

Δικαστική Γλωσσολογία:
Αυτόματη Εξαγωγή Υφολογικών Γλωσσικών Χαρακτηριστικών

Κατερίνα Θ. Φραντζή
Τμήμα Μεσογειακών Σπουδών, Πανεπιστήμιο Αιγαίου
frantzi@rhodes.aegean.gr

Abstract: The aim of this paper is to facilitate authorship identification by applying corpus based techniques on electronic textual resources. Our domain of application is the set of proclamations and testimonies by the ‘17th November’ Greek terrorist group. The use of computational techniques provides efficient, systematic and precise information which is otherwise unavailable on the basis of human judgment alone. Focusing on the textual processing of the testimonies, we apply corpus linguistics analysis on each of the 17N members’ testimonies, to acquire the linguistic and extra-linguistic information characteristic of each person’s use of language.

Λέξεις-κλειδιά: γλωσσολογία σωμάτων κειμένων, δικαστική γλωσσολογία, εύρεση συγγραφέα, υφολογολογία, τρομοκρατία

1. Εισαγωγή

Το αντικείμενο της εργασίας αυτής εντάσσεται στο γνωστικό πεδίο της Δικαστικής Γλωσσολογίας και συγκεκριμένα στην αυτόματη αναγνώριση συγγραφέα (Olsson, 2004) με τη χρήση τεχνικών της γλωσσολογίας σωμάτων κειμένων. Η χρήση υπολογιστικών τεχνικών της γλωσσολογίας σωμάτων κειμένων παρέχει συστηματική, πλήρη και ακριβή πληροφορία, η απόκτηση της οποίας δεν θα ήταν δυνατή με την χρήση παραδοσιακών (μη υπολογιστικών) μεθόδων επεξεργασίας (Wynne & Stockwell 2006). Το πεδίο εφαρμογής της εργασίας αυτής είναι το σύνολο των προκηρύξεων της τρομοκρατικής οργάνωσης “17 Νοέμβρη” (“17N”). Το καλοκαίρι του 2002, μέλη της οργάνωσης 17N συνελήφθησαν και οδηγήθηκαν σε δίκη. Μέχρι τότε, 84 προκηρύξεις που έφεραν την υπογραφή της 17N είχαν χαρακτηριστεί γνήσιες. Ωστόσο, και μετά το τέλος των συλλήψεων, προκηρύξεις με την υπογραφή της 17N στάλθηκαν σε μέσα μαζικής ενημέρωσης. Ο τελικός στόχος της μελέτης είναι η αυτόματη αναγνώριση των συγγραφέων των προκηρύξεων. Η εύρεση των συγγραφέων των προκηρύξεων έχει άμεση σχέση με το αν έχουν συλληφθεί όλα τα μέλη της οργάνωσης αλλά και με το ποιος ή ποιοι από τους συλληφθέντες κατείχαν ηγετική θέση στη οργάνωση.

Βασιζόμενοι σε γλωσσολογικές αλλά και στατιστικές παραμέτρους, εξάγουμε πληροφορία η οποία χαρακτηρίζει το ύφος της γλώσσας του κειμένου. Έτσι, μπορούμε να χαρακτηρίσουμε κάθε άτομο από γλωσσολογικές και στατιστικές παραμέτρους που χαρακτηρίζουν με τη σειρά τους κείμενα που έχουν παραχθεί από το συγκεκριμένο άτομο (Frantzi 2006; Frantzi & Ananiadou 2006). Χρησιμοποιούμε τις απολογίες (και δευτερολογίες όπου αυτές υπάρχουν) των ατόμων-κατηγορουμένων σαν κείμενα που χαρακτηρίζουν τα άτομα αυτά. Καθώς κάθε προκήρυξη χαρακτηρίζεται από γλωσσολογικές και στατιστικές παραμέτρους, ο έλεγχος για ομοιότητες και διαφορές μεταξύ των απολογιών (και δευτερολογιών) και των προκηρύξεων δίνει το βαθμό με τον οποίο θα μπορούσε να συσχετιστεί μία προκήρυξη με μία απολογία (και δευτερολογία) και κατ’έκταση με τον πιθανό της συγγραφέα. Στην εργασία αυτή παρουσιάζουμε την εφαρμογή των παραπάνω τεχνικών στο σύνολο των απολογιών και δευτερολογιών της οργάνωσης 17N.

2. Γλωσσολογία Σωμάτων Κειμένων

Η γλωσσολογία σωμάτων κειμένων (corpus linguistics) επικεντρώνεται στη μελέτη της πραγματικής χρήσης της γλώσσας και όχι στη μελέτη του τι είναι θεωρητικά πιθανό σε μία γλώσσα. Οι τεχνικές της γλωσσολογίας σωμάτων κειμένων εφαρμόζονται σε μεγάλες συλλογές κειμένων για την αυτόματη εξαγωγή πληροφορίας διαφόρων τύπων. Η γνωστική αυτή περιοχή περιλαμβάνει μελέτες για τη μορφολογία, σύνταξη, σημασιολογία, φωνολογία, λεξικογραφία, ορολογία κ.ά. (McEnery & Wilson 2000). Μεταξύ των εφαρμογών της γλωσσολογίας σωμάτων κειμένων συγκαταλέγονται και η μελέτη της γλώσσας προφορικού και γραπτού λόγου και γλωσσικών ποικιλιών, το ύφος συγκεκριμένων συγγραφέων, οι διάλεκτοι κ.ά. Μια προσέγγιση βασισμένη στις τεχνικές της γλωσσολογίας σωμάτων κειμένων κάνει εκτενή χρήση του ηλεκτρονικού υπολογιστή σε ένα σώμα κειμένων (corpus), μία μεγάλη συλλογή πραγματικών κειμένων (Biber et al. 1998), χρησιμοποιώντας αυτόματες αλλά και διαδραστικές τεχνικές. Ο υπολογιστής κάνει δυνατή την αποθήκευση και επεξεργασία μεγάλου όγκου κειμένων καθώς και τη γρήγορη παραγωγή ακριβούς και πλήρους πληροφορίας η οποία δεν θα μπορούσε να παραχθεί με τις παραδοσιακές χειρωνακτικές μεθόδους (Sinclair 1991).

Η μελέτη του ύφους έχει κυρίως εφαρμοστεί σε λογοτεχνικά κείμενα (Clark 1996; Semino & Culpeper 2002). Επικεντρώνεται στην εξαγωγή και περιγραφή της χρήσης της γλώσσας που επιλέγεται από συγκεκριμένα άτομα ή κοινωνικές ομάδες. Η χρήση της άρχισε στα μέσα του 19^{ου} αιώνα για την πιστοποίηση συγγραφέων βιβλικών κειμένων της Παλαιάς και Καινής Διαθήκης. Μεταξύ των πρώτων εφαρμογών ήταν και η πιστοποίηση κειμένων που είχαν αποδοθεί στον Shakespeare. Η εφαρμογή τεχνικών της γλωσσολογίας σωμάτων κειμένων στη μελέτη του ύφους προσφέρει ένα συστηματικό, γρήγορο και ακριβή τρόπο για τη γλωσσολογική διερεύνηση λογοτεχνικών κειμένων, εφαρμόζοντας αυτόματη ή ημι-αυτόματη επεξεργασία σε ηλεκτρονικά κείμενα και χρησιμοποιώντας γλωσσολογική μαζί με εμπειρική γνώση.

3. Η Μεθοδολογία

Τα μέσα μαζικής ενημέρωσης λαμβάνουν μηνύματα από τρομοκρατικές οργανώσεις, τα οποία συνοδεύουν χτυπήματα τρομοκρατικών οργανώσεων ή προειδοποιούν για μελλοντικά χτυπήματα. Συχνά η αυθεντικότητα των μηνυμάτων είναι υπό αμφισβήτηση. Η *εύρεση του συγγραφέα* (authorship identification) είναι μία σημαντική ερευνητική περιοχή της *δικαστικής γλωσσολογίας* (forensic linguistics) (Shy 1998; McMenemy 2002; Gibbons 2003; Olsson 2004). Σε αυτήν τη μελέτη περιγράφουμε ένα εργαλείο για την εύρεση του συγγραφέα κειμένου μεταξύ ενός συνόλου πιθανών συγγραφέων. Αν και το εργαλείο δεν θα μπορούσε να είναι πανάκεια, μπορεί να χρησιμοποιηθεί σε συνδυασμό με άλλα στοιχεία που παρουσιάζονται σε μία δίκη για την σωστότερη έκβαση της τελικής κρίσης.

Χαρακτηρίζουμε κάθε πιθανό συγγραφέα από ένα σύνολο υφολογικών γνωρισμάτων τα οποία φαίνονται στα κείμενα του. Ας υποθέσουμε ότι έχουμε κείμενα των οποίων γνωρίζουμε τους συγγραφείς. Αντιστοιχίζουμε τα κείμενα σε σύνολα τιμών υφολογικών γνωρισμάτων τις οποίες ταξινομούμε σε πίνακες. Ας υποθέσουμε για παράδειγμα ότι έχουμε τρία κείμενα. Για τα δύο από αυτά γνωρίζουμε τους συγγραφείς. Για το τρίτο γνωρίζουμε μόνο ότι ανήκει στον έναν από τους δύο συγγραφείς αλλά δεν γνωρίζουμε σε ποιον. Υπολογίσουμε την “απόσταση”, όσο αφορά στα υφολογικά τους γνωρίσματα, για τα δύο ζευγάρια κειμένων: τα (γνωστό-κείμενο₁, άγνωστο-κείμενο) και (γνωστό-κείμενο₂, άγνωστο-κείμενο). Το ζευγάρι που δίνει την μικρότερη “υφολογική” απόσταση δίνει τον πιθανότερο συγγραφέα του άγνωστου-κειμένου. Για τον

χαρακτηρισμό των κειμένων και τον υπολογισμό της απόστασης τους, εξετάζουμε τα ακόλουθα υφολογικά γνωρίσματα:

Μέσο μήκος πρότασης	Χρήση των συνδέσμων
Μέσο μήκος λέξης	Χρήση των προθέσεων
Αριθμός των διαφορετικών ουσιαστικών	Χρήση των αντωνυμιών
Αριθμός των διαφορετικών επιθέτων	Χρήση των μορίων
Αριθμός των διαφορετικών κυρίων ονομάτων	Χρήση των σημείων στίξης
Αριθμός των διαφορετικών ενεργητικών ρημάτων	Χρήση των ερωτηματικών λέξεων
Αριθμός των διαφορετικών παθητικών ρημάτων	Χρήση των λέξεων πληροφορίας
Αριθμός των διαφορετικών επιρρημάτων	Χρήση των οριστικών άρθρων
Αριθμός των αναφορών στο μέλλον	Χρήση των αόριστων άρθρων

Οι τιμές στα παραπάνω υφολογικά γνωρίσματα εκχωρούνται στα κελιά ενός πίνακα ο οποίος ονομάζεται *πίνακας υφολογικών γνωρισμάτων*. Ο πίνακας 1 είναι ένας τέτοιος *πίνακας υφολογικών γνωρισμάτων*. Κάθε κελί του κρατά την τιμή που αντιστοιχεί σε ένα συγκεκριμένο υφολογικό γνώρισμα (από αυτά που δώσαμε παραπάνω).

Πίνακας 1: Ένας πίνακας υφολογικών γνωρισμάτων

μόρια	οριστικά άρθρα	Αόριστα άρθρα	σύνδεσμοι	αντωνυμίες
μήκος πρότασης	ερωτηματικές λέξεις	μελλοντικές αναφορές	Κύρια ονόματα	παθητικά ρήματα
λέξεις πληροφορίας	επίθετα	επιρρήματα	ουσιαστικά	ενεργητικά ρήματα
προθέσεις	χρήση του “;”	χρήση του “!”	χρήση του “:”	μήκος λέξης

Έπειτα, για κάθε ζεύγος κειμένων που θέλουμε να συγκρίνουμε, κατασκευάζουμε έναν πίνακα τον οποίο ονομάζουμε *πίνακα αποστάσεων*. Κάθε κελί του πίνακα περιέχει έναν αριθμό ο οποίος αντιστοιχεί στην υφολογική διαφορά μεταξύ των δύο κειμένων όσο αφορά σε συγκεκριμένο υφολογικό γνώρισμα. Ο πίνακας 2 δίνει τα περιεχόμενα ενός *πίνακα αποστάσεων*.

Πίνακας 2. Ένας πίνακας αποστάσεων.

απόσταση στα μόρια	απόσταση στα οριστικά άρθρα	απόσταση στα αόριστα άρθρα	απόσταση στους σύνδεσμοι	Απόσταση στις αντωνυμίες
απόσταση στο μήκος πρότασης	απόσταση στις ερωτηματικές λέξεις	απόσταση στις μελλοντικές αναφορές	απόσταση στα κύρια ονόματα	απόσταση στα παθητικά ρήματα
απόσταση στις λέξεις πληροφορίας	απόσταση στα επίθετα	απόσταση στα επιρρήματα	απόσταση στα ουσιαστικά	απόσταση στα ενεργητικά ρήματα
απόσταση στις προθέσεις	απόσταση στη χρήση του “;”	απόσταση στη χρήση του “!”	απόσταση στη χρήση του “:”	απόσταση στο μήκος λέξης

Υπολογίζουμε τον *πίνακα αποστάσεων* για κάθε ζευγάρι κειμένων τα οποία θέλουμε να συγκρίνουμε. Για να υπολογίσουμε τον *πίνακα αποστάσεων* χρειάζεται να δώσουμε μία τιμή σε κάθε κελί του, η οποία και θα αντιστοιχεί σε συγκεκριμένο υφολογικό γνώρισμα. Η τιμή αυτή παρέχει τη διαφορά (απόσταση) μεταξύ των δύο κειμένων για το συγκεκριμένο υφολογικό γνώρισμα. Έπειτα, από τον *πίνακα αποστάσεων* των δύο κειμένων, υπολογίζουμε την απόσταση μεταξύ των δύο κειμένων από τη μέση τιμή όλων των αποστάσεων των υφολογικών γνωρισμάτων.

Η *απόσταση γνωρίσματος* η οποία δίνεται σε κάθε κελί υπολογίζεται από τη μέση τιμή όλων των αποστάσεων των περιστατικών του γνωρίσματος. Οι αποστάσεις των περιστατικών του γνωρίσματος υπολογίζονται από την κανονικοποίηση των συχνοτήτων των γνωρισμάτων των κειμένων που συγκρίνονται. Ο πίνακας 3 δίνει τον αλγόριθμο για την επεξεργασία των *αποστάσεων γνωρισμάτων* τα οποία και εκχωρούνται στον *πίνακα αποστάσεων*. Στην ενότητα 4 δίνεται η εφαρμογή του αλγορίθμου σε πραγματικά κείμενα.

Πίνακας 3. Ο αλγόριθμος.

για κάθε ζεύγαρι(κείμενο₁, κείμενο₂)
 για κάθε κελί i του *πίνακα-αποστάσεων*(κείμενο₁, κείμενο₂) εκχώρησε
 απόσταση-γνωρίσματος _{i} (κείμενο₁, κείμενο₂)=
 avg(για κάθε i_j , *απόσταση-γνωρίσματος* _{i} (κείμενο₁, κείμενο₂) [i_j])
 όπου i_j είναι το j περιστατικό του γνωρίσματος i και
 για κάθε i_j περιστατικό του γνωρίσματος i
 απόσταση-γνωρίσματος _{i} (κείμενο₁, κείμενο₂) [i_j]= $|f1i_j - f2i_j| / (f1i_j + f2i_j)$
 όπου $f1i_j$ και $f2i_j$ είναι οι συχνότητες εμφάνισης του περιστατικού
 i_j του γνωρίσματος i για τα κείμενο₁ και κείμενο₂

απόσταση(κείμενο₁, κείμενο₂)=
 avg(για κάθε i , *απόσταση-γνωρίσματος* _{i} (κείμενο₁, κείμενο₂))

4. Η Εφαρμογή

Εφαρμόζουμε τη μέθοδο στο σύνολο των απολογιών και δευτερολογιών των κατηγορουμένων ως μέλη της οργάνωσης “17 Νοέμβρη”. Η “17 Νοέμβρη” ήταν ενεργή στην Ελλάδα από το 1975 έως το 2002 με πρώτη επίθεση τον φόνο του Richard Welch το 1975. Το 2002 η δράση της οργάνωσης τερματίστηκε με τη σύλληψη μελών της. Αναλυτικά στοιχεία για την οργάνωση και τις επιθέσεις της μπορούν να βρεθούν από τον Kassimeris (1997), τους Παπαχελά και Τέλλογλου (2002) και τον Πρετεντέρη (2002). Το θέμα της “17 Νοέμβρη” είναι ακόμα ανοικτό. Η οργάνωση συνήθιζε να στέλνει προκηρύξεις στα μέσα ενημέρωσης μετά από τις επιθέσεις. Μέχρι το καλοκαίρι του 2002 οπότε και έγινε σύλληψη μελών, 84 προκηρύξεις οι οποίες και είχαν σταλεί στα μέσα ενημέρωσης είχαν χαρακτηρίστηκαν ως γνήσιες (Κάκτος, 2002).

Ο τελικός μας στόχος είναι το ταίριασμα κάθε προκήρυξης στον πιθανότερο συγγραφέα της. Προσεγγίζουμε το πρόβλημα χρησιμοποιώντας τη μέθοδο του *πίνακα-αποστάσεων*. Κάθε συγγραφέας χαρακτηρίζεται από τα κείμενα που έχει παράγει, τα οποία στην περίπτωση μας είναι οι απολογίες και δευτερολογίες των κατηγορουμένων ως μελών της οργάνωσης. Η ιδέα είναι ότι, αν μπορούμε να ταιριάξουμε μία προκήρυξη με μία απολογία (και δευτερολογία) σύμφωνα με τα υφολογικά τους γνωρίσματα, τότε μπορούμε πιθανόν να την ταιριάξουμε με τον συγγραφέα της. Η μικρότερη απόσταση ταιριάζει μία προκήρυξη με μία απολογία και κατ' επέκταση με τον συγγραφέα της. Προκηρύξεις οι οποίες φαίνεται να απέχουν πολύ όσο αφορά στον υφολογικό τους

χαρακτήρα, θα μπορούσαν να υποδεικνύουν κάποιο άτομο το οποίο δεν έχει ακόμα βρεθεί.

Βρίσκουμε την *απόσταση* μεταξύ μίας προκήρυξης και μίας απολογίας βάσει των τιμών των υφολογικών τους χαρακτηριστικών. Καθώς υπάρχουν διαφορές στο ύφος γραπτού και προφορικού γλωσσικού υλικού (οι απολογίες και δευτερολογίες ανήκουν στη δεύτερη κατηγορία – θεωρητικά τουλάχιστον) υπάρχει υφολογική πληροφορία την οποία και δεν μπορούμε να χρησιμοποιήσουμε για τις συγκρίσεις μας με τις προκηρύξεις, π.χ. η ορθογραφία, κάποια σημεία στίξης, χαρακτηριστικά των παραγράφων, κτλ. (πληροφορία η οποία θα μπορούσε να αυξήσει το μέγεθος του *πίνακα-αποστάσεων*).

Εξετάζουμε το πώς η μέθοδος δουλεύει σε αυτού του τύπου τα δικαστικά κείμενα. Εδώ θα δώσουμε μία εφαρμογή του *πίνακα-αποστάσεων* σε κείμενα για τα οποία ξέρουμε ήδη τους συγγραφείς για να δείξουμε πώς δουλεύει η μέθοδος. Για την οργάνωση “17 Νοέμβρη” το γλωσσικό υλικό που θα χρησιμοποιήσουμε είναι το σύνολο των απολογιών και δευτερολογιών των κατηγορουμένων ως μέλη της. Εφαρμόζουμε τη μέθοδο για συγκρίσεις μεταξύ των απολογιών και δευτερολογιών ξέροντας ήδη σε ποιους ανήκουν. Έτσι, μπορούμε να δούμε κατά πόσον η μέθοδος μπορεί να ταιριάζει μία δευτερολογία σε απολογία του ίδιου συγγραφέα.

Ο πίνακας 4 δίνει τον αριθμό των λέξεων των 16 απολογιών και 6 δευτερολογιών της πρώτης δίκης για τα άτομα που κατηγορήθηκαν ως μέλη της οργάνωσης. Για να δώσουμε ένα παράδειγμα του πώς δουλεύει ο *πίνακας-αποστάσεων* θα τον εφαρμόσουμε σε τρία κείμενα: μία απολογία και δύο δευτερολογίες. Ήδη ξέρουμε τους συγγραφείς της απολογίας και των δευτερολογιών και ακόμα ότι η μία από τις δευτερολογίες ανήκει στο άτομο στο οποίο ανήκει και η δευτερολογία. Θα δείξουμε πώς ο *πίνακας-αποστάσεων* κάνει το ταίριασμα μεταξύ των δευτερολογιών και της απολογίας.

Έχουμε λοιπόν την απολογία και τη δευτερολογία του Γιωτόπουλου (κατηγορούμενος για αρχηγός της οργάνωσης) και τη δευτερολογία του Κουφοντίνα (επίσης από τα ηγετικά μέλη). Θα υπολογίσουμε τον *πίνακα-αποστάσεων* για τα δύο ζευγάρια κειμένων:

1. απολογία και δευτερολογία του Γιωτόπουλου
2. απολογία του Γιωτόπουλου και δευτερολογία του Κουφοντίνα.

Πίνακας 4. Το σύνολο των απολογιών και δευτερολογιών

Απολογίες				Δευτερολογίες	
Άτομο	λέξεις	άτομο	λέξεις	άτομο	λέξεις
Γεωργιάδης	12479	Π_Σερίφης	4095	Κώσταρης	1594
Καρατσώλης	6483	Θ_Σερίφης	7078	Κουφοντίνας	326
Κονδύλης	13688	Σωτηροπούλου	14723	Ψαραδέλλης	502
Κώσταρης	14257	Τσελέντης	42778	Γ_Σερίφης	703
Κουφοντίνας	5699	Τζορτζάτος	13906	Τζορτζάτος	8111
Παπαναστασίου	6974	Β_Ξηρός	454	Γιωτόπουλος	2036
Ψαραδέλλης	20957	Χ_Ξηρός	12205		
Γ_Σερίφης	1114	Γιωτόπουλος	18321		

Ας υπολογίσουμε ένα από τα κελιά των δύο *πινάκων-αποστάσεων*. Ο υπολογισμός των υπόλοιπων κελιών γίνεται με τον ίδιο τρόπο¹. Ας θεωρήσουμε το κελί [1,1] του

¹ Δεν θα παρουσιάσουμε εδώ τους υπολογισμούς όλων των κελιών των πινάκων για λόγους χώρου.

πίνακα-αποστάσεων, αυτό που περιέχει πληροφορίες όσο αφορά το υφολογικό χαρακτηριστικό της χρήσης των μορίων. Πέντε μόρια βρέθηκαν στα τρία κείμενα που εξετάζουμε: “να”, “για”, “θα”, “ας” και “μα”. Ο πίνακας 5 δίνει τις συχνότητες εμφάνισης και τις κανονικοποιημένες συχνότητες (σε 10.000 λέξεις) για τα πέντε μόρια που απαντώνται στην απολογία και δευτερολογία του Γιωτόπουλου και τη δευτερολογία του Κουφοντίνα.

Πίνακας 5: Συχνότητες εμφάνισης (*fr*) και κανονικοποιημένες συχνότητες (*n-fr*) των μορίων για τα τρία κείμενα

Γ_{α} – Γιωτόπουλος απολογία			Γ_{δ} – Γιωτόπουλος δευτερολογία			K_{α} – Κουφοντίνας δευτερολογία		
	<i>fr</i>	<i>n-fr</i>		<i>f</i>	<i>n-fr</i>		<i>fr</i>	<i>n-fr</i>
να	450	245,619	να	60	294,695	για	6	184,049
για	183	99,885	θα	22	108,055	να	4	122,699
θα	165	90,060	για	20	98,232	θα	3	92,024
ας	6	3,274	ας	0	0	ας	0	0
μα	0	0	μα	0	0	μα	0	0

Ας υπολογίσουμε το κελί [1,1] του πίνακα-αποστάσεων για το ζεύγος κειμένων της απολογίας και δευτερολογίας του Γιωτόπουλου. Το κελί παρέχει την απόσταση των δύο κειμένων όσο αφορά στη χρήση των μορίων. Σύμφωνα με τον αλγόριθμο που δίνεται στον πίνακα 3 υπολογίζουμε την απόσταση μεταξύ των δύο κειμένων για καθένα από τα πέντε μόρια και εκχωρούμε στο κελί [1,1] τη μέση τιμή των αποστάσεων αυτών. Μετά επαναλαμβάνουμε τη διαδικασία για το κελί [1,1] του πίνακα-αποστάσεων του ζεύγους κειμένων της απολογίας του Γιωτόπουλου και της δευτερολογίας του Κουφοντίνα. Οι αποστάσεις σχετικά με το morio “να” για το ζεύγος κειμένων της απολογίας του Γιωτόπουλου (Γ_{α}) και της δευτερολογίας του Γιωτόπουλου (Γ_{δ}) και για το ζεύγος της απολογίας του Γιωτόπουλου (Γ_{α}) και της δευτερολογίας του Κουφοντίνα (K_{δ}) υπολογίζονται:

$$\text{απόσταση-μορίου}(Y_{\alpha}, Y_{\delta})[\text{να}] = |f_{Yp}[\text{να}] - f_{Yr}[\text{να}]| = 0,091$$

$$\text{απόσταση-μορίου}(Y_{\alpha}, K_{\delta})[\text{να}] = |f_{Yp}[\text{να}] - f_{Kr}[\text{να}]| = 0,334$$

Πίνακας 6. Πίνακας-Μορίων

Μόριο	απόσταση-μορίων(Γ_{α} , Γ_{δ})	απόσταση-μορίων(Γ_{α} , K_{δ})
να	0,091	0,334
για	0,0088	0,251
θα	0,0587	0,051
ας	1	1
μα	0	0
avg	0,289	0,409

Ο πίνακας 6 δίνει τα αποτελέσματα των παραπάνω υπολογισμών για όλες τις αποστάσεις μορίων των ζευγαριών κειμένων της απολογίας και δευτερολογίας του Γιωτόπουλου και της απολογίας του Γιωτόπουλου και της δευτερολογίας του Κουφοντίνα. Παίρνουμε τις μέσες τιμές για τις δύο στήλες του πίνακα 6, δηλ. τα 0,289 και 0,409. Αυτοί οι δύο αριθμοί, οι μέσες τιμές όλων των αποστάσεων των μορίων για τα δύο ζεύγη κειμένου, εκχωρούνται στα αντίστοιχα κελιά των δύο πινάκων-αποστάσεων (θέση [1,1] όπως έχει δοθεί στον πίνακα 2).

Πίνακας 7. Ο πίνακας-αποστάσεων για το ζεύγος κειμένων (Γιωτόπουλος-απολογία, Γιωτόπουλος-δευτερολογία)

<i>Ο πίνακας-αποστάσεων για το (Γιωτόπουλος-απολογία, Γιωτόπουλος-δευτερολογία)</i>				
0,289	0,225	0,191	0,552	0,63
0,32	0,225	0,09	0,305	0,34
0,232	0,04	0,17	0,017	0,009
0,616	0,477	1	1	0,036

Πίνακας 8. Ο πίνακας-αποστάσεων για το ζεύγος κειμένων (Γιωτόπουλος-απολογία, Κουφοντίνας-δευτερολογία)

<i>Ο πίνακας-αποστάσεων για το (Γιωτόπουλος-απολογία, Κουφοντίνας-δευτερολογία)</i>				
0,409	0,923	0,667	0,922	0,849
0,622	0,776	0,010	0,575	0,474
0,176	0,27	0,115	0,485	0,073
0,798	1	1	1	0,042

Επαναλαμβάνουμε την ίδια διαδικασία για τον υπολογισμό των αποστάσεων των γνωρισμάτων για τα υπόλοιπα κελιά και για τους δύο πίνακες-αποστάσεων. Ο πίνακας 7 δίνει τον πίνακα-αποστάσεων για τα κείμενα της απολογίας και της δευτερολογίας του Γιωτόπουλου, ενώ ο πίνακας 8 δίνει τον πίνακα-αποστάσεων για τα κείμενα της απολογίας του Γιωτόπουλου και της δευτερολογίας του Κουφοντίνα. Όταν οι πίνακες αυτοί έχουν συμπληρωθεί, μπορούμε να υπολογίσουμε τις συνολικές αποστάσεις για τα δύο ζευγάρια κειμένων, οι οποίες θα είναι οι μέσες τιμές όλων των τιμών των κελιών των δύο πινάκων:

απόσταση-πίνακα(Γ_a , K_δ)= 0,5593 και

απόσταση-πίνακα(Γ_a , Γ_δ)= 0,3382

Σαν αποτέλεσμα, η απολογία του Γιωτόπουλου συνδέεται με τη δευτερολογία του Γιωτόπουλου (λόγω της μικρότερης διαφοράς), το οποίο και σημαίνει ότι τα δύο αυτά κείμενα φαίνεται να ταιριάζουν περισσότερο όσο αφορά στο ύφος της γλώσσας, σε σχέση με το ταίριασμα που έχουν τα κείμενα της απολογίας του Γιωτόπουλου και της δευτερολογίας του Κουφοντίνα.

5. Μελλοντική Δουλειά

Σε αυτήν τη μελέτη περιγράψαμε μία μεθοδολογία η οποία με τη χρήση τεχνικών της γλωσσολογίας σωμάτων κειμένων και υφολογικής πληροφορίας προσεγγίζει το θέμα της εύρεσης του συγγραφέα. Εφαρμόσαμε το εργαλείο που προτείνουμε στις απολογίες και δευτερολογίες των κατηγορούμενων ως μέλη της τρομοκρατικής οργάνωσης “17 Νοέμβρη”. Το εργαλείο σε καμία περίπτωση δεν αποτελεί πανάκεια, μπορεί όμως να χρησιμοποιηθεί για την παροχή πληροφορίας, η οποία αφού αξιολογηθεί μπορεί να προστεθεί στα υπόλοιπα στοιχεία με σκοπό την εύρεση του πραγματικού συγγραφέα αμφισβητούμενου κειμένου.

Στο μέλλον, στα υφολογικά γνωρίσματα θα προστεθεί η χρήση των τυποποιημένων εκφράσεων και ιδιωτισμών αλλά και όρων (Frantzi & Ananiadou 1999). Επίσης θα δοκιμάσουμε την χρήση βαρών στα υφολογικά χαρακτηριστικά.

Η μέθοδος θα χρησιμοποιηθεί για τη σύγκριση των απολογιών και δευτερολογιών και των προκηρύξεων, για την εξαγωγή στοιχείων που θα μπορούσαν να διευκολύνουν την εύρεση των πραγματικών συγγραφέων των προκηρύξεων. Για τον ίδιο σκοπό, θα

γίνουν συγκρίσεις μεταξύ των προκηρύξεων για την ομαδοποίηση τους σύμφωνα με το ύφος τους.

Στα πιο μακρινά μας σχέδια είναι η εφαρμογή της πρότασης σε άλλες γλώσσες για τη σύγκριση του ύφους της γλώσσας διαφορετικών τρομοκρατικών οργανώσεων.

Βιβλιογραφία

- Biber, D., S. Conrad & R. Reppen (1998). *Corpus linguistics - investigating language structure and use*. Cambridge, U.K.: Cambridge University Press.
- Clark, U. (1996). *An introduction to stylistic - investigating English language*. United Kingdom: Stanley Thornes LTD.
- Frantzi, K. T. & S. Ananiadou (1999). "The C-value / NC-value domain independent method for multi-word term extraction". *Journal of Natural Language Processing* 6(3): 145–179.
- Frantzi, K. T. (2006). "Language resources: What can they do with terrorism?". *The International Journal of Humanities*, 2(2): 1603-1608.
- Frantzi, K. T. & S. Ananiadou (2006). "Automatic authorship identification". *Book of Abstracts of the International Association of Forensic Linguistics 2nd European Conference on Forensic Linguistics / Language & the Law*, 14-16 September, Barcelona, Spain: 8.
- Gibbons, J. (2003). *Forensic linguistics – An introduction to language in the justice system*. Blackwell Publishing Ltd.
- Kassimeris, G. (1997). *Europe's last red terrorists - the revolutionary organisation 17 November*. C. Hurst & Co Ltd.
- Κάκτος (εκδότης) (2002). *17N – Οι προκηρύξεις – Όλα τα κείμενα της οργάνωσης*. Αθήνα: Κάκτος.
- McEnery, T. & A. Wilson (1996). *Corpus linguistics*. Edinburgh, U.K.: Edinburgh University Press.
- McMenamin, G. R. (2002). *Forensic linguistics: Advances in Forensic Stylistics*. CRC Press.
- Olsson, J. (2004). *Forensic linguistics – an introduction to language, crime and the law*. Continuum.
- Παπαχελάς, Α. & Τ. Τέλλογλου (2002). *Φάκελος 17 Νοέμβρη*. Αθήνα: Εστία.
- Πρετεντέρης, Γ. 2002. *Η αναμέτρηση – ζωή και θάνατος της 17N*. Βιβλιοπωλείον της Εστίας, Αθήνα.
- Semino, E. & J. Culpeper (eds.) (2002). *Cognitive stylistics: language and cognition in text analysis*. John Benjamin Publishing Company.
- Shy, R. W. (1998). *The language of confession, interrogation and deception. Empirical Linguistics Series*. U.S.A.: Sage Publications.
- Sinclair, J. (1991). "Corpus, concordance, collocation". In *Sinclair, J.; Carter, R. (eds.)*. Oxford: Oxford University Press.
- Wynne, M. & P. Stockwell (2006). "Corpus stylistics: a public inquiry". Keynote talk at *The Third Inter-Varietal Applied Corpus Studies (IVACS) group International Conference on Language at the Interface*, University of Nottingham, U.K., 23rd – 24th June.