

ΕΛΕΝΗ ΠΟΛΙΤΟΥ ΜΑΡΜΑΡΙΝΟΥ - ΓΕΩΡΓΙΟΣ ΜΙΚΡΟΣ -
ΤΙΤΙΚΑ ΔΗΜΗΤΡΟΥΛΙΑ

ΕΦΑΡΜΟΓΗ ΥΦΟΜΕΤΡΙΚΩΝ ΤΕΧΝΙΚΩΝ
ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΠΑΤΡΟΤΗΤΑΣ ΚΕΙΜΕΝΟΥ:
ΠΡΩΤΟΤΥΠΑ ΕΡΓΑ ΚΑΙ ΜΕΤΑΦΡΑΣΕΙΣ ΤΟΥ ΠΑΠΑΔΙΑΜΑΝΤΗ

Πρόλογος

Τὸ 1991, ἤδη, ὅταν στὸ Α΄ Διεθνὲς Συνέδριο γιὰ τὸν Παπαδιαμάντη μὲ τὴν ἀνακοίνωσή μου «Παπαδιαμάντης, Μωπασσάν καὶ Τσέχωφ. Ἀπὸ τὴ Σκιαθὸ στὴν Εὐρώπη» (Πολίτου-Μαρμαρινού, 1996) ἔδειχνα τὸ ἐνδιαφέρον μου γιὰ τὶς μεταφράσεις τοῦ Παπαδιαμάντη ξένων λογοτεχνικῶν ἔργων, εἶχα ὑποβάλει στὸ Πανεπιστήμιο Ἀθηνῶν μιὰ σχετικὴ ἐρευνητικὴ πρόταση. Μὲ αὐτὴν ζητοῦσα ἓνα μικρὸ ποσό, προκειμένου νὰ συγκροτήσω ἐρευνητικὴ ὁμάδα εἰδικῶν πού θὰ ἐπιχειροῦσε νὰ ἐξακριβώσῃ μὲ τὴ βοήθεια μεθόδων τῆς Ὑπολογιστικῆς Ὑφολογίας ποιῆς ἀπὸ τὶς ἀνώνυμες μεταφράσεις λογοτεχνικῶν ἔργων τὶς δημοσιευμένες σὲ περιοδικὰ καὶ ἐφημερίδες, ὅπου ὁ Σκιαθίτης ἐργαζόταν ὡς μεταφραστὴς, ἦταν παπαδιαμαντικές. Ἡ πρόταση αὐτὴ τότε, δυστυχῶς, δὲν εἶχε ἐγκριθεῖ.

Μὲ τὴν ἀνακοίνωση τοῦ θέματος τοῦ τωρινοῦ Συνεδρίου ἡ παλιὰ ἐκείνη ἰδέα ξαναἦρθε, ἐπιτακτικὴ θὰ ἔλεγα, στὸ προσκῆνιο. Ἔτσι, ἀπευθύνθηκα στὴν κυρία Τιτίκα Δημητρούλια, Ἐπίκουρη Καθηγήτρια Θεωρίας καὶ Πράξης τῆς Μετάφρασης στὸ Πανεπιστήμιο Θεσσαλονίκης, καὶ στὸν κύριο Γεώργιο Μικρό, Ἀναπληρωτὴ Καθηγητὴ Ὑπολογιστικῆς καὶ Ποσοτικῆς Γλωσσολογίας στὸ Πανεπιστήμιο Ἀθηνῶν, τοὺς ἐξέθεσα τὴν πρότασή μου καὶ ζήτησα τὴ συνεργασία τους. Τοὺς εὐχαριστῶ καὶ ἀπὸ τὴ θέση αὐτὴ γιὰ τὴ θερμὴ ἀνταπόκριση καὶ τὴν ἀγαστὴ συνεργασία μας γιὰ περισσότερο ἀπὸ ἓνα χρόνο τώρα.

Εὐχαριστίες θὰ ἤθελα νὰ ἀπευθύνω, ἐκ μέρους καὶ τῶν τριῶν μας, στὶς ἐκδόσεις Δόμος, πού ἔθεσαν στὴ διάθεσή μας τὴν ψηφιακὴ μορφή τοῦ κειμένου τῶν Ἀπάντων Παπαδιαμάντη τῆς ἔκδοσης Ν. Δ. Τριανταφυλλόπουλου, ὅπως καὶ στὶς ἐκδόσεις Ἰνδικτος καὶ Κυριακίδη πού μᾶς

παραχώρησαν την ψηφιακή μορφή τῶν μεταφράσεων τοῦ Παπαδιαμάντη Ὁ Μαξιώτης καί Ὁ ἰατρός Ραμῶ ἀντίστοιχα.

Ἐλ. Π.-Μ.

1.Εἰσαγωγικά: Ἡ Ὑφομετρία, τὸ ὕφος καί ἡ ἔρευνά μας

Στόχος τῆς ἔρευνάς μας εἶναι νὰ ἐξακριβωθεῖ μὲ μεθόδους τῆς Ὑπολογιστικῆς Γλωσσολογίας καί εἰδικὰ τῆς Ὑπολογιστικῆς Ὑφολογίας, δηλαδή μὲ τὸν ἀντικειμενικὸ τρόπο πὺ ἐξασφαλίζουν μαθηματικοὶ ὑπολογισμοί, ποιὲς εἶναι παπαδιαμαντικὲς ἀπὸ τίς ἀνώνυμες μεταφράσεις ξένων λογοτεχνικῶν ἔργων τίς δημοσιευμένες σὲ ἐφημερίδες καί περιοδικά, ὅπου ὁ Παπαδιαμάντης ἐργαζόταν ὡς μεταφραστῆς.

Ἡ ἔρευνα ἔχει διακλαδικὸ χαρακτῆρα, καθὼς γιὰ τὴν πραγματοποίησή της συνεργάζονται ἡ Φιλολογία, ἡ Μεταφρασεολογία καί ἡ Ὑπολογιστικὴ Γλωσσολογία, καί ἰδιαίτερα ἡ Ὑφομετρία, κλάδος αὐτῆς τῆς τελευταίας. Ὁ κλάδος τῆς Ὑφομετρίας, ἐφαρμόζοντας μαθηματικὲς καί στατιστικὲς μεθόδους, ἐξάγει ποσοτικὲς μετρήσεις ἀπὸ ἓνα κείμενο. Ὡς κλάδος πῆρε γενναία ὠθηση μετὰ τὴν ἀνάπτυξη τῆς τεχνολογίας καί τῶν ὑπολογιστῶν πὺ μποροῦν νὰ ἐπεξεργάζονται τεράστια σώματα κειμένων, τὰ ὅποια θὰ ἦταν ἀδύνατον νὰ τιθασεύσει ὁ ἀνθρώπινος νοῦς. Στὴν ἔρευνα χρησιμοποιοῦνταν σώματα κειμένων, πὺ ἔγιναν ἀντικείμενο ἐπεξεργασίας μὲ μαθηματικὲς καί στατιστικὲς μεθόδους, οἱ ὅποιες ποσοτικοποιοῦν τὸ ὕφος καί μετροῦν τὰ χαρακτηριστικὰ του.

Τί εἶναι ὁμως «ὕφος» καί πῶς ὀρίζεται; Ὁ ὅρος περιβάλλεται ἀπὸ ἀμφισημία, γοητευτικὴ ὡστόσο. Φιλολόγοι, γλωσσολόγοι, κριτικοὶ τῆς λογοτεχνίας καί ψυχολόγοι τὸν χρησιμοποιοῦν μὲ προφανῶς διαφορετικὸ περιεχόμενο. Κοινὸς παρονομαστῆς, πάντως, ὄλων τῶν προσεγγίσεων εἶναι ἡ ἀντίληψη ὅτι τὸ ὕφος ἐξατομικεύει τὸν τρόπο μὲ τὸν ὅποιον ὁ καθένας χρησιμοποιεῖ τὴ γλῶσσα, ὅταν μιλάει ἢ ὅταν γράφει.

Γιὰ τοὺς φιλόλογους καί τοὺς κριτικούς τῆς λογοτεχνίας τὸ ὕφος εἶναι κατηγορία λογοτεχνική (τὸ ὕφος ἐνὸς λογοτεχνικοῦ κειμένου), καθὼς ὁμως γίνεται ἀντιληπτὸ μὲ ὑποκειμενικὸ καί ἐμπρεσιονιστικὸ τρόπο, οἱ ὑπολογιστικὲς παρατηρήσεις δὲν εἶναι δυνατὸν νὰ ἐπαληθευτοῦν μὲ ἀντικειμενικὲς μεθόδους. Ἐτσι, τὸ ὕφος ἐνὸς λογοτέχνη ἢ ἐνὸς λογοτεχνήματος περιγράφεται, συχνά, μὲ ὅρους μεταφορικούς (ὑψηλό, ταπεινό, περίτεχνο, λόγιο, διαυγές, σκοτεινό, κ.ἄ.).

Ἀκόμη, ἀπὸ τῆ φιλολογία καὶ τὴν κριτικὴ τὸ ὕφος προσδιορίζεται μὲ δεῖκτες ποὺ ἀναζητοῦνται μὲν στὸ κείμενο (π.χ. λέξεις/σημασίες, συντακτικὲς δομὲς κ.ἄ.) ἀλλὰ ἡ παρουσία τους ἐρμηνεύεται, συνήθως, μὲ βάση ἐξωκειμενικὰ δεδομένα ποὺ ἀφοροῦν τὸν λογοτέχνη ὡς πρόσωπο (βιωματικὲς ἐμπειρίες τοῦ συγγραφέα, γνώσεις του, ἰδεολογικὲς στάσεις του, κ.ἄ.). Π.χ. ἡ ρυθμικότητα τμημάτων (φράσεων, περιόδων) τοῦ πεζοῦ λόγου τοῦ Παπαδιαμάντη, χαρακτηριστικὴ πλευρὰ τοῦ ὕφους του, συνδέεται μὲ τὴν ἐξοικείωση τοῦ ἀνθρώπου Παπαδιαμάντη μὲ τὴ λειτουργικὴ ὕμνογραφία καὶ τὴν ψαλτικὴ.

Τέλος, τὰ κειμενικὰ ὑφολογικὰ στοιχεῖα ποὺ προέχουν γιὰ ἕναν φιλόλογο/κριτικὸ ἀντανακλοῦν συνειδητὲς ὑφολογικὲς ἀποφάσεις τοῦ συγγραφέα, εἶναι δηλαδὴ ἀποτέλεσμα ἐπιλογῆς ἀνάμεσα σὲ λεξικὲς, συντακτικὲς ἢ ἄλλες ἐναλλακτικὲς δυνατότητες ποὺ παρέχει ὁ γενικὸς γλωσσικὸς κώδικας στὸν ὁποῖον γράφει τὸ ἔργο του, στὴν περίπτωσή μας ἡ ἑλληνικὴ γλῶσσα. Ὅμως, οἱ συνειδητὲς αὐτὲς ὑφολογικὲς ἐπιλογὲς μποροῦν ἄνετα νὰ γίνουν ἀντικείμενο μίμησης, σκόπιμης ἢ ὄχι (Craig, 2004), νὰ γράψει κάποιος π.χ. *à la manière de* Παπαδιαμάντη. Ἡ δυνατότητα ἢ ἡ πιθανότητα αὐτὴ ἀποτελεῖ πρόβλημα γιὰ τὶς μελέτες ποὺ ἀποβλέπουν, εἰδικά, στὴν ἀνίχνευση τῆς πατρότητας ἀνώνυμων κειμένων.

Σὲ ἀντιδιαστολὴ μὲ τὰ παραπάνω, γιὰ τὴν Ὑφομετρία τὸ ὕφος εἶναι κατηγορία γλωσσικὴ, χαρακτηρίζει κάθε εἶδος λόγου, γραπτοῦ ἢ προφορικοῦ, καὶ ὄχι μόνον τὸν λογοτεχνικό. Ἡ Γνωσιακὴ Ὑφολογία, μάλιστα, κλάδος τῆς Γνωσιακῆς Ψυχολογίας, ὑποστηρίζει πὼς τὰ λογοτεχνικὰ κείμενα δὲν ἔχουν ὕφος καὶ διατείνεται πὼς τὸ ὕφος ἀνήκει στὸν ἀτομικὸ νοῦ, ὅταν αὐτός, κατὰ τὴ διάρκεια τῆς νοητικῆς γλωσσικῆς διαδικασίας καὶ κάτω ἀπὸ τὴν ἐπίδραση νευρικῶν περιορισμῶν, ἐκφέρει λόγο (Lancashire, 2004).

Γιὰ τὴν Ὑφομετρία, λοιπόν, τὸ ὕφος τείνει νὰ διακρίνει τὴ γραφὴ κάθε ἀτόμου, ἄρα καὶ τοῦ συγγραφέα, ἀπὸ ὅλες τὶς ἄλλες. Προσδιορίζεται ἀπὸ μορφὲς τῶν βασικῶν ἐπιπέδων τῆς γλώσσας (φωνητικὸ, φωνολογικὸ, μορφολογικὸ, συντακτικὸ, σημασιολογικὸ), οἱ ὁποῖες συνδέονται μὲ τὴ διαδικασία γλωσσικῆς παραγωγῆς καὶ βρίσκονται ἔξω ἀπὸ τὸν συνειδητὸ ἔλεγχο τοῦ ὁμιλητῆ/συγγραφέα. Ὅπως ἔχουν δεῖξει ἐρευνες ψυχολόγων καὶ ψυχολογολόγων, μέγα μέρος τῆς γλωσσικῆς παραγωγῆς πραγματοποιεῖται ἀπὸ τμήματα τοῦ ἐγκεφάλου ποὺ ἐνεργοῦν μὲ ἄμεσο καὶ σύνθετο τρόπο. Μορφὲς βασικῶν γλωσσικῶν ἐπιπέδων, προερχόμενες ἀπὸ ἕνα

«γλωσσικό άσυνείδητο», εμφανίζονται άθρομήςτως στον προσωπικό λόγο του καθενός, με την επίδραση ενός είδους συνειρμικής μνήμης. Η Ύφομετρία στοχεύει άκριβώς στο να έντοπίσει στη γλώσσα ενός όποιουδήποτε λόγου αυτές τις μορφές, τις δομές και τὰ σχήματα που συνδέονται με τή διαδικασία γλωσσικής παραγωγής και γραφής, άρα και με τὸ «ύφος» ενός κειμένου υπό εύρύτατη έννοια. Η ύπαρξη τών σχημάτων αυτών, που είναι μεν άόρατα στον άναγνώστη αλλά χαρακτηρίζουν τὸ ύφος ενός κειμένου με τή συχνότητα, τούς συνδυασμούς και τή διασπορά τους, δέν είναι άποδείξιμη χωρίς ύπολογιστικές και στατιστικές μεθόδους (Craig, 2004b). Οί γλωσσικές αυτές μορφές και οί δομές τους, που άποτελοϋν άκατέργαστο πρωτογενές γλωσσικό ύλικό, μπορεί να είναι άδιάφορες για τόν κριτικό και τόν φιλόλογο, για τήν Ύφομετρία όμως άποτελοϋν άσυνείδητους μεν αλλά ύπαρκτους και μετρήσιμους δείκτες προσωπικού και άτομικού ύφους, συνιστοϋν δείκτες-κλειδιά τής ιδιολέκτου του καθενός, έπομένως και του συγγραφέα.

Τὸ έτσι νοούμενο ύφος ενός κειμένου, καθώς αναλύεται σε μετρήσιμα ποσοτικά στοιχεΐα, μπορεί να μελετηθεΐ και να περιγραφεί με τρόπο άντικειμενικό. Ο προσδιορισμός του δέν έπαφίεται μόνον στην έξοικείωση, τήν έξυδέρκεια και γενικά στις όπωςδήποτε ύποκειμενικές έντυπώσεις του άναγνώστη/κριτικού, όσο έπαρκής κι άν είναι αυτός. Οί ταπεινοί αυτοί —για να χρησιμοποιήσουμε και έμεις μία μεταφορά όρίζοντάς τους— ύφολογικοί δείκτες, παρακολουθοϋν τόν συγγραφέα σε όλη τή ζωή του σαν ένα είδος «ύφολογικού δακτυλικού άποτυπώματος», όσο κι άν ή Ύφομετρία, λόγω έλλειψης, άκόμη, άρκετών έρευνητικών δεδομένων, δέν είναι σε θέση παρά μόνον να καταγράψει «άντικειμενικά μετρήσιμα στοιχεΐα ύφους που είναι μοναδικά σε συγκεκριμένο συγγραφέα, με τήν έννοια ότι εμφανίζονται σε όλα τὰ γραπτά του και σε κανέναν άλλον από όσους έχουν μελετηθεΐ ως σήμερα» (Kenny, 1982).

Η Ύφομετρία με τις στατιστικές μεθόδους της δέν μπορεί να προχωρήσει σε βάθος αυτό που συνήθως λέγεται «άνάλυση ενός λογοτεχνήματος». Μπορεί, όμως, να επικουρήσει σε αυτήν, προσφέροντας στον μελετητή τής λογοτεχνίας μία ποσοτική και στατιστική άνάλυση, μετατρέποντας σε αριθμούς και πίνακες κειμενικά δεδομένα. Δημιουργεί, περίπου, ένα άλλο κείμενο, τὸ όποιο, όμως, συνδέεται με συστηματικό και έπαναλαμβανόμενο τρόπο με τὸ πρώτο κείμενο και μπορεί να είναι άποκαλυπτικό

γι' αυτό. Όπως έχει παρατηρηθεί, αποτελεί μόνον «πρελούδιο τῆς ἐρμη-
νείας» (Craig, 2004c).

Ἡ Ὑφομετρία εἶναι βοηθητικὴ γιὰ διάφορους τρόπους μελέτης τῆς λο-
γοτεχνίας, ὅπως εἶναι ὁ προσδιορισμὸς τοῦ χρόνου συγγραφῆς ἑνὸς ἔργου,
ὁ προσδιορισμὸς τῆς σχέσης ὕφους καὶ λογοτεχνικοῦ εἴδους ἢ ὕφους καὶ θέ-
ματος. Ἰδιαιτέρα ἀποτελεσματικὴ ὅμως ἀποδείχτηκε στὴν ἀναγνώριση
τῆς πατρότητας ἀνώνυμων ἢ ἀμφισβητούμενης πατρότητας ἔργων ἢ γε-
νικῶς κειμένων, π.χ. σημειωμάτων ἀνθρώπων ποὺ ἀναγγέλλουν τὴν αὐτο-
κτονία τους ἢ διαθηκῶν. Αὐτὴ ἀκριβῶς ἢ συμβολῆ τῆς ἐνδιαφέρει ἐμᾶς
ἐδῶ, στὴν ἐξακρίβωση, δηλαδή, ποιὲς ἀπὸ τίς ἀνώνυμες μεταφράσεις εἶναι
(ἢ δὲν εἶναι) παπαδιαμαντικές.

Ἐπιλέξαμε νὰ ἐλέγξουμε τίς παρακάτω ἑξι ἀνώνυμες μεταφράσεις:

1. Γουὸ δὲ Μωπασσάν, «Ὁ δασοφύλαξ», *Ἐφημερίς*, 20.12.1886-22.12.1886.
2. Ἰβάν Τουργκιένεφ, «Ὁ ἐπαρχιακὸς ἰατρός», *Ἐφημερίς*, 23.12.1886-
24.12.1886.
3. Φραγκίσκου Κοππέ, «Ἀρριέτα», *Ἐφημερίς*, 5.8.1889-19.8.1889.
4. Νικολάου Γόγολ, «Ἡ τρομερὰ ἐκδίκησης». (Ρωσικὸν μυθιστόρημα), *Τὸ
Ἄστυ*, 3.3.1899-15.3.1899.
5. Ἐδγάρδου Πόε, «Τὰ διηγήματα τῆς πανώλους. Τὸ προσωπεῖον τοῦ Ἐρυ-
θροῦ Θανάτου», *Τὸ Ἄστυ*, 7.9.1899-8.9.1899.
6. Νικολάου Γόγολ, «Ἡ μύτη», *Τὸ Ἄστυ*, 25.11.1899-2.12.1899.

Κριτήριον γιὰ τὴν ἐπιλογή αὐτὴ ἦταν ὅτι, καθὼς —κατὰ τὴ δική μας
ἀποψη τουλάχιστον— κέντρο τοῦ φιλολογικοῦ ἐνδιαφέροντος ἀποτελεῖ τὸ
πρωτότυπο ἔργο τοῦ Παπαδιαμάντη, ὅποιαδήποτε ἔρευνα ποὺ θὰ φέρνει
στὸ φῶς τὴ γνωριμία του (στὴν περίπτωσή μας μάλιστα μέσῳ μεταφρά-
σεῶν του) μὲ ξένους λογοτέχνες καὶ ξένα λογοτεχνικὰ ἔργα ἀνοίγει τὸν
δρόμον γιὰ γόνιμες συγκρίσεις τοῦ Σκιαθίτη μὲ Εὐρωπαίους πεζογράφους.
Οἱ ἐλάχιστες ὡς τώρα τέτοιες συγκριτικὲς φιλολογικὲς μελέτες ἀφενὸς
δείχνουν πῶς ὁ Παπαδιαμάντης δὲν κινδυνεύει στὸ ἐλάχιστον ἀπὸ τὴν ἀνα-
μέτρησή του μὲ μεγάλους ξένους ὁμοτέχνους του, τὸ ἀντίθετο μάλιστα,
καὶ ἀφετέρου ἀνατρέπουν τὴ διαστρεβλωτικὴ εἰκόνα ποὺ διαμόρφωσε μέρος
τῆς παλιότερης κριτικῆς, σύμφωνα μὲ τὴν ὁποία τὸ ἔργο του, μὲ τὸν το-
πικὸ, ἠθογραφικὸ καὶ κατὰ βάση θρησκευτικὸ χαρακτῆρα του, ἦταν πε-
ριορισμένου αἰσθητικοῦ βεληνεκοῦς — τὴν ἀνατροπὴ τῆς εἰκόνας αὐτῆς
ὑπαινίσσεται ὁ ὑπότιτλος τῆς πρώτης συγκριτικῆς ἐκείνης μελέτης «Πα-
παδιαμάντης, Μωπασσάν καὶ Τσέχωφ. Ἀπὸ τὴ Σκιάθου στὴν Εὐρώπη»

(Πολίτου-Μαρμαρινού, 1996). Αὐτὸς ἦταν ὁ λόγος ποὺ ἐπιλέξαμε νὰ ἐλέγξουμε μεταφράσεις ἔργων σημαντικῶν ξένων λογοτεχνῶν, ὅπως ὁ Μωπασσάν, ὁ Τουργκιένεφ, ὁ Γκόγκολ ἢ ὁ Πόε.

Τέλος, στὸ πρῶτο αὐτὸ δεῖγμα μεταφράσεων πρὸς ἔλεγχο συμπεριλάβαμε ἐπίτηδες τὸ διήγημα τοῦ Κοππὲ «Ἀρριέτα», τὴ μετάφραση τοῦ ὁποῖου ὁ Ν. Δ. Τριανταφυλλόπουλος καὶ ἡ Λ. Τριανταφυλλοπούλου (2006) μὲ τὰ δικά τους τεκμήρια θεωροῦν «ἀναμφισβήτητα παπαδιαμαντική», καὶ τὸ διήγημα τοῦ Μωπασσάν «Ὁ δασοφύλαξ», γιὰ τὴ μετάφραση τοῦ ὁποῖου ὁ Ν. Δ. Τριανταφυλλόπουλος καὶ ἡ Λ. Τριανταφυλλοπούλου, καὶ πάλι μὲ τὰ δικά τους κριτήρια, ἔχουν σημειώσει πῶς «κανένα στοιχεῖο δὲν μᾶς δίνει τὸ ἐνδόξιμο νὰ [τὴν] θεωρήσουμε παπαδιαμαντική» (Maupassant, 2003), προκειμένου νὰ ἐλέγξουμε τὴν ἀξιοπιστία τοῦ δικοῦ μας ἐργαλείου, σὲ σχέση μὲ δύο μεταφράσεις γιὰ τὴν πατρότητα τῶν ὁποίων οἱ δύο σημαντικότεροι καὶ παραγωγικότεροι μελετητὲς τοῦ μεταφραστῆ Παπαδιαμάντη εἶχαν ἤδη ἀποφανθεῖ.

2. Τὸ ὕφος τοῦ μεταφραστῆ

Στις μεταφραστικὲς σπουδές, ποὺ αὐτονομοῦνται σταδιακὰ ἀπὸ τὴ δεκαετία τοῦ '60 καὶ γνωρίζουν ἰδιαίτερη ἀνάπτυξη ἀπὸ τὸ 1970 καὶ ἐξῆς, τὴν τελευταία δεκαετία μόνον γίνεται συστηματικὰ λόγος γιὰ τὸ ὕφος τοῦ μεταφραστῆ. Παράλληλα, ἐπιχειρεῖται ὁ ἐντοπισμὸς του, πέραν τῆς ποιητικῆς ἀνάλυσης, μὲ ποσοτικὲς καὶ στατιστικὲς μεθόδους (Baker, 2000· Bosseaux 2004· Winters, 2004· Saldanha 2005 καὶ 2011· Rybicki, 2012). Τὸ γεγονός αὐτὸ συνδέεται μὲ τὴ συστηματικὴ, σταδιακὰ, χρησιμοποίηση τῶν σωμάτων κειμένων στις μεταφραστικὲς σπουδές (Baker, 1996 καὶ 1997· Laviosa, 1997· Tymozcko, 1998· Kenny, 2001· Olohan, 2004), σὲ συνάρτηση μὲ τὴν ἀνάπτυξη τῶν ἐργαλείων τῆς Ὑπολογιστικῆς Γλωσσολογίας καὶ εἰδικὰ τῆς Ἀνάλυσης Φυσικῆς Γλώσσας. Μὲ δεδομένη τὴ μακροαίωνη ἀφάνεια τοῦ μεταφραστῆ (Venuti, 1995), περισσότερο ἢ λιγότερο κάθε φορὰ ἐπιβεβλημένη, ἡ ἐστίαση, μὲ τὴν χρησιμοποίηση ὕφομετρικῶν μεθόδων, στὴν ἰδιαιτερότητα καὶ τὸ ὕφος τοῦ μεταφραστῆ δημιουργεῖ ἓνα μεῖζον πεδίο ἔρευνας γιὰ τις μεταφραστικὲς σπουδές. Ἐξ οὗ καὶ τὸ μεγάλο ἐνδιαφέρον ποὺ παρουσιάζει τὸ παρόν, διεπιστημονικὸ ἐγχειρημα, καθὼς ἀφορᾷ τὴν ἰδιαιτερότητα καὶ τὴν ἀναγνωρισιμότητα τοῦ ὕφους τοῦ μεταφραστῆ, ὡς «δακτυλικοῦ ἀποτυπώματός» του (Baker,

2000: 245), ως «άλλης φωνής» διακριτής μέσα στο κείμενο (Hermans, 1996: 9, τὸ παραθέτει ἢ Baker, *ibid.*). Δὲν ξεχνᾶμε, βεβαίως, ποτὲ αὐτὸ ποὺ ὑπογραμμίζει ὁ Rubicki, ὅτι στὸ πεδίο αὐτὸ ὅλα τὰ ζητήματα παραμένουν ἀνοιχτὰ ἢ κλείνουν ἀπλῶς γιὰ νὰ ξανανοίξουν μετὰ ἀπὸ λίγο (δ.π., 231).

Ἡ Baker ὀρίζει τὸ ὕφος τοῦ μεταφραστῆ «ὡς ἓνα εἶδος ἀποτυπώματος ποὺ ἐκφράζεται μὲ μία σειρά γλωσσικά —καὶ μὴ γλωσσικά— στοιχεῖα» (2000: 245), ἀπὸ τίς ἐπιλογές κειμένων τοῦ μεταφραστῆ καὶ τίς στρατηγικὲς ποὺ ἐφαρμόζει ὡς τῆ «χαρακτηριστικὴ χρῆση τῆς γλώσσας, τὸ προσωπικὸ του/της προφίλ γλωσσικῶν συνηθειῶν, σὲ σύγκριση μὲ ἄλλους μεταφραστὲς» (*ibid.*). Ἐστιάζει σὲ μοτίβα ἐπιλογῶν, κυρίως, χρησιμοποιώντας γιὰ τὸν ἐντοπισμὸ τους Wordsmith Tools σὲ ἓνα ἠλεκτρονικὸ σῶμα μεταφρασμένων κειμένων Ἐξετάζει μοτίβα ποὺ σχετίζονται μὲ τὸν γλωσσικὸ πλοῦτο καὶ τὴν ἀναλογία τύπου/λέξης (type/token ratio – T.T.R.), τὸ μέσο μῆκος τῆς πρότασης, τὴ διακύμανση στὰ διάφορα κείμενα, τὴ συχνότητα καὶ τὴν κατανομὴ τοῦ ρήματος say, ποὺ εἶναι τὸ συχνότερα ἐμφανιζόμενο ρῆμα στὰ ἀγγλικά (*ibid.*: 248 κ.έ.). Τέλος, θέτει πολὺ ἐνδιαφέροντα ἐρωτήματα μὲ ἄξονα τὴν ἔρευνα αὐτῆ, ὅπως τὸ κατὰ πόσον οἱ ἐπιλογές τοῦ μεταφραστῆ εἶναι ἄσχετες πρὸς τὸ ὕφος τοῦ πρωτοτύπου, πρὸς τίς δομὲς τῆς μητρικῆς του γλώσσας καὶ ἐκείνες μιᾶς συγκεκριμένης ποιητικῆς ἢ μιᾶς κοινωνιολέκτου καὶ ἂν οἱ ἐπιλογές αὐτές μποροῦν νὰ ἐξηγηθοῦν μὲ βάση τὴν κοινωνικὴ, πολιτιστικὴ καὶ ιδεολογικὴ θέση τοῦ μεταφραστῆ (*ibid.*: 248).

3. Ἡ συγκρότηση τῆς ὑποκειμενικότητας καὶ τοῦ λόγου τοῦ μεταφραστῆ

Μὲ τὴν Ὑφομετρία νὰ ἀναφέρεται ρητὰ γιὰ πρώτη φορὰ σὲ σχέση μὲ τίς μεταφραστικὲς σπουδὲς τὸ 2004 (Olohan, 2004: 145 κ.έ.), ὑπῆρξαν, ὅπως εἶδαμε, στὴ δεκαετία ποὺ ἀκολούθησε ἀρκετὲς ἐργασίες ἐστιασμένες στὸ ὕφος τοῦ μεταφραστῆ καὶ τὸν ἐντοπισμὸ τῆς παρουσίας του στὸ κείμενο μὲ χρῆση σωματίων κειμένων καὶ ποσοτικὲς καὶ στατιστικὲς μεθόδους. Ἡ νομιμοποίηση ὡστόσο τῶν μεθόδων αὐτῶν, ποὺ ἐπ' οὐδενὶ ἀκυρώνουν τὴν ποιοτικὴ ἀνάλυση, περνᾷ ὑποχρεωτικὰ μέσα ἀπὸ τὴν κατανόηση τῆς συγκρότησης τοῦ προσώπου καὶ τοῦ λόγου τοῦ μεταφραστῆ κατὰ τὴ μεταφραστικὴ του πρακτικὴ, ἀνάμεσα στὸ κοινωνικὸ καὶ τὸ ἀτομικὸ, τὸ συνειδητὸ καὶ τὸ μὴ συνειδητὸ ἐπίπεδο.

Προστρέχοντας καταρχάς στην κοινωνιολογία τῆς μετάφρασης, ἐπισημαίνουμε, μαζί μὲ τὸν Jean-Marc Gouanic, τὸν χαρακτήρα τῆς μετάφρασης ὡς πρακτικῆς, ὡς διαδικασίας παραγωγῆς (2006: 123). Ἐρειδόμενος στὴ γόνιμη συνεισφορά τοῦ Pierre Bourdieu ὅσον ἀφορᾷ τὴν ἔννοια τοῦ πεδίου ἀφενὸς καὶ τὴν ἔννοια τῆς ἔξης ἀφετέρου (Bourdieu, 1992) καὶ προσαρμόζοντάς τὴν στὴ μετάφραση, ὁ Gouanic ὑπογραμμίζει ὅτι τὸ πεδίο τοῦ μεταφραστῆ καὶ τοῦ μεταφράσματος, ὡς πεδίο ἀνταγωνισμῶν καὶ συγκρούσεων, εἶναι διαφορετικὸ ἀπὸ τὸ πεδίο τοῦ συγγραφέα καὶ τοῦ ξένου κειμένου, ἐφόσον ἡ θέση τῶν μεταφράσεων εἶναι στὸ λογοτεχνικὸ πεδίο ὑποδοχῆς (Gouanic, *ibid*: 125). Ὅσο γιὰ τὴν ἔξη, ὡς ὑποκειμενικὴ ἐνσωμάτωση τοῦ κοινωνικοῦ στοιχείου στὴ ζωὴ τοῦ μεταφραστῆ καὶ εἰδικότερα στὴ δουλειά του (εἰδικὴ ἔξη), πραγματώνεται μέσα στὸ ἐκάστοτε μεταφραστικὸ πεδίο, μὲ τὸ ὁποῖο τὴν συνδέει μιὰ σχέση μὴ-γραμμικῆ, καθὼς «ἡ ἔξη ἐνοικεῖ στὸ πεδίο καὶ τὸ πεδίο ἐνοικεῖ στὴν ἔξη» (*ibid*: 128).

Ἡ χρησιμοποίηση τῶν ἐννοιῶν τοῦ πεδίου καὶ τῆς ἔξης ἔχει στόχο νὰ ὑπερπηρῆθουν ποικίλα ἐρμηνευτικὰ ζεύγη στὶς μεταφραστικὲς σπουδές, ὅπως οἱ νόρμες (Toury, 1995), οἱ ἠθικὲς ἐπιλογές καὶ οἱ στρατηγικὲς τοῦ ξενισμοῦ καὶ τῆς *ἰδιοποίησης* (Berman, 1995· Venuti, 1995). Καρδιά τῆς μετάφρασης, ὡστόσο, μέσα σ' αὐτὴ τὴ συναίρεση τοῦ ἀτομικοῦ μὲ τὸ κοινωνικὸ στοιχεῖο στὴν ὑποκειμενικότητα καὶ τὴν ἱστορικότητα τοῦ μεταφραστῆ, παραμένει ὁ λόγος, ὅπως ἐκφέρεται στὸ μεταφρασμένο λογοτεχνικὸ κείμενο. Στὴν ἐκφορά του, ὑπογραμμίζει ὁ Dominique Maingueneau, «ἡ ἀναφορά στὸν κόσμο συναιρεῖται μὲ τὴν ἐγγραφή τοῦ ὁμιλοῦντος ὑποκειμένου στὸν ἴδιο τὸν λόγο του» (2004), δημιουργώντας νέες σχέσεις ἀνάμεσα στὴ γλωσσολογία καὶ τὴ λογοτεχνία. Ἡ λογοτεχνία ὡς λόγος/discours θέτει τὸ ζήτημα τοῦ ὁμιλητῆ, τῶν πολλαπλῶν ὁμιλητῶν στὴν περίπτωση τῆς μετάφρασης, καὶ ἀναδομεῖ τὴν περὶ ὕφους ἀντίληψη — σφραγίζοντας μάλιστα κατὰ τὸν Maingueneau τὸ τέλος τῆς Ἰσολογίας (*ibid.*).

Μιὰ τέτοια προσέγγιση ἀναδεικνύει τὰ πολλαπλὰ ἐπίπεδα τοῦ λόγου — καὶ τοῦ ὕφους — στὴ μετάφραση, εἰδικὰ ὅταν ὁ μεταφραστὴς εἶναι συγγραφέας ὁ ἴδιος: ὁ συγγραφέας-μεταφραστὴς, ὅπως εἶναι ὁ Παπαδιαμάντης, τοποθετεῖται ἔναντι τοῦ ὕφους αὐτοῦ μέσα ἀπὸ τὸ ἀτομικὸ, γλωσσικὸ ὕφος του, ἀλλὰ καὶ τὴ λογοτεχνικὴ του ἰδιοπροσωπία, ἡ ὁποία ἀναδραστικὰ καθορίζει τὴν ἴδια τὴ γλώσσα (Taylor, 1993).

4. Ὁ μεταφραστής Παπαδιαμάντης

Ἐπανερχόμενοι στις μεταφράσεις τοῦ Παπαδιαμάντη, σημειώνουμε λοιπὸν ὀρισμένες παραμέτρους τῆς μεταφραστικῆς του πρακτικῆς. Στὸ πεδίο τῆς μετάφρασης, ὁ Παπαδιαμάντης βιοποριζόταν ἀπὸ τὴ μετάφραση καὶ δὴ ἀπὸ τὴν ὡς ἐπὶ τὸ πλεῖστον ἀνυπόγραφη μετάφραση λογοτεχνικῶν κατὰ κύριο λόγο κειμένων γιὰ ἐφημερίδες, διηγημάτων ἢ ἐπιφυλλιδικῶν μυθιστορημάτων, σύμφωνα μὲ τὴν τρέχουσα πρακτικὴ τῆς ἐποχῆς. Μιᾶς ἐποχῆς ὅπου ἡ μετάφραση κυριαρχεῖ: «κάθε συζήτηση γιὰ τὴ γλῶσσα, τὴν κριτικὴ, τὴν ἐθνικὴ λογοτεχνία, τὰ λογοτεχνικὰ εἶδη καὶ ἤθη, περνᾶει ἀπὸ τὶς μεταφράσεις. Κάθε λογοτέχνης εἶναι καὶ μεταφραστής, συχνὰ μάλιστα κυρίως μεταφραστής. Κάθε ἔντυπο τῆς ἐποχῆς βρίθεται ἀπὸ μεταφράσεις» (Παππᾶς, 2010). Ἡ ἀφάνειά του ὡς μεταφραστῆ στις ἐφημερίδες ἀντανακλᾷ μιὰ ἀντίληψη περὶ τῆς μετάφρασης ὡς ἀδιαμεσολάβητης μεταφορᾶς, ἡ ὁποία κυριαρχεῖ, ἐντέλει, ὡς σήμερα (Saldanha, 2011). Ἀπὸ τὴν ἄλλη, ἡ συνεργασία μὲ τὶς ἐφημερίδες θέτει ἀσφυκτικὲς προθεσμίες. Ὁ συνδυασμὸς τῆς ἀωνυμίας μὲ τὴν πίεση χρόνου εἶναι πολὺ πιθανὸ ὅτι λειτουργοῦσε στὴν κατεύθυνση τῆς ἐνίσχυσης τῆς ἀτομικῆς ιδιολέκτου, ὡς ὑποστηρικτικοῦ αὐτοματισμοῦ.

Ὡς πρὸς τὴν ἔξη, ἡ πρωταρχικὴ ἔξη, ποὺ στὸν Παπαδιαμάντη ὀρίζεται ἀπὸ τὴν καταγωγή του καὶ τὴν ἰδιαίτερη σχέση του μὲ τὴν Ἐκκλησία μεταξὺ ἄλλων, συνδυάζεται μὲ τὴν εἰδικὴ ἔξη, τὸ γεγονός ὅτι εἶναι αὐτοδίδακτος στις γλῶσσες, τὴν ἀντίληψή του γιὰ τὴν ξένη λογοτεχνία καὶ τὸν ξένο πολιτισμὸ μὲ τὶς ἀναλογίες ποὺ ὁ ἴδιος, ὡς ἐκ τῆς καταγωγῆς καὶ τῆς συγκρότησής του, δημιουργεῖ (Gouanic, *ibid.*, 127 κ.έ.). Ἄν, σὲ συνδυασμὸ μὲ τὴν ἔννοια τῆς ἔξης καὶ τοῦ πεδίου, δεχτοῦμε τὴ θέση τῆς Guillemin-Flescher, σύμφωνα μὲ τὴν ὁποία ἡ διάσταση τῆς γλῶσσας εἶναι «ἡ γλωσσικὴ ἀναπαράσταση πολιτιστικῶν συμπεριφορῶν ποὺ ἐγγράφονται τόσο στὴν ἐπιλογὴ τῶν ρηματικῶν τύπων, στὶς προτιμώμενες συντακτικὲς δομὲς ὅσο καὶ στὴ συλλογικὴ ὀργάνωση τοῦ λόγου» (Flescher 1994: 38, παρατίθεται στὸ Nenopoulou, 2005: 134), ἀντιλαμβανόμεστε τὴ σημασία τῶν ἀσύνειδων ἐγγραφῶν στὸ μεταφρασμένο κείμενο, ἐγγραφῶν ποὺ προκύπτουν ἀπὸ τὴ συγκεκριμένη συγκρότησή του σὲ μιὰ συγκεκριμένη ἱστορικὴ στιγμή.

Δὲν πρέπει βεβαίως νὰ ξεχνᾷμε τὴ μεταφραστικὴ ἐντολή, τὸν προσανατολισμὸ δηλαδή τοῦ κειμένου ποὺ μεταφράζει ὁ Παπαδιαμάντης σὲ ἕνα

εύρυν κοινό, μέσα από την έφημερίδα, με δεδομένο ότι το μέσο αλλάζει το μήνυμα: η συνεξέταση βέβαιων μεταφράσεων του Παπαδιαμάντη σε έφημερίδες και αυτοτελείς εκδόσεις θα μπορούσε να δείξει αν και κατά πόσο η μεταφραστική πρακτική του διαφοροποιείται αναλόγως με την έκδοση σε βιβλίο ή την ανώνυμη δημοσίευση στην έφημερίδα.

5. Καθορισμός του έρευνητικού προβλήματος

Η ύφομετρική απόδοση της συγγραφικής πατρότητας (authorship attribution) είναι πλέον ένα καλά μελετημένο πρόβλημα που τα τελευταία χρόνια έχει ένταχθει στην έρευνητική παλέτα του κλάδου της Έξόρυξης Κειμένου (Text Mining) που με τη σειρά του ανήκει στην έπιστημονική περιοχή της Ανάκτησης της Πληροφορίας (Information Retrieval). Στο πλαίσιο αυτής της προσέγγισης ή απόδοση ενός αγνώστου πατρότητας κειμένου σε συγκεκριμένο συγγραφέα μεταξύ ενός κλειστού αριθμού πιθανών συγγραφέων αντιμετωπίζεται ως ένα πρόβλημα κειμενικής κατηγοριοποίησης (text categorization). Σύμφωνα με αυτή την προσέγγιση τα άδιαμφισβήτητα κείμενα των υποψήφιων συγγραφέων μετατρέπονται σε διανύσματα μετρήσεων που αντιπροσωπεύουν τη συχνότητα ποικίλων ύφομετρικών χαρακτηριστικών σε αυτά. Έν συνέχεια τα διανύσματα αυτά χρησιμοποιούνται ως σώμα εκπαίδευσης (training set) για να «διδάξουν» έναν συγκεκριμένο στατιστικό αλγόριθμο: ποια ύφομετρικά χαρακτηριστικά και σε τί βαθμό αποτελούν το ύφομετρικό προφίλ του κάθε υποψήφιου συγγραφέα. Στην ουσία, αναπτύσσεται για κάθε υποψήφιο συγγραφέα ένα ύφομετρικό μοντέλο το οποίο χρησιμεύει ως μήτρα σύγκρισης για κάθε ανώνυμο κείμενο του οποίου ή πατρότητα θα έλεγχθει. Η απόδοση της συγγραφικής πατρότητας στηρίζεται στη σύγκριση του ύφομετρικού προφίλ του ανώνυμου κειμένου με τα ύφομετρικά προφίλ καθενός από τους υποψήφιους συγγραφείς. Η ακρίβεια αυτής της μεθόδου, αν και εξαρτάται σημαντικά τόσο από τα χαρακτηριστικά όσο και από τον αλγόριθμο που χρησιμοποιείται, όταν οι υποψήφιοι συγγραφείς είναι έως 4-5, μπορεί να φτάσει στο 100% (Mikros, 2014).

Όστόσο, δέν είναι πάντα όλα τα προβλήματα απόδοσης συγγραφικής πατρότητας προσδιορισμένα σε κλειστό σύνολο υποψήφιων συγγραφέων. Ύπάρχουν πολλές περιπτώσεις όπου θέλουμε να διαπιστώσουμε κατά πόσο ένα κείμενο έχει γραφτεί από κάποιον συγκεκριμένο συγγραφέα χωρίς

όμως να είμαστε σίγουροι ότι αυτός το έχει γράψει. Στην πραγματικότητα μπορεί να το έχει γράψει οποιοσδήποτε και εμείς δεν γνωρίζουμε a priori ποιός είναι ο πραγματικός συντάκτης του. Ένα τέτοιο πρόβλημα ορίζεται ως πρόβλημα πιστοποίησης συγγραφέα (authorship verification) και είναι πολύ πιο δύσκολο στην επίλυσή του από το πρόβλημα απόδοσης συγγραφικής πατρότητας (authorship attribution). Στη περίπτωση αυτή έχουμε τη δυνατότητα να δημιουργήσουμε ένα ύφομετρικό προφίλ του συγγραφέα που θέλουμε να ελέγξουμε την πιθανότητα να είναι αυτός ο συγγραφέας του κειμένου, αλλά δεν μπορούμε να δημιουργήσουμε έναλλακτικά ύφομετρικά προφίλ, αφού δεν γνωρίζουμε ποιοι άλλοι είναι οι πιθανοί συγγραφείς. Έρευνητικά το πρόβλημα της συγγραφικής πιστοποίησης έχει αντιμετωπιστεί με ποικίλες μεθόδους, όπως ή χρήση αλγόριθμων μονής κατηγορίας (one-class algorithms) (Koppel & Schler, 2004· Manevitz & Yousef, 2001), ή ανάπτυξη μεθόδων αποκάλυψης (unmasking methods) (Koppel, Schler, & Boucek-Dokow, 2007), ό έντοπισμός ακραίων τιμών (outlier detection) (Stein, Lipka, & Prettenhofer, 2011) κ.ά. Αν και ό συγκεκριμένος τομέας παρουσιάζει σημαντική έρευνητική δραστηριότητα τα τελευταία χρόνια, ή ακρίβεια αναγνώρισης υπολείπεται σημαντικά τών αποτελεσμάτων που επιτυγχάνονται στις μελέτες απόδοσης πατρότητας (Koppel, Schler, & Argamon, 2009).

Η περίπτωση τών ανώνυμων μεταφράσεων που πιθανόν να έγραψε ό Παπαδιαμάντης αποτελούν μια πραγματική πρόκληση στις ύφομετρικές τεχνικές για δύο λόγους:

α) Η μετάφραση ως κειμενικό είδος περιέχει ύφομετρικά ίχνη τόσο του μεταφραζόμενου συγγραφέα όσο και του μεταφραστή. Δεν μπορούμε να γνωρίζουμε ποιά από τα ύφομετρικά χαρακτηριστικά του συγγραφέα επιβιώνουν στη μετάφραση, ποιά τροποποιούνται λόγω κωδικοποίησης του κειμένου σε διαφορετική γλώσσα αλλά παραμένουν χαρακτηριστικά του συγγραφέα, και ποιά εισάγονται από τόν μεταφραστή και άπηχοϋν την ιδιόλεκτό του.

β) Η ύφομετρική απόδοση της πατρότητας μεταφράσεων σε μεταφραστή είναι ένα σύνθετο πρόβλημα προσδιορισμού συγγραφικού ύφους που ανάγεται στο πρόβλημα της συγγραφικής πιστοποίησης. Αυτό σημαίνει ότι οι ανώνυμες μεταφράσεις που θα εξετάσουμε δυνητικά θα μπορούσαν να έχουν γραφτεί από οποιοδήποτε συγγραφέα της εποχής (συμπεριλαμβανομένου και του Παπαδιαμάντη).

Λαμβάνοντας υπόψη τὰ παραπάνω σημεία, βασικές μεθοδολογικές επιλογές τῆς παρούσας ἔρευνας εἶναι οἱ ἀκόλουθες:

1. Προσδιορισμὸς καὶ μέτρηση μεγάλου ἀριθμοῦ ὑφομετρικῶν χαρακτηριστικῶν ποὺ ἀνήκουν σὲ πολλαπλὰ γλωσσικά ἐπίπεδα καὶ παρουσιάζουν ὑψηλὴ συχνότητα ἐμφάνισης στὴν ἑλληνικὴ γλώσσα. Στόχος μας εἶναι ἡ ἀνάλυση τῶν κειμένων στὸ εὐρύτερο δυνατὸ ἐπίπεδο, ἔτσι ὥστε οἱ ἀλγόριθμοι κατηγοριοποίησης νὰ μποροῦν νὰ ἀξιολογήσουν καὶ νὰ ἐπιλέξουν μόνον ἐκεῖνα τὰ χαρακτηριστικά ποὺ σχετίζονται μὲ τὴ μεταφραστικὴ ιδιόλεκτο τοῦ Παπαδιαμάντη καὶ ὄχι ἐκεῖνα ποὺ συνδέουν τίς μετα-φράσεις μὲ τὸν συγγραφέα τοῦ πρωτότυπου κειμένου. Ὁ ἐξαναγκασμὸς τῶν ἀλγορίθμων νὰ χρησιμοποιήσουν μόνον τὰ ὑφομετρικὰ χαρακτηριστικά ποὺ εἰσάγει ὁ Παπαδιαμάντης στὶς μεταφράσεις ποὺ κάνει, γίνεται ἐφικτὸς ἐπειδὴ ἔχουμε στὴ διάθεσή μας βέβαιες μεταφράσεις τοῦ Παπαδιαμάντη. Αὐτές, σὲ συνδυασμὸ μὲ τὰ αὐθεντικὰ ἔργα του, ἀξιοποιοῦνται ὥστε νὰ ἀναπτυχθεῖ ἓνα ἐνοποιημένο ὑφομετρικὸ προφίλ ποὺ στηρίζεται σὲ μία εὐρεία βάση ὑφομετρικῶν χαρακτηριστικῶν, τὰ ὁποῖα χαρακτηρίζουν ποσοτικά τὸν παπαδιαμαντικὸ τρόπο γραφῆς ἀνεξαρτήτως κειμενικοῦ γένους.
2. Μετατροπὴ τοῦ προβλήματος συγγραφικῆς πιστοποίησης σὲ πρόβλημα ἀπόδοσης συγγραφικῆς πατρότητας. Δεδομένου ὅτι οἱ μέθοδοι ἀπόδοσης συγγραφικῆς πατρότητας εἶναι σημαντικὰ πιὸ ἀκριβεῖς, ἐπιχειρήσαμε νὰ ἐπαναδιατυπώσουμε τὸ συγκεκριμένο ἐρευνητικὸ πρόβλημα ὡς πρόβλημα ἐπιλογῆς πιθανοῦ συγγραφέα ἀπὸ κλειστὸ σύνολο ἐπιλογῶν. Αὐτὸ ἐπιτεύχθηκε δημιουργώντας μία ξεχωριστὴ κατηγορία πιθανοῦ συγγραφέα ποὺ τὴν ὀνομάσαμε «Σύγχρονοι τοῦ Παπαδιαμάντη». Πρόκειται γιὰ τὴν ἀνάπτυξη ἐνὸς Ἡλεκτρονικοῦ Σώματος Κειμένων (ΗΣΚ) μὲ δείγματα γραφῆς συγγραφέων ποὺ ἦταν σύγχρονοι τοῦ Παπαδιαμάντη καὶ ἔγραφαν σὲ παρόμοιο κειμενικὸ γένος. Μὲ τὴν ἀνάπτυξη ἐνὸς τέτοιου ΗΣΚ ἀναφορᾶς ἐπιτρέπουμε στὸν ἀλγόριθμο κατηγοριοποίησης νὰ διαλέξει κατὰ πόσο μία μετάφραση ἀνήκει στὸν Παπαδιαμάντη ἢ ὄχι, ἔχοντας ὡς ἐναλλακτικὴ λύση τὸ γενικὸ ὑφομετρικὸ προφίλ πολλῶν διαφορετικῶν συγγραφέων τῆς ἴδιας ἐποχῆς. Ἡ ἀπόδοση μιᾶς μετάφρασης στὴ γενικὴ κατηγορία «Σύγχρονοι τοῦ Παπαδιαμάντη» δὲν συνεπάγεται ὅτι ὑποχρεωτικά κάποιος ἀπὸ τοὺς συγγραφεῖς ποὺ ἐμπεριέχονται σὲ αὐτὴν εἶναι ὁ πραγματικὸς συγγραφέας τῆς. Ὁ ρόλος τῆς κατηγορίας αὐτῆς εἶναι ἡ προσφορὰ δυνατότητας στὸν ἀλγόριθμο νὰ μπορεῖ νὰ ἀποδώσει σὲ κάποια ἐναλλακτικὴ κατηγορία τὴ μετάφραση ποὺ δὲν ἔκανε ὁ Παπαδιαμάντης.

6. Ανάπτυξη και προεπεξεργασία τῶν Ἡλεκτρονικῶν Σωμάτων Κειμένων τῆς ἔρευνας

Ἡ σύγχρονη ὑφομετρικὴ ἀνάλυση ἀπαιτεῖ τὴ χρήση ψηφιοποιημένων ἐκδόσεων τῶν κειμένων, γιὰ νὰ καταστεῖ ἐφικτὴ ἡ ὁποιαδήποτε περαιτέρω ὑπολογιστικὴ ἐπεξεργασία τους. Στὴν παρούσα ἔρευνα χρησιμοποιήθηκαν τέσσερα διαφορετικὰ εἶδη ΗΣΚ, τὰ μεγέθη καὶ ἡ σύσταση τῶν ὁποίων βρίσκονται στὸ Παράρτημα:

1. Τὸ ΗΣΚ τῶν Ἀπάντων τοῦ Παπαδιαμάντη ἀπὸ τὴν κριτικὴ ἔκδοση Ν. Δ. Τριανταφυλλόπουλου.
2. Τὸ ΗΣΚ τεσσάρων βέβαιων μεταφράσεων τοῦ Παπαδιαμάντη¹. Ἀπὸ τίς τέσσερις αὐτὲς βέβαιες μεταφράσεις τοῦ Παπαδιαμάντη οἱ δύο (*Ὁ ἰατρός Ραμῶ καὶ Ὁ Μαξιώτης*) μᾶς παραχωρήθηκαν σὲ ἠλεκτρονικὴ μορφή, οἱ ἄλλες δύο (*Ἐγκλημα καὶ Τιμωρία καὶ Ταρταρίνος ὁ ἐκ Ταρασκῶνος*) ψηφιοποιήθηκαν γιὰ τίς ἀνάγκες τῆς παρούσας ἔρευνας.
3. Τὸ ΗΣΚ «Σύγχρονοι τοῦ Παπαδιαμάντη» ποὺ ἀποτελεῖται ἀποκλειστικὰ ἀπὸ ἤδη ψηφιοποιημένα κείμενα συγχρόνων συγγραφέων τοῦ Παπαδιαμάντη, τὰ ὁποῖα βρίσκονται ἐλεύθερα στὸ Διαδίκτυο².
4. Τὸ ΗΣΚ ἀνώνυμων μεταφράσεων ποὺ ἀποτελεῖται ἀπὸ ἕξι συνολικὰ κείμενα ποὺ ψηφιοποιήθηκαν γιὰ τίς ἀνάγκες τῆς παρούσας ἔρευνας³.

Ἡ ὑπολογιστικὴ ἐπεξεργασία ΗΣΚ ἀπαιτεῖ ὑψηλὸ βαθμὸ ὁμογενοποίησης τῶν κειμένων ποὺ τὰ ἀπαρτίζουν, σὲ ἓνα μεγάλο εὖρος παραμέτρων, ὅπως ἡ χρῆση ἴδιου προτύπου ἀποθήκευσης ἀρχείου, κοινοῦ προτύπου κωδικοποίησης τῆς γλώσσας, κ.ἄ. Γιὰ νὰ γίνεῖ ἐφικτὴ μιὰ τέτοια κειμενικὴ «κανονικοποίηση», ὄρισάμε μιὰ σειρά ἀπὸ βήματα κειμενικῆς προεπεξεργασίας, τὰ ὁποῖα εἶναι τὰ ἀκόλουθα:

1. Μετατροπὴ ὁποιων ἀρχείων ἦταν διαθέσιμα σὲ πρότυπο ἐγγράφου Microsoft Word (κατάληξη ἀρχείου doc) σὲ πρότυπο ἀπλοῦ κειμένου (κατάληξη ἀρχείου txt).

¹ Βλ. παρακ. Πίνακα 7. Ἡ βεβαιότητα τῆς πατρότητας τῶν μεταφράσεων αὐτῶν προκύπτει εἴτε ἀπὸ τὸ γεγονός ὅτι, ζῶντος τοῦ Παπαδιαμάντη, εἶχαν ἐκδοθεῖ σὲ βιβλίο μὲ τὸ ὄνομα τοῦ μεταφραστῆ, εἴτε ἀπὸ πληροφορίες ἀπὸ συγχρόνους του γιὰ τὸ ὄνομα τοῦ μεταφραστῆ δημοσιευμένες ἐπίσης ζῶντος τοῦ Παπαδιαμάντη.

² Βλ. παρακ. Πίνακα 6.

³ Βλ. παρακ. Πίνακα 8.

2. Μετατροπή όλων των αρχείων άπλου κειμένου στο πρότυπο κωδικοποίησης ελληνικής γλώσσας Windows cp1253.
3. Μετατροπή όποιων αρχείων ήταν πολυτονικά σε μονοτονικά χρησιμοποιώντας λογισμικό που αναπτύχθηκε ειδικά για την παρούσα έρευνα.
4. «Καθαρισμός» ορισμένων κειμένων από συστηματικά λάθη που είχαν προκύψει κατά τη διάρκεια της ψηφιοποίησής τους.

Μετά την ολοκλήρωση κάθε βήματος της προεπεξεργασίας γινόνταν συστηματικός δειγματοληπτικός έλεγχος στα κείμενα, για να αποκλειστεί ή πιθανότητα εισαγωγής λαθών από λανθασμένη ή απρόβλεπτη συμπεριφορά των λογισμικών που χρησιμοποιούσαμε.

Έν συνεχεία τα ΗΣΚ αναλύθηκαν με τη βοήθεια εξειδικευμένων εργαλείων Έπεξεργασίας Φυσικής Γλώσσας (Natural Language Processing – NLP), έτσι ώστε να είναι έφικτη ή δυνατότητα μέτρησης ενός μεγάλου αριθμού ύφομετρικών χαρακτηριστικών. Ειδικότερα τα στάδια ανάλυσης που ακολουθήσαμε είναι τα εξής:

1. Χωρισμός των κειμένων σε ισομεγέθη τμήματα των 1000 λέξεων (chunking). Ο τεμαχισμός των κειμένων έξυπηρετεί δύο σημαντικές έρευνητικές ανάγκες:
 - i. Πολλαπλασιάζει το μέγεθος του δείγματος, αφού ένα μακροσκελές κείμενο μπορεί να μετατραπεί σε πολλές δεκάδες μικρότερων αρχείων. Αυτό με τη σειρά του είναι ιδιαίτερα εύρηγετικό στη στατιστική ανάλυση των δεδομένων που προκύπτουν, αφού το μεγαλύτερο μέγεθος του δείγματος αυξάνει την αξιοπιστία και την έγκυρότητα της στατιστικής ανάλυσης. Έτσι, ενώ συνολικά τα 4 ΗΣΚ της έρευνας απαρτίζονταν από 207 κείμενα, μετά τον τεμαχισμό ο αριθμός των συνολικών κειμένων ανήλθε σε 1508 κειμενικά τεμάχια.
 - ii. Κανονικοποιεί τις μετρήσεις των ύφομετρικών χαρακτηριστικών σε ίσα μεγέθη κειμένου. Πολλά ύφομετρικά χαρακτηριστικά επηρεάζονται σημαντικά από το μέγεθος του κειμένου (Tweedie & Baayen, 1998) και η χρήση ισομεγεθών τεμαχίων αίρει τον συγκεκριμένο περιορισμό. Το κατώφλι των 1000 λέξεων ανά τεμάχιο όρίστηκε με βάση τα πορίσματα σχετικών έρευνών (Juola, Sofko, & Brennan, 2006· Mikros, 2007· Stamatatos, Fakotakis, & Kokkinakis, 1999· Zhao & Zobel, 2007) που έχουν δείξει ότι οι περισσότερες ύφομετρικές μετρήσεις σταθεροποιούνται σε αυτό το μέγεθος.
2. Ανάλυση σε βασικές λεξικές μονάδες (tokenization). Τα κειμενικά τεμά-

για μετατράπηκαν σε άρχεϊα κειμένου με τήν κάθε λεξική μονάδα να καταλαμβάνει μία ξεχωριστή σειρά.

3. *Ανάλυση σε διγράμματα και τριγράμματα χαρακτήρων* (character bi-gram and tri-gram analysis). Έκτος τής ανάλυσης σε λεξικές μονάδες, τὰ κειμενικά τεμάχια αναλύθηκαν σε ύπολεξικές ενότητες που ονομάζονται ν-γράμματα χαρακτήρων. Τò ν είναι κάθε φορά ο αριθμός τών γειτονικῶν χαρακτήρων που συναποτελοῦν τò ν-γράμμα. Έτσι, ὅταν ἡ ανάλυση γίνεται με ν=2, δηλαδή διγράμματα χαρακτήρων, τότε τò κείμενο τεμαχίζεται σε ὅλες τις παρατηρούμενες ἀκολουθίες δύο γειτονικῶν χαρακτήρων. Π.χ. ἔστω ὅτι ἔχουμε τò ἀκόλουθο κείμενο: *Η βροχή είναι δυνατή.* Τὰ διγράμματα χαρακτήρων τοῦ συγκεκριμένου κειμένου είναι τὰ ἀκόλουθα 16:

[Η β], [βρ], [ρο], [οχ], [χή], [ή ε], [εί], [ίν], [να], [αι], [ιδ], [δου], [υν], [να], [ατ], [τή.]

7. *Υφομετρικά χαρακτηριστικά που μετρήθηκαν στην παρούσα ἔρευνα*

Σύμφωνα με τὸν Rudman (1997) τὰ ὑφομετρικά χαρακτηριστικά που κατὰ καιρούς ἔχουν χρησιμοποιηθεῖ σε ἔρευνες ἀπόδοσης πατρότητας κειμένων ξεπερνῶν τὰ 1000 και ἀνήκουν σε ποικίλα ἐπίπεδα γλωσσικῆς ανάλυσης. Στην παρούσα μελέτη ἐπιλέξαμε νὰ συνδυάσουμε ὑφομετρικά χαρακτηριστικά που ἔχουν μεγάλη ἱστορία στὸν χῶρο τῆς ἀπόδοσης πατρότητας ἀνώνυμων κειμένων (π.χ. τὸ μέσο μῆκος λέξης και τὸ μέσο μῆκος πρότασης) και μία σειρά ἀπὸ χαρακτηριστικά που ἔχουν προκύψει μέσα ἀπὸ τὴ σύγχρονη ἔρευνα στὸν χῶρο τῆς Ἀνάκτησης Πληροφορίες (π.χ. ἡ κανονικοποιημένη συχνότητα τῶν πιὸ συχνῶν διγραμμάτων και τριγραμμάτων χαρακτήρων). Πιὸ ἀναλυτικά τὰ ὑφομετρικά χαρακτηριστικά που μετρήθηκαν είναι τὰ ἀκόλουθα:

A. «Κλασικά» ὑφομετρικά χαρακτηριστικά

1. *Γλωσσικός «Πλοῦτος»*

- i. Yule's K: Ὑπολογίζεται ὁ δείκτης K ὁ ὁποῖος ἀποτελεῖ τὸν πιὸ ἀξιόπιστο δείκτη λεξιλογικοῦ «πλούτου» στὴ σχετικὴ βιβλιογραφία. Ὁ ὑπολογισμὸς του ἀκολουθεῖ τὸν τύπο τῶν Tweedie & Baayen (1998, p. 330). Ὅσο πιὸ μικρὸς εἶναι ὁ δείκτης τόσο μεγαλύτερη λεξιλογικὴ ποικιλία παρουσιάζει τὸ κείμενο.
- ii. Λεξιλογικὴ Πυκνότητα: Ὁ λόγος τοῦ ποσοστοῦ τῶν λέξεων «παραχομένου» (content words) πρὸς τὸ ποσοστὸ τῶν λειτουργικῶν λέξεων (function words). Ὁ ὅρος «Λεξιλογικὴ Πυκνότητα» (Lexical Density)

χρησιμοποιείται στη θεωρία της λειτουργικής γραμματικής του Halliday με έλαφρά διαφορετικό τρόπο υπολογισμού. Έδω ακολουθούμε την έκδοση που χρησιμοποιείται αρκετά συχνά σε μελέτες έντοπισμού άγνωστου συγγραφέα και απαντάται συχνά και με τον όρο «Λειτουργική Πυκνότητα» (Functional Density) (Miranda & Calle, 2007). Όσο μεγαλύτερος είναι ο δείκτης τόσο μεγαλύτερο τμήμα του κειμένου αποτελείται από λέξεις περιεχομένου.

- iii. Άπαξ Λεγόμενα: Υπολογίζεται το ποσοστό των λέξεων που εμφανίζονται στο κείμενο με συχνότητα 1.
- iv. Δις Λεγόμενα: Υπολογίζεται το ποσοστό των λέξεων που εμφανίζονται στο κείμενο με συχνότητα 2.
- v. Λόγος Δις πρὸς Άπαξ Λεγόμενα: Ο λόγος αυτός έχει προταθεί στη βιβλιογραφία ως ένδεικτικός του συγγραφικού ύφους (Hoover, 2003).
- vi. Έντροπία (Entropy): Υπολογίζεται η έντροπία του κάθε κειμένου, ο βαθμός δηλαδή της οργάνωσης και της προβλεψιμότητας των λεξικών συχνοτήτων. Ο υπολογισμός της ακολουθεί τον τύπο του Oakes (1998, p. 61). Όσο μεγαλύτερη είναι η τιμή της έντροπίας τόσο λιγότερο αναμενόμενες, με τη στατιστική σημασία του όρου, είναι οι λέξεις που προκύπτουν στο κείμενο.
- vii. Σχετική Έντροπία (Relative Entropy): Ο λόγος της θεωρητικά μέγιστης έντροπίας ενός κειμένου με την παρατηρηθείσα έντροπία. Μέγιστη έντροπία παρουσιάζει ένα κείμενο που κάθε λέξη που θα περιελάμβανε θα εμφανιζόταν 1 φορά και άρα στο κείμενο αυτό όλες οι λέξεις θα ήταν άπαξ λεγόμενα. Όσο μεγαλύτερη είναι η Σχετική Έντροπία, τόσο λιγότερο τυποποιημένο είναι το κείμενο, και άρα περισσότερο «πλούσιο» λεξιλογικά.

2. Μήκος Λέξης

- i. Μέσο Μήκος Λέξης: Υπολογίζεται ο μέσος όρος του μήκους των λέξεων του κάθε κειμένου με βασική μονάδα μέτρησης τον χαρακτήρα.
- ii. Τυπική Απόκλιση του Μέσου Μήκους Λέξης: Η τυπική απόκλιση του μέσου όρου του μήκους των λέξεων του κειμένου.
- iii. Φάσμα Συχνότητας Μήκους Λέξεων: Η συχνότητα με την οποία εμφανίζονται λέξεις με 1, 2, 3 ... 14 χαρακτήρες στο κείμενο.

3. Μήκος Πρότασης

- i. Μέσο Μήκος της Πρότασης: Ο μέσος όρος του μήκους των προτάσεων του κειμένου μετρημένο σε λέξεις.
- ii. Τυπική Απόκλιση του Μέσου Μήκους της Πρότασης: Η τυπική απόκλιση του μέσου μήκους των προτάσεων του κάθε κειμένου.

B. Υφομετρικά χαρακτηριστικά ανάκτησης πληροφορίας

1. Συχνότητα χαρακτήρων: Μετρήθηκε ή συχνότητα τῶν γραμμάτων σὲ ὅλα τὰ ΗΣΚ καὶ κανονικοποιήθηκε σὲ μέγεθος κειμένου 100 λέξεων.
2. Συχνότητα διγραμμάτων καὶ τριγραμμάτων χαρακτήρων: Τὰ πιὸ συχνὰ διγράμματα καὶ τριγράμματα χαρακτήρων ἔχουν χρησιμοποιηθεῖ μὲ ἐπιτυχία στὸν ἐντοπισμὸ τῆς πατρότητας ἀνώνυμων κειμένων (Grieve, 2007· Koppel, Schler, & Argamon, 2011· Luyckx & Daelemans, 2011) καὶ ἀποτελοῦν ἕναν ἀπὸ τοὺς πιὸ ἀξιόπιστους καὶ εὐρωστους δείκτες συγγραφικῆς πατρότητας. Στὴν παρούσα ἔρευνα χρησιμοποιήσαμε τὴν κανονικοποιημένη συχνότητα τῶν 2000 πιὸ συχνῶν διγραμμάτων καὶ τριγραμμάτων χαρακτήρων στὰ ΗΣΚ.
3. Συχνότητα συχνῶν λέξεων: Οἱ πιὸ συχνές λέξεις ἐνὸς κειμένου καλύπτουν ἕνα μεγάλο μέρος τῆς ὑφομετρικῆς ταυτότητας τοῦ συγγραφέα καὶ ἔχουμε πλῆθος ἐρευνῶν ποὺ ἐπιβεβαιώνουν αὐτὴ τὴν παρατήρηση, π.χ. (Allison & Guthrie, 2008· Coyotl-Morales, Villasenor-Pineda, Montes-y-Gomez, & Rosso, 2006· Diederich, Kindermann, Leopold, & Paass, 2003). Στὴν παρούσα ἔρευνα χρησιμοποιήσαμε τὴν κανονικοποιημένη συχνότητα τῶν 2000 πιὸ συχνῶν λέξεων τῶν ΗΣΚ.

Ὁ συνολικὸς πίνακας διανυσμάτων περιεῖχε συνολικὰ 6.054 χαρακτηριστικά ποὺ μετρήθηκαν σὲ 1508 κειμενικὰ τεμάχια.

8. Ἀλγόριθμος ταξινόμησης καὶ μεθοδολογία ἀξιολόγησης τῆς ἀκρίβειάς του

Τὸ μέγεθος τῶν δεδομένων ποὺ δημιουργήθηκε, ἡ ἐξαιρετικὰ στρεβλωμένη κατανομὴ τῶν περισσότερων χαρακτηριστικῶν ποὺ μετρήθηκαν ἀλλὰ καὶ ἡ διαστασιμότητα (dimensionality), δηλαδή ἡ ὑπαρξὴ μεγαλύτερου ἀριθμοῦ ὑφομετρικῶν χαρακτηριστικῶν ἀπὸ κείμενα πρὸς ταξινόμηση, καθιστᾷ ἀπαγορευτικὴ ὅποιαδήποτε χρῆση παραδοσιακῶν στατιστικῶν μεθόδων. Γιὰ τὸν λόγο αὐτὸ ἀποφασίσαμε νὰ χρησιμοποιήσουμε μεθόδους ἀνάλυσης ποὺ ἀνήκουν στὸν εὐρύτερο χῶρο τῆς Τεχνητῆς Νοημοσύνης (Artificial Intelligence) καὶ εἰδικότερα στὸν κλάδο τῆς Μηχανικῆς Μάθησης (Machine Learning). Στὸ πλαίσιο τῆς σχετικῆς ἔρευνας ἔχουν ἀναπτυχθεῖ μιὰ σειρά ἀπὸ ἀλγόριθμους ταξινόμησης (classification algorithms), οἱ ὁποῖοι μποροῦν νὰ χειριστοῦν δεδομένα ὑψηλῆς διαστασιμότητας μὲ ἀξιόπιστία.

Ἐνας ἀπὸ τοὺς πλέον εὐρωστους ἀλγόριθμους ταξινόμησης εἶναι οἱ

Μηχανές Διανυσμάτων Ύποστήριξης - ΜΔΥ (Support Vector Machines - SVM), οι οποίες αναπτύχθηκαν αρχικά για δυαδικά προβλήματα ταξινόμησης (Vapnik, 1995). Προβάλλουν τα σημεία των δεδομένων με τα όποια εκπαιδεύονται σε έναν χώρο περισσότερων διαστάσεων και βρίσκουν το υπερεπίπεδο, το οποίο διαχωρίζει βέλτιστα τα σημεία των δύο τάξεων. Αυτή η ιδιότητά τους τις κάνει εξαιρετικά χρήσιμες στην ταξινόμηση μη γραμμικών προβλημάτων, όπως είναι οι περισσότερες περιπτώσεις έντοπισμού συγγραφέα ανώνυμων κειμένων. Στην παρούσα έρευνα χρησιμοποιήσαμε μια βελτιωμένη έκδοση των ΜΔΥ που ονομάζεται Sequential Minimal Optimization (SMO) (Platt, 1999) που βελτιώνει σημαντικά την απόδοση των ΜΔΥ σε ταχύτητα εκτέλεσης.

Ένα σημαντικό κομμάτι της μεθοδολογίας χρήσης των αλγόριθμων μηχανικής ταξινόμησης σχετίζεται με την αξιολόγηση της έγκυρότητάς τους. Για τον λόγο αυτό έχουν αναπτυχθεί διάφορες μέθοδοι, οι οποίες προσπαθούν να ελέγξουν κατά πόσο ο αλγόριθμος «έμαθε» πραγματικά να διακρίνει το ύφος ενός συγγραφέα σε σχέση με άλλους και κατά πόσο η «μάθηση» που επιτεύχθηκε είναι αρκετά γενικευμένη ώστε ο αλγόριθμος να μπορεί να αναγνωρίζει κείμενα του συγγραφέα που δεν περιέχονταν στο σώμα εκπαίδευσης που του δόθηκε. Η μέχρι τώρα έρευνα (Witten & Frank, 2005) έχει καταλήξει ότι η πλέον αξιόπιστη μέθοδος αξιολόγησης της μηχανικής μάθησης είναι η 10-πτυχη διασταυρούμενη επικύρωση (10-fold cross-validation). Σε αυτήν, το 90% των δεδομένων χρησιμοποιείται ως δείγμα εκπαίδευσης (training sample) και το 10% ως δείγμα επικύρωσης (validation sample). Ο αλγόριθμος εκπαιδεύεται στο 90% των δεδομένων και ελέγχονται οι προβλέψεις του στο δείγμα επικύρωσης (10%). Για την αποφυγή τυχαίων διακυμάνσεων στην απόδοση του αλγόριθμου, η διαδικασία αυτή επαναλαμβάνεται άλλες 9 φορές χωρίζοντας τα δεδομένα με τυχαίο τρόπο σε 9 διαφορετικά δείγματα εκπαίδευσης και 9 διαφορετικά δείγματα επικύρωσης. Η τελική ακρίβεια του αλγόριθμου είναι ο μέσος όρος των 10 επιμέρους αποτελεσμάτων που έδωσε στα 10 δείγματα επικύρωσης.

Οι μετρικές που χρησιμοποιούνται για να ελέγξουμε την ποιότητα των ταξινομήσεων που παράγει ο αλγόριθμος είναι οι ακόλουθες:

- Ακρίβεια (accuracy): Ποσοστό των κειμένων που αποδόθηκε ορθώς στον συγγραφέα.
- Ορθότητα (precision): Το ποσοστό των ορθώς ταξινομημένων κειμέ-

νων σὲ σχέσηη μὲ τὸ σύνολο τῶν κειμένων πού ἀπέδωσε ὁ ἀλγόριθμος στὸν συγκεκριμένο συγγραφέα.

- Ἀνάκληση (recall): Τὸ ποσοστὸ τῶν ὀρθῶς ταξινομημένων κειμένων σὲ σχέσηη μὲ τὸν συνολικὸ ἀριθμὸ πού ἔχει γράψει ὁ συγκεκριμένος συγγραφέας.
- F1: Ὁ γεωμετρικὸς μέσος ὄρος τῆς ὀρθότητος καὶ τῆς ἀνάκλησης. Δίνεται ἀπὸ τὸν ἀκόλουθο τύπο:

$$F1 = 2 \cdot \frac{\text{Ὄρθότητα} \cdot \text{Ἀνάκληση}}{\text{Ὄρθότητα} + \text{Ἀνάκληση}}$$

Ὅλα τὰ ποσοστὰ πού θὰ ἀναφερθοῦν στὰ πειράματα ταξινομήσεως πού ἀκολουθοῦν εἶναι ὁ μέσος ὄρος τῶν παραπάνω μετρικῶν στὶς 10 ἐπαναλήψεις τῆς διασταυρούμενης ἐπικύρωσης.

9. Πειράματα ἀναγνώρισης πατρότητας

Πρὶν προβοῦμε στὴ χρῆση τῶν ΜΔΥ γιὰ τὴν ἀπόδοση πατρότητας στὶς ἀνώνυμες μεταφράσεις, εἶναι ἀπαραίτητο νὰ ἐλέγξουμε τὴν ἀποτελεσματικότητά τους σὲ μιὰ σειρὰ ἀπὸ προβλήματα ἀπόδοσης τῶν ὁποίων τὸ ἀποτέλεσμα γνωρίζουμε ἤδη. Γιὰ τὸν λόγο αὐτὸ ὀργανώθηκε ἓνα πείραμα προσομοίωσης μὲ τὴ χρῆση κειμένων τῆς Νέας Ἑλληνικῆς. Ἐνας ἐκ τῶν τριῶν συγγραφέων τοῦ παρόντος ἄρθρου (Τιτίκα Δημητρούλια) εἶναι ἐπαγγελματίας μεταφραστὴς καὶ παράλληλα συγγραφέας κειμένων σχετικῶν μὲ τὸν πολιτισμὸ καὶ τὸ βιβλίο σὲ ἐφημερίδα μεγάλης κυκλοφορίας (*Ἡ Καθημερινή*). Αὐτὴ ἡ εὐτυχὴ συγκυρία μᾶς ἐπέτρεψε νὰ ἀναδημιουργήσουμε ἓνα ΗΚΣ πού μοιάζει σημαντικὰ ὡς πρὸς αὐτὸ πού χρησιμοποιοῦμε γιὰ τὸν Παπαδιαμάντη. Ἐπιπλέον, ὅπως καὶ στὸν Παπαδιαμάντη, ἔτσι καὶ ἐδῶ ἀναπτύξαμε ἓνα ΗΚΣ συγχρόνων συγγραφέων πού ἀποτελεῖται ἀπὸ συγγραφεῖς πού ἀρθρογραφοῦν συστηματικὰ στὸν Τύπο μὲ ἀνάλογο κειμενικὸ θέμα καὶ τύπο μὲ τὴν Τ. Δημητρούλια.

Σὲ αὐτὸ τὸ ΗΚΣ ὑπολογίσαμε τοὺς ἴδιους ὑφομετρικὸς δείκτες καὶ ἀξιοποιήσαμε τὸν ἴδιο ἀλγόριθμο γιὰ νὰ ἐκτιμήσουμε τὴν ἀποτελεσματικότητα τῆς μεθόδου στὴν ἀναγνώριση τῶν μεταφράσεων τῆς Τ. Δημητρούλια. Τὸ συγκεκριμένο πείραμα στέφθηκε μὲ ἐπιτυχία ἀφοῦ οἱ μεταφράσεις τῆς ἀναγνωρίστηκαν μὲ μεσοσταθμικὴ ἀκρίβεια 97,4%, γεγονός πού πιστοποίησε τὴν ἐγκυρότητα τῆς μεθοδολογικῆς μας προσέγγισης.

Ἐν συνέχειᾳ ἐπιστρέψαμε στὸ ΗΚΣ τοῦ Παπαδιαμάντη καὶ τῶν συγχρόνων του καὶ ὀργανώσαμε τρία διαφορετικὰ πειράματα μὲ κειμενικὰ τεμάχια, τῶν ὁποίων ἤδη γνωρίζαμε τὴν πατρότητά τους. Εἰδικότερα ἐλέγξαμε τὴν ἀπόδοση τῶν ΜΔΥ στὰ ἀκόλουθα:

1. Διάκριση μεταξύ αὐθεντικῶν ἔργων τοῦ Παπαδιαμάντη καὶ αὐθεντικῶν ἔργων τῶν συγχρόνων του.
2. Διάκριση μεταξύ συνολικοῦ ἔργου τοῦ Παπαδιαμάντη (αὐθεντικὰ ἔργα + βέβαιες μεταφράσεις) καὶ αὐθεντικῶν ἔργων τῶν συγχρόνων του.
3. Διάκριση μεταξύ αὐθεντικῶν ἔργων τοῦ Παπαδιαμάντη ~ βέβαιων μεταφράσεων τοῦ Παπαδιαμάντη ~ αὐθεντικῶν ἔργων τῶν συγχρόνων του.

Στὸ πρῶτο πείραμα ὁ στόχος ἦταν νὰ ἐλεγχθεῖ κατὰ πόσο οἱ ΜΔΥ διακρίνουν τὰ κείμενα ποὺ ἔχει γράψει ὁ Παπαδιαμάντης (ὄχι μεταφράσεις) καὶ κείμενα ἄλλων συγγραφέων τῆς ἐποχῆς. Τὰ ἀποτελέσματα τῆς ταξινομήσεως τοῦ ἀλγόριθμου δίνονται στὸν παρακάτω πίνακα (Πίνακας 1):

Πίνακας 1: Πίνακας ταξινομήσεως τοῦ πρώτου πειράματος

Σύγχρονοι	Παπαδιαμάντης	< = Ταξινομήθηκαν ὡς
227	2	Σύγχρονοι
1	737	Παπαδιαμάντης

Ὁ παραπάνω πίνακας διαβάζεται ἀπὸ δεξιὰ πρὸς τὰ ἀριστερά. Οἱ ΜΔΥ ταξινομήσαν σωστὰ 227 ἀπὸ τὰ 229 κειμενικὰ τεμάχια ποὺ ἀνήκαν στὴν κατηγορία «Σύγχρονοι τοῦ Παπαδιαμάντη» (Ὁρθότητα = 0,996, Ἀνάκληση = 0,991, F1 = 0,993) καὶ 737 ἀπὸ τὰ 738 κειμενικὰ τεμάχια ποὺ ἀνήκαν στὸν Παπαδιαμάντη (Ὁρθότητα = 0,997, Ἀνάκληση = 0,999, F1 = 0,998). Ἡ μεσοσταθμικὴ ἀκρίβεια τῆς ταξινομήσεως ἀνῆλθε στὸ 0,997 ἢ 99,7%, ἀποτέλεσμα ποὺ δείχνει ὅτι ὁ συγκεκριμένος ἀλγόριθμος χρησιμοποιοῦντας τὰ ὑφομετρικὰ χαρακτηριστικὰ ποὺ ὑπολογίσαμε ἔχει μάθει σχεδὸν τέλεια τὸ παπαδιαμαντικὸ ὑφομετρικὸ προφίλ καὶ τὸ διακρίνει σχεδὸν πάντα ἀπὸ τὸ ὑφομετρικὸ προφίλ ἄλλων συγγραφέων τῆς ἐποχῆς του.

Στὸ δεύτερο πείραμα αὐξήσαμε τὴν ἐκθεση τοῦ ἀλγόριθμου στὸ παπαδιαμαντικὸ ἔργο προσθέτοντας καὶ τὶς βέβαιες μεταφράσεις του. Τὰ ἀποτελέσματα τῆς ταξινομήσεως τοῦ ἀλγόριθμου δίνονται στὸν παρακάτω πίνακα (Πίνακας 2):

Πίνακας 2: Πίνακας ταξινόμησης του δεύτερου πειράματος

Σύγχρονοι	Παπαδιαμάντης	< = Ταξινομήθηκαν ως
224	5	Σύγχρονοι
2	1197	Παπαδιαμάντης

Οι ΜΔΥ ταξινόμησαν όρθα τα 224 από τα 229 κειμενικά τεμάχια που άνηκαν στην κατηγορία «Σύγχρονοι του Παπαδιαμάντη» (Όρθότητα = 0,991, Ανάκληση = 0,978, F1 = 0,985) και 1197 από τα 1199 κειμενικά τεμάχια που άνηκαν στον Παπαδιαμάντη (Όρθότητα = 0,996, Ανάκληση = 0,998, F1 = 0,997). Η μεσοσταθμική ακρίβεια τής ταξινόμησης άνηλθε στο 0,995 ή 99,5%.

Τα συγκεκριμένο αποτέλεσμα άξιζει να σχολιαστεί τουλάχιστον ως προς δύο σημεία:

1. Η πρόσθεση του μεταφραστικού λόγου στα υπόλοιπα έργα του Παπαδιαμάντη δέν επηρεάζει σημαντικά την ακρίβεια τής ταξινόμησης. Τό γεγονός αυτό είναι ιδιαίτερα ένθαρρυντικό, γιατί καταδεικνύει ότι ό συνδυασμός άλγόριθμου και ύφομετρικών χαρακτηριστικών τής παρούσας έρευνας μπορεί να ένοποιήσει τό ύφομετρικό προφίλ τών μεταφράσεων του Παπαδιαμάντη με τα υπόλοιπα έργα του και να υπερκεράσει τις ιδιαιτερότητες τής μεταφραστικής ύφομετρικής ταυτότητας που ήδη έπισημάνθησαν στην ένότητα 5.α
2. Η μείωση τής τιμής τής Ανάκλησης στην κατηγορία «Σύγχρονοι του Παπαδιαμάντη» κατά 0,013 ή 1,3% είναι μικρή και μπορεί να έξηγηθεί από τό γεγονός ότι σε αυτό τό πείραμα όι ΜΔΥ τροφοδοτήθηκαν με σημαντικά μεγαλύτερο όγκο κειμενικών τεμαχίων από τόν Παπαδιαμάντη (άύξηση 62%). Η άύξημένη έκθεση του άλγόριθμου σε περισσότερα κειμενικά δείγματα τής κατηγορίας «Παπαδιαμάντης» σε σχέση με την κατηγορία «Σύγχρονοι του Παπαδιαμάντη» προκάλεσε «ύπερεκπαίδευση» (overtraining) με αποτέλεσμα ό άλγόριθμος να άύξησει έλαφρά την τάση του να κατηγοριοποιεί κείμενα άλλων ως Παπαδιαμαντικά. Πρέπει να υπογραμμιστεί όμως ότι αυτή ή τάση είναι πολύ μικρή και δέν επηρεάζει ουσιαστικά τα αποτελέσματα τής κατηγοριοποίησης, άφου συνολικά ή ακρίβεια έπεσε μόνον κατά 0,02%, ποσοστή μη σημαντικό στατιστικά.

Στό τρίτο πείραμα προχωρούμε σε μιá τριμερή ταξινόμηση. Έκπαιδεύουμε τις ΜΔΥ σε τρεις ξεχωριστές κατηγορίες. Ειδικότερα, χωρίζουμε

τά κειμενικά τεμάχια σέ έργα τοῦ Παπαδιαμάντη πού δέν εἶναι μεταφράσεις, σέ βέβαιες μεταφράσεις τοῦ Παπαδιαμάντη καί σέ έργα συγχρόνων τοῦ Παπαδιαμάντη. Τά ἀποτελέσματα τῆς τριμεροῦς ταξινόμησης δίνονται στὸν παρακάτω πίνακα (Πίνακας 3):

Πίνακας 3: Πίνακας ταξινόμησης τοῦ τρίτου πειράματος

Σύγχρονοι	Παπαδιαμάντης	Βέβαιες μεταφράσεις Παπαδιαμάντη	< = Ταξινομήθηκαν ὡς
227	2	0	Σύγχρονοι
1	737	0	Παπαδιαμάντης
0	0	461	Βέβαιες μεταφράσεις Παπαδιαμάντη

Οἱ ΜΔΥ ταξινόμησαν ὀρθά τὰ 227 ἀπὸ τὰ 229 κειμενικά τεμάχια πού ἀνήκαν στὴν κατηγορία «Σύγχρονοι τοῦ Παπαδιαμάντη» (Ὄρθότητα = 0,996, Ἀνάκληση = 0,991, $F1 = 0,993$), 737 ἀπὸ τὰ 738 κειμενικά τεμάχια πού ἀνήκαν στὸν Παπαδιαμάντη (χωρὶς τίς βέβαιες μεταφράσεις του) (Ὄρθότητα = 0,997, Ἀνάκληση = 0,999, $F1 = 0,998$). Ταξινόμησαν ἐπίσης ὅλα τὰ κειμενικά τεμάχια πού ἀνήκαν στίς βέβαιες μεταφράσεις τοῦ Παπαδιαμάντη στὴ σωστή τους κατηγορία (461/461) ἐπιτυγχάνοντας τέλειες μετρικὲς ταξινόμησης (Ὄρθότητα = 1, Ἀκρίβεια = 1, $F1 = 1$). Ἡ μεσοσταθμικὴ ἀκρίβεια τῆς ταξινόμησης ἀνῆλθε στὸ 0,998 ἢ 99,8%.

Τά ἀποτελέσματα στὸ τρίτο πείραμα δείχνουν ὅτι οἱ βέβαιες μεταφράσεις τοῦ Παπαδιαμάντη μποροῦν νὰ ἀναγνωριστοῦν ἀπὸ τὸν ἀλγόριθμο ἀκόμα καλύτερα καὶ ἀπὸ τὰ ἄλλα έργα του. Αὐτὸ δείχνει ὅτι ὁ συγκεκριμένος ἀλγόριθμος (ΜΔΥ) σέ συνδυασμὸ μὲ τὰ ὑφομετρικὰ χαρακτηριστικὰ πού χρησιμοποιήθηκαν μπορεῖ νὰ λειτουργήσει εὐέλκτα καὶ εἴτε νὰ ἐνοποιήσει τὸ ὑφομετρικὸ προφίλ τῶν μεταφράσεων καὶ τῶν ἄλλων ἔργων τοῦ Παπαδιαμάντη, εἴτε νὰ ἐκπαιδευτεῖ νὰ ἀναγνωρίζει τίς μεταφράσεις ὡς ξεχωριστὴ κατηγορία καὶ νὰ τίς ἀντιδιαστέλλει ὡς πρὸς τὰ ἄλλα έργα τοῦ Παπαδιαμάντη καὶ τὰ έργα τῶν συγχρόνων του. Αὐτὸ ἐπιτυγχάνεται λόγῳ τοῦ μεγάλου ἀριθμοῦ τῶν ὑφομετρικῶν χαρακτηριστικῶν πού χρησιμοποιοῦμε στὴν ταξινόμηση. Οἱ ΜΔΥ στὴν πρώτη περίπτωση ἐπιλέγουν συγκεκριμένο ὑποσύνολο χαρακτηριστικῶν γιὰ νὰ «μάθουν» νὰ ἐνοποιοῦν τὸ συνολικὸ ἔργο τοῦ Παπαδιαμάντη (μεταφράσεις καὶ μὴ), ἐνῶ στὴ δεύ-

τερη περίπτωση αξιοποιούν διαφορετικό ύποσύνολο χαρακτηριστικών για να διακρίνουν τις βέβαιες μεταφράσεις από τα άλλα έργα του Παπαδιαμάντη και των συγχρόνων του.

10. Απόδοση της πατρότητας των ανώνυμων μεταφράσεων

Τὰ πειράματα που περιγράφηκαν στην προηγούμενη ενότητα έδειξαν ότι οι ΜΔΥ έχουν εκπαιδευθεί ικανοποιητικά και μπορούν να διακρίνουν τὸ ύφομετρικὸ προφίλ τοῦ Παπαδιαμάντη ἀπὸ αὐτὸ τῶν συγχρόνων του συγγραφέων. Ἐπίσης, ἀπὸ τὰ ἀποτελέσματα τοῦ δευτέρου καὶ τοῦ τρίτου πειράματος εἶδαμε ὅτι ὁ ἀλγόριθμος μπορεῖ νὰ λειτουργήσῃ ἐξίσου καλὰ εἴτε ὡς πρὸς τὴν αὐτόνομη ἀναγνώριση τοῦ μεταφραστικοῦ λόγου (Πείραμα 3), εἴτε ὡς πρὸς τὴν ἐνοποιημένη ἀπεικόνιση τοῦ παπαδιαμαντικοῦ ὕφους καὶ τὴν ἀντιδιαστολή του μετὰ τὸ ὑφομετρικὸ προφίλ ἄλλων συγγραφέων τῆς ἐποχῆς.

Γιὰ τὸν ἔλεγχο τῆς πατρότητας τῶν ἀνώνυμων μεταφράσεων χρησιμοποίησαμε τὸ μοντέλο ἐκπαίδευσης που περιγράφεται στὸ δεῦτερο πείραμα καὶ συνίσταται στὴν ἐκπαίδευση τῶν ΜΔΥ μετὰ ὅλα τὰ κειμενικὰ τεμάχια τοῦ Παπαδιαμάντη (βέβαιες μεταφράσεις καὶ τὰ ἄλλα ἔργα του) σὲ ἀντιδιαστολή μετὰ τὰ ἔργα τῶν συγχρόνων του. Ἐν συνεχείᾳ οἱ ΜΔΥ κλήθηκαν νὰ προβλέψουν σὲ ποιά ἀπὸ τὶς δύο κατηγορίες (Παπαδιαμάντης ~ Σύγχρονοι τοῦ Παπαδιαμάντη) ἀνήκουν τὰ κειμενικὰ τεμάχια τῶν ἀνώνυμων μεταφράσεων. Τὰ ἀποτελέσματα δίνονται στὸν παρακάτω πίνακα (Πίνακας 4):

Πίνακας 4: Ἀποτελέσματα πρόβλεψης τοῦ ἀλγόριθμου
στὶς ἀνώνυμες μεταφράσεις

Ἀνώνυμη Μετάφραση	Πρόβλεψη		Ποσοστὸ ἀπόδοσης στὸν Παπαδιαμάντη
	Παπαδιαμάντης (κειμ. τεμάχια)	Σύγχρονοι τοῦ Παπαδιαμάντη (κειμ. τεμάχια)	
Κοσπέ, «Ἄρριέτα»	20	0	100%
Γόγολ, «Ἡ μύτη»	6	0	100%
Γόγολ, «Ἡ προμερὰ ἐκδίκησις»	6	3	67%
Τουργκιένεφ, «Ὁ ἐπαρχιακὸς ἰατρός»	3	1	75%

Πόε, «Τὰ διηγήματα τῆς πανώλους - Τὸ προσωπεῖον τοῦ Ἐρυθροῦ Θανάτου»	0	2	0%
Μωπασσάν, «Ὁ δασοφύλαξ»	0	2	0%

Ὁ ἀλγόριθμος ἀναγνωρίζει τὸ ὑφομετρικὸ προφίλ τῆς μετάφρασης τοῦ Κοππέ «Ἀρριέτα», ὡς ἀμιγῶς παπαδιαμαντικὸ. Ἐπίσης, ὑφομετρικὰ ἀποδίδονται ὅλα τὰ κειμενικὰ τεμάχια τῆς μετάφρασης τοῦ Γόγολ «Ἡ μύτη» στὸ ὑφομετρικὸ προφίλ τοῦ Παπαδιαμάντη.

Στὶς μεταφράσεις «Ὁ ἐπαρχιακὸς ἰατρός» τοῦ Τουργκιένεφ καὶ «Ἡ τρομερὰ ἐκδίκησις» τοῦ Γόγολ, οἱ ΜΔΥ ἀποδίδουν τὴν πλειονότητα τῶν κειμενικῶν τεμαχίων στὸ παπαδιαμαντικὸ προφίλ (75% καὶ 67% ἀντίστοιχα). Ὡστόσο, δεδομένου τοῦ πολὺ ὑψηλοῦ ποσοστοῦ ἀκρίβειας ποὺ παρουσίασε ὁ ἀλγόριθμος στὰ πειράματα ποὺ προηγήθηκαν (δὲν ἔπεσε κάτω ἀπὸ 99,5%), τὰ ποσοστὰ ἀπόδοσης στὸ ὑφομετρικὸ προφίλ τοῦ Παπαδιαμάντη εἶναι συγκριτικὰ πολὺ μικρότερα. Γιὰ τὸν λόγο αὐτὸ ὑποστηρίζουμε ὅτι οἱ συγκεκριμένες μεταφράσεις δὲν μποροῦν νὰ συσχετιστοῦν μὲ τὸ ὑφομετρικὸ προφίλ τοῦ Παπαδιαμάντη, παρὰ τὸ γεγονὸς ὅτι παρουσιάζονται κάποιες ὁμοιότητες.

Τέλος, οἱ μεταφράσεις «Τὰ διηγήματα τῆς πανώλους - Τὸ προσωπεῖον τοῦ Ἐρυθροῦ Θανάτου» τοῦ Πόε καὶ «Ὁ δασοφύλαξ» τοῦ Μωπασσάν, ὑφομετρικὰ ταξινομοῦνται μὲ βεβαιότητα ὡς μὴ παπαδιαμαντικῆς.

Συμπεράσματα - προοπτικὲς

Ἡ ὀργάνωση καὶ ἡ ἐφαρμογὴ τοῦ ἐρευνητικοῦ αὐτοῦ προγράμματος, λόγῳ τοῦ καινοτόμου χαρακτήρα του, μιὰ καὶ εἶναι τὸ πρῶτο στὴν Ἑλλάδα καὶ ἀπὸ τὰ ἐλάχιστα διεθνῶς μὲ τὸ ὁποῖο ἐπιχειρεῖται ἡ χρῆση ὑφομετρικῶν μεθόδων γιὰ τὴν ἀναγνώριση τῆς πατρότητας ὄχι πρωτότυπων κειμένων ἀλλὰ μεταφράσεων, ἀπαιτοῦσε προσεκτικὰ καὶ διαρκῶς σὲ κάθε φάση ἐλεγχόμενα καὶ ἀναπροσαρμοζόμενα βήματα. Ἀπὸ τὴν ἄλλη πλευρά, τὸ ἐγχείρημα ὑπῆρξε καὶ ἐξαιρετικὰ χρονοβόρο, ὄχι μόνον γιὰτὶ μεγάλα σώματα κειμένων ἀπὸ αὐτὰ ποὺ χρησιμοποιήθηκαν ἔπρεπε νὰ ψηφιοποιηθοῦν, ὅπως π. χ. ἡ μετάφραση τοῦ *Ἐγκλημα καὶ Τιμωρία*, βάρους ποὺ ἔπεσε στοὺς ὤμους μας, ἀλλὰ καὶ γιὰτὶ ὅλα τὰ σώματα, ἀκόμη καὶ αὐτὰ ποὺ μᾶς δόθηκαν ψηφιοποιημένα, ὅπως τὰ πρωτότυπα ἔργα τοῦ Παπαδιαμάν-

τη, έπρεπε να γίνονταν αντικείμενο ειδικής κειμενικής επεξεργασίας, προκειμένου να όμογενοποιηθούν, για να γίνονταν αναγνωρίσιμα από τὰ λογισμικά που χρησιμοποιούσαμε.

Όπως φάνηκε από τον παραπάνω Πίνακα 4, ό έλεγχος τής πατρότητας των έξι αυτών πρώτων μεταφράσεων έδειξε, πρώτον, ότι οι μεταφράσεις του διηγήματος του Γκόγκολ «Η μύτη» και του Κοππέ «Άρριέτα» είναι παπαδιαμαντικές, με ποσοστό απόδοσης 100%. Το γεγονός ότι το πρόγραμμά μας αναγνώρισε ως άμιγώς παπαδιαμαντική τή μετάφραση τής «Άρριέτας» του Κοππέ, τήν όποία ό Ν. Δ. Τριανταφυλλόπουλος και ή Λαμπρινή Τριανταφυλλοπούλου είχαν αποδώσει και αυτοί στον Παπαδιαμάντη με τὰ δικά τους, δύσκολα άμφισβητούμενα, κριτήρια (Τριανταφυλλόπουλος - Τριανταφυλλοπούλου, 2006), ένισχύει τήν άξιοπιστία του ως προς τήν απόδοση στον Παπαδιαμάντη τής μετάφρασης του διηγήματος του Γκόγκολ «Η μύτη».

Δεύτερον, ύστερα από τον έλεγχο μας, οι μεταφράσεις «Τὰ διηγήματα τής πανώλους - Το προσωπίον του Έρυθρου Θανάτου» του Πόε και «Ό δασοφύλαξ» του Μωπασσάν, ταξινομήθηκαν με βεβαιότητα, όπως είδαμε, ως μη παπαδιαμαντικές. Και πάλι, ή σύμπτωση των εύρημάτων μας ως προς τήν μη παπαδιαμαντική πατρότητα τής μετάφρασης του διηγήματος του Μωπασσάν «Ό δασοφύλαξ» με τή θέση του Ν. Δ. Τριανταφυλλόπουλου και τής Λαμπρινής Τριανταφυλλοπούλου (Maupassant, 2003), ένισχύει τήν άξιοπιστία του ύφομετρικού εργαλείου μας ως προς τὰ εύρήματά του για τήν μη παπαδιαμαντική μετάφραση του έργου «Τὰ διηγήματα τής πανώλους - Το προσωπίον του Έρυθρου Θανάτου» του Πόε.

Τέλος, το γεγονός ότι με το ύφομετρικό εργαλείο μας οι μεταφράσεις «Ό έπαρχιακός ιατρός» του Τουργκιένεφ και «Η τρομερά έκδίκησις» του Γκόγκολ αποδόθηκαν στον Παπαδιαμάντη με ποσοστό 75% και 67% αντίστοιχα, ποσοστό ιδιαίτερα χαμηλότερο από τὰ ποσοστά ακρίβειας με τὰ όποία αποδόθηκαν ή δέν αποδόθηκαν στον Παπαδιαμάντη οι τέσσερις άλλες μεταφράσεις, μάς αναγκάζει να άμφιβάλλουμε για τήν παπαδιαμαντική πατρότητα των δύο αυτών μεταφράσεων.

Τελειώνοντας θά θέλαμε να έπισημάνουμε ότι το έγχείρημά μας και τὰ αποτελέσματά του μπορούν βέβαια να τροφοδοτήσουν και ποικίλες άλλες προσεγγίσεις του Παπαδιαμάντη —άλλά και άλλων συγγραφέων— με άφετηρία έρευνητικά ενδιαφέροντα άλλων κλάδων, όπως είναι π. χ. ή Μεταφρασεολογία, πέραν δηλαδή τής Συγκριτικής Φιλολογίας, στο πλαίσιο

της οποίας τοποθετήσαμε από την αρχή τον στόχο αυτής της έρευνας. Πάντως, σέ ό,τι μάς αφορά και έχοντας διαρκώς κατά νοῦν τὸ πρωτότυπο ἔργο τοῦ Παπαδιαμάντη, ἐλπίζουμε νὰ μπορέσουμε νὰ συνεχίσουμε τὴν ἔρευνά μας μὲ τὸν ἔλεγχο τῆς πατρότητας καὶ ἄλλων μεταφράσεων, ἀρχίζοντας ἀπὸ τὶς μεταφράσεις ἔργων σημαντικῶν καὶ γνωστῶν ξένων λογοτεχνῶν κατὰ τὴ χρονολογικὴ σειρὰ τῆς δημοσίευσής τους καὶ προχωρώντας, σταδιακά, μὲ τὸν ἔλεγχο τῶν μεταφράσεων λιγότερο γνωστῶν ἢ καὶ ἀσήμαντων καὶ ἀγνωστῶν μερικῆς φορὸς λογοτεχνικῶν ἔργων.

ΠΑΡΑΡΤΗΜΑ

Σύσταση καὶ μέγεθος τῶν ΗΣΚ ποὺ χρησιμοποιήθηκαν στὴν ἔρευνα

Πίνακας 5: Ἀριθμὸς κειμένων καὶ λέξεων τῶν ΗΣΚ τῆς ἔρευνας

ΗΣΚ	Αριθμὸς κειμένων (τεμάχια 1000 λέξεων)	Ἀριθμὸς λέξεων
Σύγχρονοι τοῦ Παπαδιαμάντη	16 (229)	233.226
Παπαδιαμάντης (Ἄπαντα)	181 (738)	811.289
Βέβαιες μεταφράσεις τοῦ Παπαδιαμάντη	4 (498)	462.674
Ἀνώνυμες μεταφράσεις	6 (43)	45.306
Σύνολο	207 (1508)	1.552.495

Πίνακας 6: Σύσταση καὶ μέγεθος κειμένων στὸ ΗΣΚ

«Σύγχρονοι τοῦ Παπαδιαμάντη»

Συγγραφέας	Τίτλος ἔργου	Λέξεις
Δροσίνης, Γεώργιος	Τὸ βοτάνι τῆς ἀγάπης (ἀπόσπασμα)	1.206
Καρκαβίτσας, Ἀνδρέας	Διηγήματα	46.089
Καρκαβίτσας, Ἀνδρέας	Ἡ λυγερή	40.490
Κονδυλάκης, Ἰωάννης	Ὁ Πατούχας	55.847
Μητσάκης, Μιχαήλ	Αυτόχειρ	5.597
Μητσάκης, Μιχαήλ	Τὸ φίλημα	681
Μητσάκης, Μιχαήλ	Τὸ ὀμορφότερο πράγμα τοῦ κόσμου	481

Μητσάκης, Μιχαήλ	Ή Παναγιά ή Μεγαλομάτα	695
Μωραϊτίδης, Ἀλέξανδρος	Ἀρφανούλα	5.320
Μωραϊτίδης, Ἀλέξανδρος	Ἀθωνας	1.363
Μωραϊτίδης, Ἀλέξανδρος	Ἵπεράνω τῶν νεφελῶν	2.791
Βιζυηνός, Γεώργιος	Τὸ ἀμάρτημα τῆς μητρός μου	7.884
Βιζυηνός, Γεώργιος	Ποῖος ἦτον ὁ φονεὺς τοῦ ἀδελφοῦ μου	15.155
Βιζυηνός, Γεώργιος	Μοσκῶβ Σελήμ	17.508
Βιζυηνός, Γεώργιος	Αἱ συνέπειαι τῆς παλαιᾶς ἱστορίας	20.983
Βιζυηνός, Γεώργιος	Τὸ μόνον τῆς ζωῆς του ταξίδιον	11.136

Πίνακας 7: Σύσταση καὶ μέγεθος κειμένων στὸ ΗΣΚ
«Βέβαιες μεταφράσεις τοῦ Παπαδιαμάντη»

Συγγραφέας	Τίτλος	Λέξεις
Ντοστογιέφσκι, Φιοντόρ	Τὸ ἔγκλημα καὶ ἡ τιμωρία	170.857
Κέιν, Χώλλ	Ὁ Μαξιώτης	196.251
Ὅνέ, Ζώρζ	Ὁ ἰατρός Ραμώ	69.468
Δωδέ, Ἀλφόνσος	Ταρταρίνος ὁ ἐκ Ταρασκῶνος	26.098

Πίνακας 8: Σύσταση καὶ μέγεθος κειμένων στὸ ΗΣΚ
«Ἀνώνυμες μεταφράσεις»

Συγγραφέας	Τίτλος	Ἐφημερίδα	Ἡμερομηνία	Λέξεις
Γουὸ δὲ Μωπασσάν	Ὁ δασοφύλαξ	Ἐφημερίς	20-22/12/1886	2.036
Ἰβάν Τουργκιένεφ	Ὁ ἐπαρχιακὸς ἰατρός	Ἐφημερίς	23-24/12/1886	4.362
Φραγκίσκου Κοπέ	Ἀρριέτα	Ἐφημερίς	5-19/8/1889	20.066
Νικολάου Γόγολ	Ἡ τρομερὰ ἐκδίκησις. (Ρωσσικὸν μυθιστόρημα)	Τὸ Ἄστυ	3-15/3/1899	6.953
Ἐδγάρδου Πόε	Τὰ διηγήματα τῆς πανώ- λους. Τὸ προσωπεῖον τοῦ Ἐρυθροῦ Θανάτου	Τὸ Ἄστυ	7-8/9/1899	2.280
Νικολάου Γόγολ	Ἡ μύτη	Τὸ Ἄστυ	25/11-2/12/ 1899	9.609

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

- Allison, B., & Guthrie, L. (2008). Authorship attribution of E-Mail: Comparing classifiers over a new corpus for evaluation. *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA).
- Baker, M. (1995). Corpora in Translation Studies. An Overview and Suggestions for Future. *Research. Target*, 7 (2), 223-243.
- Baker, M. (1996). Corpus-based Translation Studies. The Challenges that Lie Ahead. In: Somers, Harold (Ed.), *Terminology, LSP and Translation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 175-186.
- Bosseaux, C. (2004). Translating point of view: A corpus-based study. *Corpus-based Translation Studies: Research Applications, special issue of Language Matters, Studies in the Languages of Africa*, 35 (1), guest edited by Alet Kruger, 39-53.
- Bourdieu, P. (1992). *Les Règles de l'art. Genèse et structure du champ littéraire*. Paris: Seuil.
- Coyotl-Morales, R., Villasenor-Pineda, L., Montes-y-Gomez, M., & Rosso, P. (2006). Authorship Attribution Using Word Sequences. In: J. Martinez-Trinidad, J. Carrasco Ochoa & J. Kittler (Eds.), *Progress in Pattern Recognition. Σεμρά: Image Analysis and Applications* (Vol. 4225, pp. 844-853). Berlin/Heidelberg: Springer.
- Craig, Hugh (2004a). Stylistic Analysis and Authorship Studies. In: *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens and John Unsworth. Blackwell, 285.
- Craig, Hugh (2004b). Stylistic Analysis and Authorship Studies. In: *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens and John Unsworth. Blackwell, 273.
- Craig, Hugh (2004c). Stylistic Analysis and Authorship Studies. In: *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens and John Unsworth. Blackwell, 280.
- Diederich, J., Kindermann, J., Leopold, E., & Paass, G. (2003). Authorship Attribution with Support Vector Machines. *Applied Intelligence*, 19 (1), 109-123.
- Gouanvic, J.-M. (2006). Au-delà de la pensée binaire en traductologie: esquisse d'une analyse sociologique des positions traductives en traduction littéraire. *TTR: traduction, terminologie, rédaction*, 19 (1), 123-134.
- Grammenidis, S. (2011). Traduire: transcrire, ré-écrire ou ré-énoncer? *Travaux Linguistiques du CerLiCo*, 24, 179- 199.
- Grieve, J. W. (2007). Quantitative Authorship Attribution: an Evaluation of Tech-

- niques. *Literary and Linguistic Computing*, 22 (3), 251-270.
- Hoover, D. (2003). Another Perspective on Vocabulary Richness. *Computers and the Humanities*, 37, 151-178.
- Juola, P., Sofko, J., & Brennan, P. (2006). A Prototype for Authorship Attribution Studies. *Literary and Linguistic Computing*, 21 (2), 169-178.
- Kenny, Anthony (1982). *The Computation of Style*. Oxford: Pergamon Press, 12-13.
- Kenny, D. (2001). *Lexis and Creativity in Translation*. A Corpus-based Study. Manchester, UK: St Jerome.
- Koppel, M., & Schler, J. (2004). Authorship Verification as a one-class Classification Problem. *Proceedings of 21st International Conference on Machine Learning*, July 2004 (pp. 489-495). Banff, Canada.
- Koppel, M., Schler, J., & Argamon, S. (2009). Computational Methods in Authorship Attribution. *Journal of the American Society for Information Science and Technology*, 60 (1), 9-26. doi: 10.1002/asi.20961.
- Koppel, M., Schler, J., & Argamon, S. (2011). Authorship Attribution in the Wild. *Language Resources and Evaluation*, 45 (1), 83-94. doi: 10.1007/s 10579-009-9111-2.
- Koppel, M., Schler, J., & Boucek-Dokow, E. (2007). Measuring Differentiability: Unmasking Pseudonymous Authors. *Journal of Machine Learning Research*, 8, 1261-1276.
- Lancashire, Ian (2004). Cognitive Stylistics and the Literary Imagination. In: *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens and John Unsworth. Blackwell, 406.
- Laviosa, S. (1997). How Comparable Can 'Comparable Corpora' Be? *Target*, 9 (2), 289-319.
- Leech, G. N. and Short, M. H. (1981). *Style in Fiction*, London: Longman
- Luyckx, K., & Daelemans, W. (2011). The effect of Author Set Size and Data Size in Authorship Attribution. *Literary and Linguistic Computing*, 26 (1), 35-55. doi: 10.1093/lc/fqq013.
- Maingueneau, D. (2004). Linguistique et littérature : le tournant discursive. *Vox poetica*, www.vox-poetica.org, *ανάρτηση* 5 'Ιουνίου 2004 (15.3.2012).
- Manevitz, L. M., & Yousef, M. (2001). One-Class SVMs for Document Classification. *Journal of Machine Learning Research*, 2, 139-154.
- Maupassant, Guy de (2003). *Ἡ κληρονομία*. Μετάφρασις Ἀλ. Παπαδιαμάντης. Πρόλογος Σταῦρος Ζουμπουλάκης. Φιλολογικὴ ἐπιμέλεια Ν. Δ. Τριανταφυλλόπουλος - Α. Τριανταφυλλοπούλου. Ἀθήνα: Δόμος, 115, σημ. 3.
- Mikros, G. K. (2007). *Authorship Attribution Using Discriminant Function Analysis: Exploring Literary Style Variation in Five Modern Greek Novels*. Paper

- presented at the 5th Trier Symposium on Quantitative Linguistics, Trier, Germany.
- Mikros, G. K. (2011). *Author Profiling in Greek Blogs: Gender Identification Using Machine Learning Methods*. Paper presented in the QUALICO 2012, 26-29 Απριλίου 2012, Βελιγράδι (Σερβία).
- Miranda, G. A., & Calle, M. J. (2007). Function Words in Authorship Attribution Studies. *Literary and Linguistic Computing*, 22 (1), 49-66.
- Nenopoulou, T. (2005). L'expression d'(im)politesse en grec et en français. In: Hui-Lan Chao et Kyoto Kuroda (Eds), *Divergence dans la traduction entre les langues orientales et le français*, Bulag, 30, 123-137.
- Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Olohan, M. (2004). *Introducing Corpora in Translation Studies*. Oxford: Routledge.
- Platt, J. C. (1999). Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In: B. Scholkopf, C. J. C. Burges & A. J. Smola (Eds.), *Advances in kernel methods* (pp. 185-208). Cambridge: MIT Press.
- Rudman, J. (1997). The State of Authorship Attribution Studies: Some Problems and solutions. *Computers and the Humanities*, 31 (4), 351-365.
- Rybicki J. (2012). The Great Mystery of the (almost) Invisible Translator. Stylometry in translation. In: Michael P. Oakes and Meng Ji (Eds), *Quantitative Methods in Corpus-Based Translation Studies. A Practical Guide to Descriptive Translation Research*. University of Sunderland / University of Tokyo, Studies in Corpus Linguistics, 51, 231-248.
- Saldanha, G. (2004). Accounting for the Exception to the Norm: Split Infinitives in Translated English. In: *Corpus-Based Translation Studies: Research Applications, special issue of Language Matters, Studies in the Languages of Africa*, 35, (1), guest edited by Alet Kruger, 39-53.
- Saldanha G. (2011). Translator Style: Methodological Considerations. *The Translator*, 17, 1, 25-50.
- Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (1999). *Automatic authorship attribution*. Paper presented at the Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics, Bergen, Norway.
- Stein, B., Lipka, N., & Prettenhofer, P. (2011). Intrinsic Plagiarism Analysis. *Language Resources and Evaluation*, 45 (1), 63-82. doi: 10.1007/s10579-010-9115-y.
- Toury, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins.
- Tweedie, F. J., & Baayen, H. R. (1998). How Variable May a Constant Be? Meas-

- ures of Lexical Richness in Perspective. *Computers and the Humanities*, 32 (5), 323-352.
- Tymoczko, M. (1998). Computerized Corpora and the Future of Translation Studies. *Meta*, 43 (4), 652-660.
- Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer-Verlag.
- Venuti, L. (1995). *The Translator's Invisibility: A History of Translation*. London and New York: Routledge.
- Winters, M. (2004). F.Scott Fitzgerald's Die Schonen und Verdammten. A Corpus-based Study of Loan Words and Code Switches as Features of Translators's Style. In: *Corpus-Based Translation Studies: Research Applications, special issue of Language Matters, Studies in the Languages of Africa*, guest edited by Alet Kruger, 39-53.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, A: Elsevier.
- Zhao, Y., & Zobel, J. (2007). *Search with Style: Authorship Attribution in Classic Literature*. Paper presented at the Proceedings of the 30th Australasian Computer Science Conference (ACSC 2007), Ballarat, Australia.
- Παππᾶς, Φ. (2010). «Ίστορία τῶν μεταφράσεων καὶ τῶν μεταφραστῶν στὴν Ἑλλάδα (1830-1909)»: Ἀπόπειρα προσέγγισης. www.apiliotis.gr (15.3.2012).
- Πολίτου-Μαρμαρινοῦ, Ἑλένη (1996). «Παπαδιαμάντης, Μωπασσὰν καὶ Τσέχωφ. Ἀπὸ τὴ Σκιαθὸ στὴν Εὐρώπη», *Πρακτικὰ Α' Διεθνοῦς Συνεδρίου γιὰ τὸν Ἀλέξανδρο Παπαδιαμάντη (Σκιαθὸς 20-24 Σεπτεμβρίου 1991)*. Ἀθήνα: Δόμος, 423-456 [= Πολίτου-Μαρμαρινοῦ, Ἑλένη (2009). *Συγκριτικὴ Φιλολογία. Ἀπὸ τὴ θεωρία στὴν πράξη*. Ἀθήνα: Ἑλληνικὰ Γράμματα, 163-189].
- Τριανταφυλλόπουλος, Ν. Δ. - Τριανταφυλλοπούλου, Λ. (2006). «Τεκμήρια γιὰ τὴν παπαδιαμαντικὴ πατρότητα ἀνυπόγραφων μεταφράσεων». *Παπαδιαμαντικά Τετράδια. Ἀφιέρωμα στὸν μεταφραστὴ Παπαδιαμάντη*, 7, 100-101.