,0565	,0247
ΕΠΙΠΕΔΟ Γ	ΕΠΙΠΕΔΟ Α	,04854*	,01815	,022	,0057	,0914
	ΕΠΙΠΕΔΟ Β	,01588	,01720	,626	-,0247	,0565

*. The mean difference is significant at the 0.05 level.

Multiple Comparisons

Dependent Variable: ΑΠΑΞ

Tukey HSD

(I) ΕΠΙΠΕΔΟ	(J) ΕΠΙΠΕΔΟ	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
ΕΠΙΠΕΔΟ Α	ΕΠΙΠΕΔΟ Β	-,02763*	,00951	,011	-,0501	-,0052
	ΕΠΙΠΕΔΟ Γ	-,03044*	,01064	,013	-,0555	-,0053
ΕΠΙΠΕΔΟ Β	ΕΠΙΠΕΔΟ Α	,02763*	,00951	,011	,0052	,0501
	ΕΠΙΠΕΔΟ Γ	-,00281	,01008	,958	-,0266	,0210
ΕΠΙΠΕΔΟ Γ	ΕΠΙΠΕΔΟ Α	,03044*	,01064	,013	,0053	,0555
	ΕΠΙΠΕΔΟ Β	,00281	,01008	,958	-,0210	,0266

*. The mean difference is significant at the 0.05 level.

ΠΑΡΑΡΤΗΜΑ Γ

Από την εφαρμογή της εξίσωσης πολλαπλής παλινδρόμησης (Multiple Regression Analysis- MRA) προκύπτουν τέσσερις πίνακες: Correlations, Model Summary, ANOVA και Coefficients. Οι τιμές που αποδεικνύουν τη συσχέτιση μεταξύ εξαρτημένης μεταβλητής (βαθμολογία γραπτών ανά επίπεδο γλωσσομάθειας) και ανεξάρτητων μεταβλητών (Ποσοστό Άπαξ Λεγομένων (HL), R1, Lambda (Λ), h-point, εντροπία, μέσο μήκος λεξικού τύπου) είναι το R2, το F και το significance (sig.) από τον πίνακα ANOVA και t και significance (sig.) από τον πίνακα Coefficients. Ακολουθούν οι συγκεκριμένοι πίνακες ανά επίπεδο γλωσσομάθειας.

ΕΠΙΠΕΔΟ Α

Correlations								
		ΒΑΘΜΟΣ	Η-ΡΟΙΝΤ	ΕΝΤΡΟΠΙΑ	ΜΕΣΟ ΜΗΚΟΣ	R1	LAMBDA (Λ)	ΑΠΑΞ
Pearson Correlation	ΒΑΘΜΟΣ	1,000	-,049	-,018	-,142	,041	,095	,123
	Η-ΡΟΙΝΤ	-,049	1,000	-,056	-,301	-,466	-,480	-,500
	ΕΝΤΡΟΠΙΑ	-,018	-,056	1,000	,314	,490	,651	,463
	ΜΕΣΟ ΜΗΚΟΣ	-,142	-,301	,314	1,000	,336	,429	,406
	R1	,041	-,466	,490	,336	1,000	,326	,302
	LAMBDA (Λ)	,095	-,480	,651	,429	,326	1,000	,936
	ΑΠΑΞ	,123	-,500	,463	,406	,302	,936	1,000
Sig. (1-tailed)	ΒΑΘΜΟΣ	.	,342	,439	,117	,365	,213	,151
	Η-ΡΟΙΝΤ	,342	.	,320	,005	,000	,000	,000
	ΕΝΤΡΟΠΙΑ	,439	,320	.	,004	,000	,000	,000
	ΜΕΣΟ ΜΗΚΟΣ	,117	,005	,004	.	,002	,000	,000
	R1	,365	,000	,000	,002	.	,003	,005
	LAMBDA (Λ)	,213	,000	,000	,000	,003	.	,000
	ΑΠΑΞ	,151	,000	,000	,000	,005	,000	.
N	ΒΑΘΜΟΣ	72	72	72	72	72	72	72
	Η-ΡΟΙΝΤ	72	72	72	72	72	72	72
	ΕΝΤΡΟΠΙΑ	72	72	72	72	72	72	72
	ΜΕΣΟ ΜΗΚΟΣ	72	72	72	72	72	72	72
	R1	72	72	72	72	72	72	72
	LAMBDA (Λ)	72	72	72	72	72	72	72
	ΑΠΑΞ	72	72	72	72	72	72	72

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,281 ^a	,079	-,006	102,51900

a. Predictors: (Constant), ΑΠΑΞ, R1, ΜΕΣΟ ΜΗΚΟΣ, ΕΝΤΡΟΠΙΑ, Η-POINT, LAMBDA (Λ)

b. Dependent Variable: ΒΑΘΜΟΣ

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	58404,618	6	9734,103	,926	,482 ^b
	Residual	683159,402	65	10510,145		
	Total	741564,020	71			

a. Dependent Variable: ΒΑΘΜΟΣ

b. Predictors: (Constant), ΑΠΑΞ, R1, ΜΕΣΟ ΜΗΚΟΣ, ΕΝΤΡΟΠΙΑ, Η-POINT, LAMBDA (Λ)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	176,552	631,950		,279	,781
	Η-POINT	24,569	33,496	,143	,733	,466
	ΕΝΤΡΟΠΙΑ	-169,037	164,386	-,284	-1,028	,308
	ΜΕΣΟ ΜΗΚΟΣ	-104,522	57,866	-,245	-1,806	,076
	R1	860,472	804,483	,203	1,070	,289
	LAMBDA (Λ)	387,565	581,656	,380	,666	,508
	ΑΠΑΞ	14,425	742,366	,009	,019	,985

a. Dependent Variable: ΒΑΘΜΟΣ

Β ΕΠΙΠΕΔΟ

Correlations

		ΒΑΘΜΟΣ	Η-POINT	ΕΝΤΡΟΠΙΑ	ΜΕΣΟ ΜΗΚΟΣ	ΑΠΑΞ	LAMBDA (Λ)	R1
Pearson Correlation	ΒΑΘΜΟΣ	1,000	,140	,184	,115	,074	,126	-,026
	Η-POINT	,140	1,000	,130	-,044	,112	,155	,104
	ΕΝΤΡΟΠΙΑ	,184	,130	1,000	,226	,487	,741	,404
	ΜΕΣΟ ΜΗΚΟΣ	,115	-,044	,226	1,000	,325	,362	,082
	ΑΠΑΞ	,074	,112	,487	,325	1,000	,889	,347
	LAMBDA (Λ)	,126	,155	,741	,362	,889	1,000	,270
	R1	-,026	,104	,404	,082	,347	,270	1,000
Sig. (1-tailed)	ΒΑΘΜΟΣ	.	,091	,039	,136	,240	,114	,404
	Η-POINT	,091	.	,107	,337	,142	,069	,159
	ΕΝΤΡΟΠΙΑ	,039	,107	.	,015	,000	,000	,000
	ΜΕΣΟ ΜΗΚΟΣ	,136	,337	,015	.	,001	,000	,217
	ΑΠΑΞ	,240	,142	,000	,001	.	,000	,000
	LAMBDA (Λ)	,114	,069	,000	,000	,000	.	,004
	R1	,404	,159	,000	,217	,000	,004	.
N	ΒΑΘΜΟΣ	93	93	93	93	93	93	93
	Η-POINT	93	93	93	93	93	93	93
	ΕΝΤΡΟΠΙΑ	93	93	93	93	93	93	93
	ΜΕΣΟ ΜΗΚΟΣ	93	93	93	93	93	93	93
	ΑΠΑΞ	93	93	93	93	93	93	93
	LAMBDA (Λ)	93	93	93	93	93	93	93
	R1	93	93	93	93	93	93	93

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,279 ^a	,078	,013	,80158

a. Predictors: (Constant), R1, ΜΕΣΟ ΜΗΚΟΣ, Η-POINT, ΕΝΤΡΟΠΙΑ, ΑΠΑΞ, LAMBDA (Λ)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4,659	6	,776	1,208	,310 ^b
	Residual	55,258	86	,643		
	Total	59,917	92			

a. Dependent Variable: ΒΑΘΜΟΣ

b. Predictors: (Constant), R1, ΜΕΣΟ ΜΗΚΟΣ, Η-POINT, ΕΝΤΡΟΠΙΑ, ΑΠΑΞ, LAMBDA (Λ)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1,115	4,008		-,278	,781
	H-POINT	,125	,090	,147	1,382	,171
	ΕΝΤΡΟΠΙΑ	1,599	,944	,382	1,695	,094
	ΜΕΣΟ ΜΗΚΟΣ	,445	,448	,111	,993	,324
	ΑΠΑΞ	3,473	4,639	,242	,749	,456
	LAMBDA (Λ)	-3,203	3,495	-,385	-,917	,362
	R1	-5,543	4,048	-,184	-1,369	,174

a. Dependent Variable: ΒΑΘΜΟΣ

Γ ΕΠΙΠΕΔΟ**Correlations**

		ΒΑΘΜΟΣ	H-POINT	ΕΝΤΡΟΠΙΑ	ΜΕΣΟ ΜΗΚΟΣ	R1	LAMBDA (Λ)	ΑΠΑΞ
Pearson Correlation	ΒΑΘΜΟΣ	1,000	,201	,146	,147	-,112	-,058	-,109
	H-POINT	,201	1,000	,012	-,118	-,121	-,196	-,225
	ΕΝΤΡΟΠΙΑ	,146	,012	1,000	,473	,127	,724	,617
	ΜΕΣΟ ΜΗΚΟΣ	,147	-,118	,473	1,000	,158	,632	,637
	R1	-,112	-,121	,127	,158	1,000	,242	,343
	LAMBDA (Λ)	-,058	-,196	,724	,632	,242	1,000	,943
	ΑΠΑΞ	-,109	-,225	,617	,637	,343	,943	1,000
Sig. (1-tailed)	ΒΑΘΜΟΣ	.	,063	,135	,134	,198	,331	,205
	H-POINT	,063	.	,465	,188	,180	,068	,044
	ΕΝΤΡΟΠΙΑ	,135	,465	.	,000	,169	,000	,000
	ΜΕΣΟ ΜΗΚΟΣ	,134	,188	,000	.	,116	,000	,000
	R1	,198	,180	,169	,116	.	,033	,004
	LAMBDA (Λ)	,331	,068	,000	,000	,033	.	,000
	ΑΠΑΞ	,205	,044	,000	,000	,004	,000	.
N	ΒΑΘΜΟΣ	59	59	59	59	59	59	59
	H-POINT	59	59	59	59	59	59	59
	ΕΝΤΡΟΠΙΑ	59	59	59	59	59	59	59
	ΜΕΣΟ ΜΗΚΟΣ	59	59	59	59	59	59	59
	R1	59	59	59	59	59	59	59
	LAMBDA (Λ)	59	59	59	59	59	59	59
	ΑΠΑΞ	59	59	59	59	59	59	59

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,406 ^a	,165	,068	,75037

a. Predictors: (Constant), ΑΠΑΞ, H-POINT, R1, ΜΕΣΟ ΜΗΚΟΣ, ΕΝΤΡΟΠΙΑ, LAMBDA (Λ)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5,774	6	,962	1,709	,137 ^b
	Residual	29,279	52	,563		
	Total	35,053	58			

a. Dependent Variable: ΒΑΘΜΟΣ

b. Predictors: (Constant), ΑΠΑΞ, Η-POINT, R1, ΜΕΣΟ ΜΗΚΟΣ, ΕΝΤΡΟΠΙΑ, LAMBDA (Λ)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3,496	3,793		-,922	,361
	Η-POINT	,114	,116	,131	,986	,329
	ΕΝΤΡΟΠΙΑ	1,004	,679	,290	1,479	,145
	ΜΕΣΟ ΜΗΚΟΣ	,907	,470	,321	1,930	,059
	R1	-,358	1,307	-,038	-,274	,786
	LAMBDA (Λ)	-,701	3,133	-,105	-,224	,824
	ΑΠΑΞ	-4,081	5,003	-,352	-,816	,418

a. Dependent Variable: ΒΑΘΜΟΣ