

Ανάλυση Αυθόρμητου Λόγου Χρησιμοποιώντας Τεχνικές Επεξεργασίας της Γλώσσας για Εντοπισμό Αναφορών σε Συμπτώματα του Εγκεφαλικού Επεισοδίου

Ε. Σ. ΡΗΓΑΣ¹, Μ. ΠΑΠΟΥΤΣΟΓΛΟΥ¹, Κ. ΤΣΑΚΠΟΥΝΙΔΟΥ², Α. ΣΕΡΝΤΑΡΗ³, Ε. ΜΟΥΖΑ⁴, Χ. ΠΡΩΙΟΥ^{2,4}

Περίληψη

Η παρούσα έρευνα μελετά τη συχνότητα που εντοπίζονται τα συμπτώματα Πρόσωπο, Χέρια, Ομιλία, που προέκυψαν από τη διαδικασία λήψης ιατρικού ιστορικού ατόμων με εγκεφαλικό επεισόδιο, χρησιμοποιώντας τεχνικές επεξεργασίας φυσικής γλώσσας. Συνολικά συμμετείχαν 58 άτομα, τα οποία προσκομίστηκαν για εξέταση και διάγνωση προβλημάτων επικοινωνίας. Όλοι οι συμμετέχοντες απάντησαν προφορικά και οι ηχογραφήσεις αποτυπώθηκαν σε γραπτό λόγο για μεταγενέστερη ανάλυση του κειμένου. Χρησιμοποιήσαμε δύο τεχνικές επεξεργασίας φυσικής γλώσσας, δηλαδή τη συχνότητα εμφάνισης όρων (tf) που δείχνει τον αριθμό που εμφανίζεται μια λέξη σε ένα κείμενο και τη λανθάνουσα κατανομή Dirichlet (Latent Dirichlet Allocation LDA) η οποία διέκρινε 3 θέματα στην ανάλυση κειμένου ως τον ιδανικό αριθμό θεμάτων. Όσον αφορά το tf, κατατάξαμε τις πιο συχνά εμφανιζόμενες ακολουθίες σε φράσεις μίας, δύο ή τριών λέξεων. Για τις μεμονωμένες λέξεις, παρατηρήσαμε ότι η «ομιλία» και το «χέρι» εμφανίστηκαν συχνότερα κατά τη λήψη ιατρικού ιστορικού (σχν. 16.1% και 6.9% αντίστοιχα), ενώ το «πρόσωπο» και άλλες ανάλογες λέξεις ήταν λιγότερο συχνές (σχν. 2.68%). Όσον αφορά την LDA, 3 θέματα, μεταξύ των οποίων το «χέρι» και ο «λόγος», ξεχώρισαν από τα υπόλοιπα. Έτσι, οι λέξεις «χέρι» και «ξαφνικά» στο

θέμα 1, «εγκεφαλικό», «αναφέρει» και «αδυναμία» στο θέμα 2, «νοσοκομείο» και «αδυναμία» στο θέμα 3, ξεχώρισαν. Τα αποτελέσματα δείχνουν ότι τα συμπτώματα του εγκεφαλικού επεισοδίου θα πρέπει να συμπεριλαμβάνονται στη λήψη αποφάσεων κατά την κλινική υγειονομική περίθαλψη.

Λέξεις-κλειδιά: Εγκεφαλικό, Συχνότητα όρων, Λανθάνουσα κατανομή Dirichlet, Συμπτώματα, Ιατρικό ιστορικό

¹Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τμήμα Πληροφορικής, Θεσσαλονίκη, Ελλάδα.

²Πανεπιστήμιο Μακεδονίας, Τμήμα Εκπαιδευτικής και Κοινωνικής Πολιτικής, Θεσσαλονίκη, Ελλάδα.

³Δημοκρίτειο Πανεπιστήμιο Θράκης, Τμήμα Ιατρικής, Πανεπιστημιακή Παιδοψυχιατρική Κλινική, Αλεξανδρούπολη, Ελλάδα.

⁴Κέντρο Αποθεραπείας-Αποκατάστασης «η Αναγέννηση», Τμήμα Λογοθεραπείας, Νέα Ραιδεστός, Φιλοθέη, Θεσσαλονίκη, Ελλάδα.

Υπεύθυνη επικοινωνίας: Χαρίκλεια Πρώιου

Τμήμα Εκπαιδευτικής και Κοινωνικής Πολιτικής, Πανεπιστήμιο Μακεδονίας, Εγνατία 156, Θεσσαλονίκη, ΤΚ: 54636, Ελλάδα, hproios@uom.edu.gr, +30 2310 891630

1. Εισαγωγή

Σύμφωνα με την Έρευνα του 2010 “Globan burden of disease” υπάρχουν περισσότεροι από 33 εκατομμύρια επιζώντες του εγκεφαλικού επεισοδίου (Feigin et al., 2014), εκ των οποίων περισσότεροι από ένα εκατομμύριο επιζήσαντες έχουν αφασία, με τον αριθμό αυτό να αντανακλά μόνο τις Ηνωμένες Πολιτείες της Αμερικής (Moffatt, Pourshahid, & Baecker, 2017). Παλαιότερες έρευνες έχουν επιχειρήσει να εντοπίσουν γενετικούς βιοδείκτες του εγκεφαλικού επεισοδίου (Bang, 2017; Damico, Müller, & Ball, 2010; Katan, & Elkind, 2018; Kim, Moon, & Bang, 2013). Αυτή η γνώση είναι εξαιρετικά σημαντική για την εξασφάλιση της υγιούς διαβίωσης σε έναν κόσμο όπου ο επιπολασμός των ατόμων που σχετίζεται με κάποια αναπηρία λόγω εγκεφαλικού επεισοδίου ολοένα και αυξάνεται (Feigin, Norrving, & Mensah, 2017; Huang et al., 2016). Υπό αυτό το πρίσμα, η λήψη ιατρικού ιστορικού είναι ένα εξαιρετικά πολύτιμο μέρος της εξέτασης, διάγνωσης και θεραπείας ενός ατόμου. Κατά τη διάρκεια λήψης του ιατρικού ιστορικού ενός ασθενούς, μια σύνθητη ερώτηση προς τον ασθενή είναι «Γιατί είστε εδώ;». Με την παρούσα μελέτη στοχεύουμε στην ανάλυση του ιατρικού ιστορικού προκειμένου να εντοπίσουμε τα συμπτώματα του προσώπου, του χεριού και του λόγου που πιθανώς να περιγράφονται από τους ασθενείς που επεβίωσαν του εγκεφαλικού επεισοδίου.

Το ακρωνύμιο ΧΟΠΑ (FAST), που αντιπροσωπεύει το χέρι, την ομιλία, το πρόσωπο και το ασθενοφόρο, χρησιμοποιείται για τον εντοπισμό συμπτωμάτων που υποδεικνύουν ότι κάποιος/α παθαίνει εγκεφαλικό επεισόδιο (Dombrowski et al., 2015) και για την άμεση κλήση ενός ασθενοφόρου. Τα πιο κοινά συμπτώματα ενός εγκεφαλικού επεισοδίου είναι το Χ: ημιπληγία/ημιπάρεση του ενός χεριού, Ο: ελλείμματα λόγου και/ή αφασία ή Π: πτώση της μίας πλευράς του προσώπου.

Για τις ανάγκες της παρούσας μελέτης δημιουργούμε ένα μοντέλο που συνδέει τα συμπτώματα του χεριού, της ομιλίας και του προσώπου με την παραγωγή λόγου από κάθε άτομο. Συχνά, όταν λαμβάνεται το ιατρικό ιστορικό ενός επιζήσαντα εγκεφαλικού επεισοδίου, η επικοινωνία είναι διαταραγμένη. Η παραγωγή λόγου μπορεί συχνά να είναι ασυνάρτητη και χωρίς συνοχή, ή ακόμη και ακατανόητη (Kurowski, & Blumstein, 2016). Σε αυτή τη μελέτη χρησιμοποιήσαμε καταγεγραμμένες απαντήσεις στην ερώτηση «Γιατί είστε εδώ;», από την ελληνική έκδοση του Aachen Aphasia Test (AAT) (Proios et al., 2006), για

να κατανοήσουμε και να προσδιορίσουμε τις σχετικές περιγραφές με τα τρία συμπτώματα που αναφέρονται από τους επιζώντες του εγκεφαλικού επεισοδίου. Αρχικά εξετάσαμε τους αυθόρμητους διαλόγους και εντοπίσαμε τα συμπτώματα που ανέφεραν στις προσωπικές τους περιγραφές οι συμμετέχοντες. Έτσι, αναλύσαμε τη σχέση μεταξύ των συμπτωμάτων και των θεμάτων που προκύπτουν από τη διαδικασία μοντελοποίησης.

Για να αποφευχθούν διαγνωστικά σφάλματα των εξοικειωμένων περιγραφών, το εύρος της λήψης ιατρικού ιστορικού επεξεργάστηκε χρησιμοποιώντας υπολογιστικές τεχνικές. Ο κύριος σκοπός της μελέτης μας ήταν η εφαρμογή μιας τυπικής τεχνικής επεξεργασίας φυσικής γλώσσας, προσδιορίζοντας τα συμπτώματα των καταγεγραμμένων απαντήσεων.

Παρόλο που υπάρχουν μελέτες για υπολογιστικές τεχνικές με θεραπευτικούς σκοπούς (Grawemeyer, Cox, & Lum, 2000; Vannobel, & Toulotte, 1992), καθώς και για διάγνωση της αφασίας (Sponsler, & Burkhart, 2016), ες' όσων γνωρίζουμε, το εύρος βιβλιογραφίας για αναφορά των συμπτωμάτων ενός εγκεφαλικού επεισοδίου κατά τη λήψη ιατρικού ιστορικού είναι περιορισμένο. Υπάρχουν βέβαια κάποιες έρευνες για τη χρήση υπολογιστικών τεχνικών με στόχο την εύρεση θεμάτων που ενδιαφέρονται οι επιζώντες εγκεφαλικού επεισοδίου να συζητήσουν (όπως «φαγητό και ροφήματα», «φύση και κηπουρική», «ψυχαγωγία» και άλλα) (Palmer, Hughes, & Chater, 2017) αλλά και την αξιολόγηση των αναγκών των επιζήσαντων για την πληροφόρησή τους σχετικά με το εγκεφαλικό επεισόδιο και τα απότοκά αυτού (Dixon, Thornton, & Young, 2007; Mangset et al., 2008; Worrall et al., 2011).

2. Μεθοδολογία

Συνολικά 58 άτομα συμμετείχαν στην έρευνα, 68.96% άνδρες και 31.04% γυναίκες, ηλικίας από 26 έως 78 ετών με μέσο όρο 58.29 και διάμεση τιμή 61, η πλειοψηφία των οποίων υπέστη ισχαιμικό αγγειακό εγκεφαλικό επεισόδιο. Σε όλους τους συμμετέχοντες τέθηκε η ερώτηση «Γιατί είστε εδώ σήμερα;». Με βάση αυτό, περιέγραψαν προφορικά τα συμπτώματά τους, τα οποία μεταμορφώθηκαν σε γραπτό κείμενο από τους συγγραφείς. Η ιστορία κάθε ασθενούς καταγράφηκε με το χέρι, προκειμένου να κρατηθούν οι σημαντικές πληροφορίες και να μειωθούν περισσια στοιχεία, με στόχο τη βελτίωση της αποτελεσμα-

τικότητα της ανάλυσης. Οι ιστορίες γράφτηκαν στην ελληνική γλώσσα. Πριν από την ανάλυση, πραγματοποιήθηκε ένα στάδιο προ-επεξεργασίας. Σε αυτό το στάδιο, ορίσαμε ένα λεξικό με λέξεις που δεν συνεισφέρουν στον ορισμό του νοήματος (stop words) (π.χ., και, του / της), προκειμένου να διατηρηθεί ένα σύνολο λέξεων που παρέχουν εννοιολογικό νόημα. Για να γίνει αυτό, 1) προστέθηκε η λίστα με τις προθέσεις, 2) προστέθηκε η λίστα με τους συνδέσμους και 3) προστέθηκαν και άλλες λέξεις που διαπιστώθηκαν στο πείραμά μας να μην προσδίδουν κάποιο επιπλέον νόημα. Σημειώστε ότι τέτοια ψηφιακά λεξικά των stop words είναι σύνθητη στην αγγλική γλώσσα, αλλά κανένα αντίστοιχο δεν υπήρχε στην ελληνική γλώσσα.

Η φάση προ-επεξεργασίας αποτελείται από τα ακόλουθα βήματα:

1. Αφαιρούμε τα ονόματα των ασθενών από το κείμενο. Αυτό το βήμα πραγματοποιήθηκε και για τους λόγους ανωνυμίας.
2. Μετατρέπουμε όλο το κείμενο σε πεζά γράμματα.
3. Αφαιρούμε όλους τους ειδικούς χαρακτήρες όπως “.” ή “;”. Για να το πετύχουμε, χρησιμοποιήσαμε την κανονική έκφραση “[[:punct:]][\$]”, όπου το punct καθορίζεται από τη γλώσσα προγραμματισμού R η οποία χρησιμοποιήθηκε σε αυτή την εργασία. Στην επιστήμη των υπολογιστών, μια κανονική έκφραση ορίζεται ως μια ακολουθία χαρακτήρων που ορίζουν ένα πρότυπο αναζήτησης. Επιλέξαμε να αφαιρέσουμε τα σύμβολα στίξης, όπως το κόμμα, τις τελείες, τα ερωτηματικά και άλλα σύμβολα, καθώς δεν παρέχουν καμία σημαντική πληροφορία για τις ανάγκες μας. Επιπλέον, αφαιρώντας τη στίξη, αφαιρούμε και τους ειδικούς χαρακτήρες που θα μπορούσαν να προκαλέσουν σφάλματα στη διαμόρφωση κειμένου.
4. Αφαιρούμε όλα τα stop words.

Μετά τη φάση προ-επεξεργασίας, πραγματοποιήθηκε η ανάλυση κειμένου. Για να αναλύσουμε το κείμενο, χρησιμοποιήσαμε δύο τυπικές τεχνικές επεξεργασίας φυσικής γλώσσας, δηλαδή τη συχνότητα όρων (tf) (Salton, & Buckley, 1998) και τη λανθάνουσα κατανομή Dirichlet (LDA) (Blei, Ng, & Jordan, 2003).

Το tf, το οποίο είναι μέρος του TFIDF - συντομογραφία του όρου συχνότητας εμφάνισης όρου (term frequency – tf) – αντίστροφης συχνότητας του εγγράφου (inverse document frequency – idf) - είναι μια αριθμητική στατιστική που έχει σκοπό να αντικατο-

πτρίζει τη σπουδαιότητα μιας λέξης σε ένα έγγραφο ενός συνόλου δεδομένων (de Zubizaray et al., 2006). Για τη συχνότητα όρων $tf(t,d)$, η απλούστερη επιλογή είναι να χρησιμοποιήσουμε τον ακριβή αριθμό ενός όρου σε ένα έγγραφο, δηλαδή τον αριθμό των φορών που εμφανίζεται ο όρος t σε ένα έγγραφο d .

Στην επεξεργασία φυσικής γλώσσας, η λανθάνουσα κατανομή Dirichlet (LDA) είναι ένα στατιστικό μοντέλο παραγωγής που επιτρέπει σε παρατηρήσιμες ομάδες στο κείμενο να επεξηγούνται από μη παρατηρήσιμες ομάδες, οι οποίες με τη σειρά τους εξηγούν γιατί ορισμένα τμήματα αυτών των δύο δεδομένων είναι παρόμοια. Στην LDA, κάθε έγγραφο μπορεί να θεωρηθεί ως ένα μείγμα διαφόρων θεμάτων, όπου κάθε έγγραφο θεωρείται ότι έχει ένα σύνολο θεμάτων που του έχουν ανατεθεί μέσω της LDA. Η εικόνα 1 δείχνει τη διαδικασία που ακολουθείται για τον υπολογισμό των θεμάτων και την εκχώρηση των όρων στα θέματα. α είναι η παράμετρος του Dirichlet πριν από τις κατανομές θεμάτων ανά έγγραφο, β είναι η παράμετρος του Dirichlet πριν από την κατανομή των λέξεων ανά θέμα, θ_d είναι η διανομή θεμάτων για τα έγγραφα, ϕ_k είναι η κατανομή λέξεων για το θέμα k , z_{dj} είναι το θέμα της λέξης j στο έγγραφο d , και το w_{dj} είναι η συγκεκριμένη λέξη. Η εικόνα 2 δείχνει ολόκληρη τη διαδικασία ανάλυσης κειμένου.

Προκειμένου να εφαρμόσουμε τα βήματα μεθοδολογίας που περιγράφηκαν προηγουμένως, επιλέξαμε τη χρήση λογισμικού ανοιχτού κώδικα. Τα βήματα της μεθοδολογίας περιέχουν ποιοτική ανάλυση του ακατέργαστου ψηφιακού περιεχομένου και χρειάζονται ειδικό χειρισμό σε αντίθεση με τις απλούστερες προσεγγίσεις που απαιτούνται για τα ποσοτικά δεδομένα. Υπάρχουν πολλές διαφορετικές λύσεις για την ποιοτική ανάλυση των δεδομένων κειμένου και μερικές από αυτές μπορούν να εξοικονομήσουν χρόνο στην πρακτική διαδικασία βοηθώντας ερευνητές που δεν έχουν εμπειρία κωδικοποίησης. Αυτή η λύση προέρχεται από εργαλεία κατάλληλα για ποιοτική ανάλυση και ένα από τα πιο συχνά χρησιμοποιούμενα στο NVivo (<https://www.qsrinternational.com/nvivo/nvivo-products>) (Abbe et al., 2016; Roundtree, 2018; Ward, Gbadebo, & Baruah, 2015). Παρόλο που το NVivo είναι ένα φιλικό προς το χρήστη λογισμικό για ανάλυση περιεχομένου, με μια εμφάνιση κατάλληλη για όλα τα επίπεδα χρηστών, η χρήση του εξακολουθεί να έχει πολλά μειονεκτήματα. Είναι εμπορικό λογισμικό και απαιτεί χρόνο για να μάθει κανείς πώς να το χρησιμοποιεί, καθώς είναι ένα ημιαυτόματο

εργαλείο (Ranney et al., 2014). Επιχειρώντας να αποφύγουμε αυτά τα μειονεκτήματα, επιλέξαμε να εφαρμόσουμε τα προτεινόμενα βήματα χρησιμοποιώντας τη γλώσσα προγραμματισμού R. Η προσέγγισή μας είναι ημι-αυτοματοποιημένη μόνο στο πρώτο βήμα της μετατροπής του ακατέργαστου χειρόγραφου περιεχομένου σε ψηφιακό. Μετά από αυτό, τα βήματα ανάλυσης περιεχομένου ήταν πλήρως αυτοματοποιημένα υπό μια λύση ανοιχτού κώδικα.

3. Αποτελέσματα

Προκειμένου να εξάγουμε αποτελέσματα με νόημα από το κείμενο, επιλέξαμε να εκτελέσουμε το τέταρτο βήμα από την Εικόνα 2 (δηλαδή την Ανάλυση κειμένου) για τρεις διαφορετικές περιπτώσεις ngram. Από το ακατέργαστο κείμενο μας συλλέξαμε μια συνεχή ακολουθία λέξεων χ στοιχείων. Για την περίπτωση μας, τα στοιχεία είναι η ακολουθία λέξεων και επιλέξαμε να εκτελέσουμε το βήμα ανάλυσης κειμένου για τρεις διαφορετικές περιπτώσεις. Ειδικότερα για το unigram ($\chi = 1$), το bigram ($\chi = 2$) και το trigram ($\chi = 3$). Επιλέξαμε την επιλογή χ -γράμμα, καθώς σε ορισμένες περιπτώσεις βρήκαμε σημαντικά αποτελέσματα για $\chi > 1$.

Αναφορικά με το tf, παρατηρούμε ότι συχνά εμφανίζονται η «ομιλία» και το «χέρι», ενώ υπάρχουν και λέξεις που σχετίζονται με το πρόσωπο ή ακόμα και με άλλες συνθήκες (βλέπε Πίνακες 1-3). Πιο λεπτομερώς, για τη συχνότητα των μονών λέξεων, οι λέξεις *ομιλία*, *αδυναμία* και *ένιωσε* εμφανίζονται πιο συχνά. Για τη συχνότητα ζευγών λέξεων, τα ζεύγη *αδυνατούσε μιλήσει*, *αδυναμία χέρι* και *δεν μπορεί* εμφανίζονται πιο συχνά. Τέλος, για τη συχνότητα των τριάδων λέξεων, οι φράσεις *δεν μπορεί μιλήσει*, *χάνει αίσθηση χέρι* και *άρχισε ατονεί τελικά* εμφανίζονται πιο συχνά. Γενικά, κάποιος μπορεί να παρατηρήσει ότι οι συχνότερες λέξεις αναφέρονται σε συμπτώματα που σχετίζονται με το λόγο και το χέρι, ενώ το πρόσωπο εμφανίζεται, αλλά σε μικρότερη συχνότητα.

Όσον αφορά την LDA, πρέπει να αναφέρουμε ότι υπάρχει συγκεκριμένος αριθμός παραγόμενων θεμάτων που είναι τα πλέον κατάλληλα για κάθε κατάσταση, σύμφωνα με την ανάλυση. Μετά από μία πειραματική διαδικασία, διαπιστώσαμε ότι ο αριθμός των θεμάτων που ξεχωρίζει και εξηγεί καλύτερα τις ανάγκες μας είναι η επιλογή 3 θεμάτων. Έτσι επιλέξαμε αυτά τα 3 θέματα τα οποία διαχωρίζονται σαφώς μεταξύ τους και παρατηρούμε ότι οι όροι «ομιλία» και

«χέρι» κατηγοριοποιούνται σε δύο από αυτά, ενώ ο τρίτος είναι πιο ασαφής, δεδομένου ότι περιέχει λέξεις που σχετίζονται με τη γενική κατάσταση των ασθενών (βλέπε Εικόνες 3-5). Με αυτό τον τρόπο, επιβεβαιώνονται τα αποτελέσματα του tf καθώς αυτά τα δύο συμπτώματα επιλέγονται και με αυτή τη μέθοδο.

4. Συζήτηση

Στην παρούσα μελέτη προσπαθούμε να διερευνήσουμε αν υπάρχει σχέση μεταξύ των συμπτωμάτων του προσώπου, του χεριού και της ομιλίας και των καταγεγραμμένων αυθόρμητων προφορικών περιγραφών των επιζώντων ενός εγκεφαλικού επεισοδίου. Οι τεχνικές επεξεργασίας tf και LDA εφαρμόστηκαν για να εκτιμηθεί η αυθόρμητη λεκτική παραγωγή των συμμετεχόντων. Αποδείχθηκε ότι οι λέξεις «ομιλία» και «χέρι» είναι οι πιο συχνά αναφερόμενες. Για τη συχνότητα των ζευγών λέξεων, οι φράσεις «αδυνατούσε μιλήσει», «αδυναμία χέρι» και «δεν μπορεί» εμφανίζονται πιο συχνά, ενώ για τη συχνότητα των τριάδων λέξεων, οι φράσεις «δεν μπορεί μιλήσει», «χάνει αίσθηση χέρι» και «άρχισε ατονεί τελικά» εμφανίζονται πιο συχνά. Τα ευρήματά μας επιβεβαιώνονται και συμφωνούν με άλλες μελέτες που δείχνουν ότι τα συμπτώματα του προσώπου, του χεριού και της ομιλίας εντοπίζονται από τη συντριπτική πλειοψηφία των επιζώντων ενός εγκεφαλικού επεισοδίου (Kleindorfer et al., 2007). Αναφέρονται και άλλα συμπτώματα του εγκεφαλικού επεισοδίου, όπως οπτικές δυσκολίες, κεφαλαλγία, ζάλη / ανισορροπία, καθώς και μερικές λιγότερο κοινές περιγραφές (Aroor, Singh, & Goldstein, 2017).

Όσον αφορά την LDA, επιλέχθηκαν τρία θέματα συζήτησης και μετρήθηκε ο αριθμός που αναφέρθηκε μία λέξη σε ολόκληρο το κείμενο (Hassabis, & Maguire, 2007). Τα θέματα «χέρι» και «ομιλία» είναι αυτά στα οποία αναφέρθηκαν περισσότερο οι επιζήσαντες του εγκεφαλικού επεισοδίου, ενώ τα υπόλοιπα είναι ασαφή καθώς περιέχουν λέξεις σχετικές με τη γενική κατάσταση των ασθενών (για παράδειγμα «νοσοκομείο», «εγκεφαλικό επεισόδιο» και άλλα). Με τον τρόπο αυτό, επιβεβαιώθηκαν τα αποτελέσματα του tf, καθώς τα συμπτώματα που επιλέχθηκαν με τη μέθοδο LDA είναι τα ίδια.

Με βάση τα αποτελέσματά μας, τα συμπτώματα της ομιλίας και του χεριού ήταν κοινά στις καταγραφές των περιγραφών των επιζώντων εγκεφαλικού επει-

σοδίου. Το ερώτημα για το εάν έχει κάποια σχέση με τη θεραπεία των ασθενών δεν μπορεί να απαντηθεί αποκλειστικά από αυτή τη μελέτη. Όπως έδειξαν προηγούμενες έρευνες, περίπου τα τρία τέταρτα των ασθενών με αγγειακό εγκεφαλικό επεισόδιο δεν ερμηνεύουν σωστά τα συμπτώματά τους ως σημάδια εγκεφαλικού επεισοδίου, με αποτέλεσμα να μην αναζητήσουν ιατρική βοήθεια αμέσως μετά την εμφάνιση των συμπτωμάτων (Elkind, 2009; Kothari et al., 1997; Williams et al., 1997). Άλλες έρευνες επικεντρώθηκαν στα υψηλά επίπεδα ανίχνευσης και διαγνωστικής ακρίβειας του εγκεφαλικού επεισοδίου με βάση το μνημονικό ΧΟΠΑ (FAST) (Harbison et al., 2003; Nor et al., 2004; Robinson et al., 2013). Τα αποτελέσματα της εργασίας μας παρέχουν μια πρωταρχική τεκμηρίωση για τη σχέση της ενσωμάτωσης των συμπτωμάτων του εγκεφαλικού επεισοδίου στην αξιολόγηση λήψης ιατρικού ιστορικού. Ενόψει των σοβαρών συνεπειών του εγκεφαλικού επεισοδίου, ο εξεταστής μπορεί να δυσκολεύεται να καταγράψει και να κατανοήσει μία ακατάληπτη ομιλία. Έτσι, δημιουργήσαμε μία μεθοδολογία επεξεργασίας φυσικής γλώσσας μέσω υπολογιστή, για να βοηθήσουμε στην αναγνώριση σχετικών πληροφοριών, ελπίζοντας να καθοδηγήσουμε τον εξεταστή να διακρίνει τη σημασία αυτών των συμπτωμάτων, και αυτό είναι το πρώτο βήμα στην καθοδήγηση της αποκατάστασης: Οι αναμνήσεις των ΧΟΠΑ (FAST) συμπτωμάτων θα πρέπει να αναγνωρίζονται ως κομμάτι της συνέχειας του εγκεφαλικού επεισοδίου και μπορούν να χρησιμοποιηθούν ως δευτερογενής πρόληψη και αυτοφροντίδα (Hassabis, & Maguire, 2007; Schacter, Addis, & Buckner, 2008; Wu, Cheng, & Chiou, 2017).

Καταληκτικά, ο εντοπισμός και η κατανόηση των συμπτωμάτων του εγκεφαλικού επεισοδίου σε έναν γραπτώς αποτυπωμένο διάλογο ασθενή-εξεταστή με τη βοήθεια τεχνικών επεξεργασίας φυσικής γλώσσας μέσω υπολογιστή, μπορεί να αποτελέσει σημείο εκκίνησης για μελλοντική έρευνα. Προκειμένου να εκπαιδευτούν αποτελεσματικά οι επιζήσαντες του εγκεφαλικού επεισοδίου για να φροντίζουν τον εαυτό τους και να βελτιώνονται οι μέθοδοι αποκατάστασης σε παγκόσμιο επίπεδο, πρέπει πρώτα να συνεχίσουμε συστηματικά να υπολογίζουμε ποσοτικά τα αναφερόμενα συμπτώματα κατά τη λήψη του ιατρικού ιστορικού, που είναι και ο βασικός κορμός της διάγνωσης. Οι μελλοντικές μελέτες θα μπορούσαν να εξετάσουν τους τύπους των προβλημάτων επικοινωνίας (αφασία), καθώς σχετίζονται με τα συμπτώματα του εγκεφαλικού επεισοδίου, χρησιμοποιώντας ανάλογες τεχνικές επεξεργασίας φυσικής γλώσσας σε μεγαλύ-

τερο δείγμα επιζώντων του εγκεφαλικού επεισοδίου.

Ηθική έγκριση

Η μελέτη αυτή διεξήχθη σύμφωνα με τα πρότυπα δεοντολογίας της θεσμικής ή/και εθνικής επιτροπής έρευνας και με τη διακήρυξη του Ελσίνκι του 1964 και τις μεταγενέστερες τροποποιήσεις της. Τα υποκείμενα προέρχονταν από το νοσοκομείο Αναγέννηση και ήταν εθελοντές, αριθμός έγκρισης # 1102-2013. Οι συμμετέχοντες υπέγραψαν ένα έντυπο συγκατάθεσης και ενημερώθηκαν στο τέλος της μελέτης. Δε δόθηκε καμία χρηματική αποζημίωση για τη συμμετοχή τους στη μελέτη.

Ευχαριστίες

Θα θέλαμε να ευχαριστήσουμε τη Χρυσανθή Σπύρου για τη συνεισφορά της στην ανάλυση των δεδομένων.

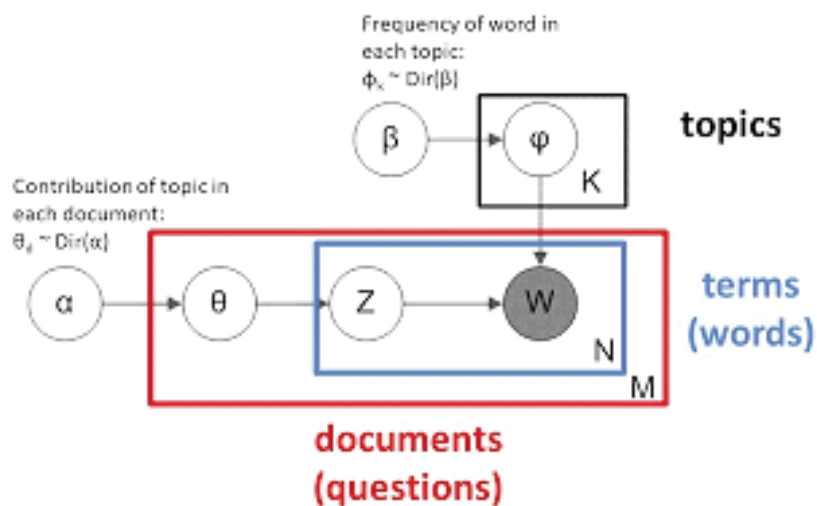
Γνωστοποίηση ενδεχόμενων συγκρούσεων συμφερόντων

Εκ μέρους όλων των συγγραφέων, η υπεύθυνη επικοινωνίας δηλώνει ότι δεν υπάρχει σύγκρουση συμφερόντων.

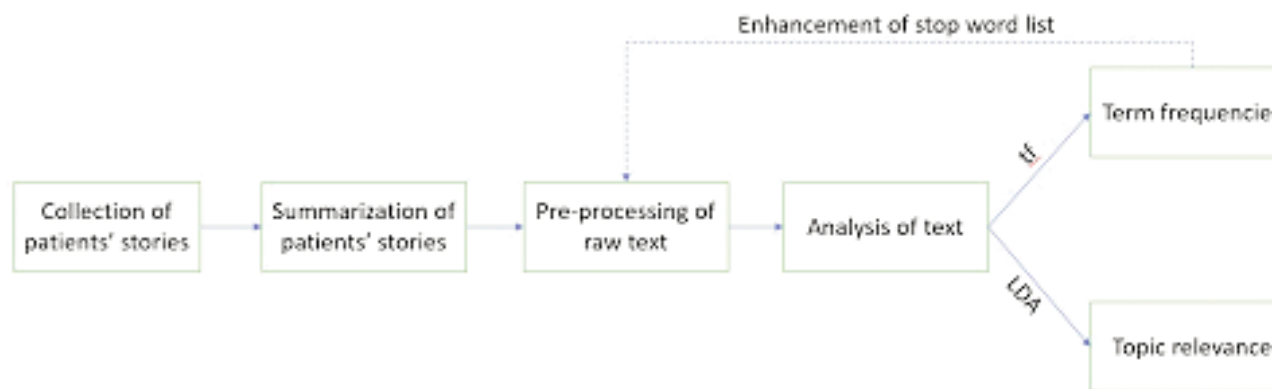
References

- Abbe, A., Grouin, C., Zweigenbaum, P., & Falissard, B. (2016). Text mining applications in psychiatry: a systematic literature review. *International Journal of Methods in Psychiatric Research*, 25(2), 86-100. doi: 10.1002/mpr.1481
- Aroor, S., Singh, R., & Goldstein, L. B. (2017). BE-FAST (Balance, Eyes, Face, Arm, Speech, Time): Reducing the Proportion of Strokes Missed Using the FAST Mnemonic. *Stroke*, 48(2), 479-81. doi: 10.1161/STROKEAHA.116.015169
- Bang, O. Y. (2017). Advances in biomarker for stroke patients: from marker to regulator. *Precision and Future Medicine*, 1(1), 32-42. doi: 10.23838/pfm.2017.00052
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Damico, J. S., Müller, N., & Ball, M. J. (2010). *The handbook of language and speech disorders*. West Sussex: Wiley-Blackwell.
- de Zubizaray, G., McMahon, K., Eastburn, M., & Pringle, A. (2006). Top down influences on lexical selection during spoken word production: A 4T fMRI investigation of refractory effects in picture naming. *Human brain mapping*, 27(11), 864-73. doi: 10.1002/hbm.20227
- Dixon, G., Thornton, E. W., & Young, C. A. (2007). Perceptions of self-efficacy and rehabilitation among neurologically disabled adults. *Clinical Rehabilitation*, 21(3), 230-40.
- Dombrowski, S. U., White, M., Machintosh, J. E., Gellert, P., Araujo-Soares, V., Thomson, R. G., et al (2015). The stroke 'Act FAST' campaign: Remembered but not understood?. *International Journal of Stroke*, 10(3), 324-330.
- Elkind, M. S. V. (2009). Outcomes After Stroke: Risk of Recurrent Ischemic Stroke and Other Events. *The American Journal of Medicine*, 122, S7-S13.
- Feigin, V. L., Forouzanfar, M. H., Krishnamurthi, R., Mensah, G. A., Connor, M., Bennett, D. A., et al. (2014). Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet (London, England)*, 383(9913), 245-254. doi: 10.1016/s0140-6736(13)61953-4
- Feigin, V. L., Norrving, B., & Mensah, G. A. (2017). Global Burden of Stroke. *Circulation Research*, 120(3), 439-448. doi: 10.1161/CIRCRESAHA.116.308413
- Grawemeyer, B., Cox, R., & Lum, C. (2000). AUDIX: A knowledge-based system for speech-therapeutic auditory discrimination exercises. *Medical infobahn für Europe*, 77, 568-72. doi: 10.3233/978-1-60750-921-9-568
- Harbison, J., Hossain, O., Jenkinson, D., Davis, J., Louw, S. J., Gary, A. (2003). Diagnostic accuracy of stroke referrals from primary care, emergency room physicians, and ambulance staff using the face arm speech test. *Stroke*, 34(1), 71-6.
- Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, 11(7), 299-306. doi: 10.1016/j.tics.2007.05.001
- Huang, S., Lv, Z., Guo, Y., Li, L. Zhang, Y. Zhou, L., et al. (2016). Identification of Blood Let-7e-5p as a Biomarker for Ischemic Stroke. *PLoS ONE*, 11(10), e0163951. doi: 10.1371/journal.pone.0163951
- Katan, M., & Elkind, M. S. V. (2018). The Potential Role of Blood Biomarkers in Patients with Ischemic Stroke: An Expert Opinion. *Clinical and Translational Neuroscience*, 1-7. doi: 10.1177/2514183X18768050
- Kim, S. J., Moon, G. J., & Bang, O. Y. (2013). Biomarkers for stroke. *Journal of Stroke*, 15(1), 27-37. doi: 10.5853/jos.2013.15.1.27
- Kleindorfer, D. O., Miller, R., Moomaw, C. J., Alwell, K., Broderick, J. P., Khoury, J., et al. (2007). Designing a Message for Public Education Regarding Stroke: Does FAST Capture Enough Stroke?. *Stroke*, 38, 2864-8. doi: 10.1161/STROKEAHA.107.484329
- Kothari, R., Sauerbeck, L., Jauch, E., Broderick, J., Brott, T., Khoury, J., & Liu, T. (1997). Patients' awareness of stroke signs, symptoms, and risk factors. *Stroke*, 28(10), 1871-5.
- Kurowski, K., & Blumstein, S. E. (2016). Phonetic basis of phonemic paraphasias in aphasia: Evidence for cascading activation. *Cortex*, 75, 193-203. doi: 10.1016/j.cortex.2015.12.005
- Mangset, M., Tor, E. D., Forde, R., & Wyller, T. B. (2008). 'We're just sick people, nothing else': Factors contributing to elderly stroke patients' satisfaction with rehabilitation. *Clinical Rehabilitation*, 22(9): 825-35.
- Moffatt, K., Pourshahid, G., & Baecker, R. M. (2017). Augmentative and alternative communication devices for aphasia: the emerging role of "smart" mobile devices. *Universal Access in the Information Society*, 16, 115. doi: 10.1007/s10209-015-0428-x
- Nor, A. M., McAllister, C., Louw, S. J., Dyker, A. G., Davis, M., Jenkinson, D., & Ford, G.A. (2004). Agreement between ambulance paramedic- and physician-recorded neurological signs with Face Arm Speech Test (FAST) in acute stroke patients. *Stroke*, 35(6), 1355-9.
- Palmer, R., Hughes, H., & Chater, T. (2017). What do people with aphasia want to be able to say? A content analysis of words identified as personally relevant by people with aphasia. *PLoS One*, 12(3), e0174065. doi: 10.1371/journal.pone.0174065
- Proios, H., Malatra, I., Christoudi, C., Willmes, K., Weniger, D., & Milonas, I. (2006). Development and validation of the Aachen Aphasia Test in the Hellenic Language, paper delivered at the International Neuropsychological Society (INS/SVNP/GNP) Meeting "From Plasticity to Rehabilitation", Zurich, Switzerland.
- Ranney, M. L., Choo, E. K., Cunningham, R. M., Spirito, A., Thorsen, M., Mello, M. J., & Morrow, K. (2014). Acceptability, language, and structure of text message-based behavioral interventions for high-risk adolescent females: a qualitative study. *Journal of Adolescent Health*, 55(1), 33-40. doi: 10.1016/j.jadohealth.2013.12.017
- Robinson, T. G., Reid, A., Haunton, V. J., Wilson, A. & Naylor, A.

- R. (2013). The face arm speech test: does it encourage rapid recognition of important stroke warning symptoms? *Emergency Medicine Journal*, 30(6), 467-71. doi: 10.1136/emered-2012-201471
- Roundtree, A. K. (2018). From Engineers' Tweets: Text Mining Social Media for Perspectives on Engineering Communication. In: *IEEE International Professional Communication Conference (ProComm)*, Toronto, Canada, 22-25 July, pp. 6-15. IEEE.
- Salton, G., & Buckley, C. (1998). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), 513-23. doi: 10.1016/0306-4573(88)90021-0
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2008). Episodic Simulation of Future Events. *Annals of the New York Academy of Sciences*, 1124, 39-60. doi: 10.1196/annals.1440.001
- Sponsler, J. L., & Burkhart, J. C. (2016). Aphasia Dx: An Expert Diagnostic System for Aphasia. In: *Telehealth and Assistive Technology / 847: Intelligent Systems and Robotics*, Zurich, Switzerland, January. doi: 10.2316/P.2016.847-005
- Vannobel, J., & Toulotte, J. M. (1992). An Expert System For Aphasia Treatment Using Exercises Concerning Alphabetically Written Numeral Forms. In: *14th annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Paris, France, paper no. 3, pp. 922-3. doi.org/10.1109/IEMBS.1992.5761309
- Ward, A., Gbadebo, A., & Baruah, B. (2015). Using job advertisements to inform curricula design for the key global technical challenges. In: *International Conference on Information Technology Based Higher Education and Training (ITHET)*, Caparica, Lisbon, Portugal, 11-13 June, pp. 1-6. IEEE.
- Williams, L. S., Bruno, A., Rouch, D., & Marriott, D. J. (1997). Stroke patients' knowledge of stroke. Influence on time to presentation. *Stroke*; 28(5), 912-5.
- Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersh, D., Ferguson, A., & Davidson, B. (2011). What people with aphasia want: Their goals according to the ICF. *Aphasiology*, 3, 309-22. doi: 10.1080/02687038.2010.508530
- Wu, W. H., Cheng, W. & Chiou, W. B. (2017). Episodic future thinking about the ideal self induces lower discounting, leading to a decreased tendency toward cheating. *Frontiers in Psychology*, 8, Article ID 287. doi: 10.3389/fpsyg.2017.00287



Εικόνα 1 Η διαδικασία επιλογής θέματος της LDA



Εικόνα 2 Η διαδικασία ανάλυσης κειμένου

Λέξη	Συχν.	Λέξη	Συχν.	Λέξη	Συχν.	Λέξη	Συχν.
ομιλία	43	περιστατικό	8	πόνος	5	ζαλάδα	3
αδυναμία	24	εγκεφαλικό	7	μετά	5	ικανότητα	2
ένικωσε	19	σύγχυση	7	αίσθηση	5	πονακέφαλος	2
χέρι	18	μεταφέρθηκε	6	σύμπτωμα	4	εγχείρηση	2
νοσοκομείο	11	παράλυση	6	έχασε	4	ισχυρίζεται	2
πόδι	11	ξεκίνησε	5	ακολούθησε	4	αστάθεια	2
θυμάται	9	μούδιασμα	5	αδιαθεσία	4	μόνος	2
έντονο	9	αναφέρει	5	πρόβλημα	4	αμέσως	2
ξαφνικά	8	μάτι	5	σώμα	3	πονακέφαλος	2

Πίνακας 1 Συχνότητα μονών λέξεων

Λέξη	Συχν.	Λέξη	Συχν.	Λέξη	Συχν.
αδυνατούσε μιλήσει	52	τέλος αδυνατούσε	8	ότι ένικωσε	4
αδυναμία χέρι	26	περπατάσει καλά	8	επιστημαίνει είναι	4
δεν μπορεί	24	επεισόδιο ξεκίνησε	4	μάτι αδυνάτισε	4
μεταφέρθηκε νοσοκομείο	18	χέρι ξεκίνησε	4	σπαστική κίνηση	4
ομιλία σύγχυση	12	πρώτο σύμπτωμα	4	ξαφνικά αισθάνθηκε	2
άρχισε αδυναμία	10	είχε έλεγχο	4	έχασε αισθήσεις	2
δεν είχε	8	ξεκίνησε αισθάνεται	4	αίσθηση χέρι	2
σφιξει ελαφριά	8	πόδι ξεκίνησε	4	προφανή αδυναμία	2
χέρι πόδι	8	σχετικά έλεγε	4	μπορούσε ακόμη	2

Πίνακας 2 Συχνότητα ζευγών λέξεων

Λέξη	Συχν.	Λέξη	Συχν.
δεν μπορεί μιλήσει	28	ψυχολογία εμφάνισε εγκεφαλικό	6
χάνει αίσθηση χέρι	24	έντονος πονοκέφαλος ζαλάδα	6
άρχισε ατονεί τελικά	14	έχει υποστεί αλλοιώσεις	4
ένωσε μάτι ατονεί	12	είχε εγχείρηση καρδιά	4
πάντα πρόβλημα διαβήτη	12	είχε αιμορραγία στομάχι	4
ακόμη σφίγγει ελαφριά	8	αισθήσεις μεταφέρθηκε νοσοκομείο	6
κάνει σπαστική κίνηση	6	πάντοτε είχε πρόβλημα	2
δεν θυμάται τίποτα	6	σφίγγω παράλυση πόδι	6
ομιλία σύγχυση σκέψη	6	τελικά σαγόνι στράβωσε	2

Πίνακας 3 Συχνότητα τριάδων λέξεων

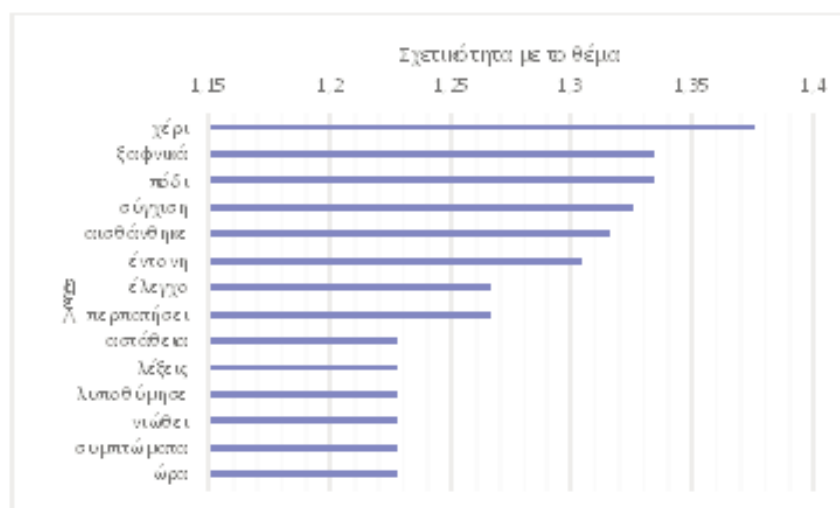


Fig1 Θέμα 1 [χέρι]

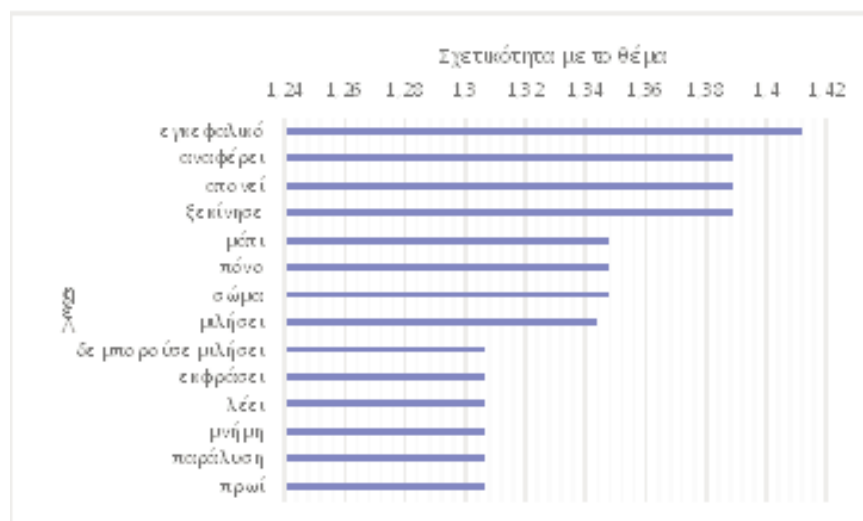


Fig 4 Θέμα 2 [εγκεφαλικό]

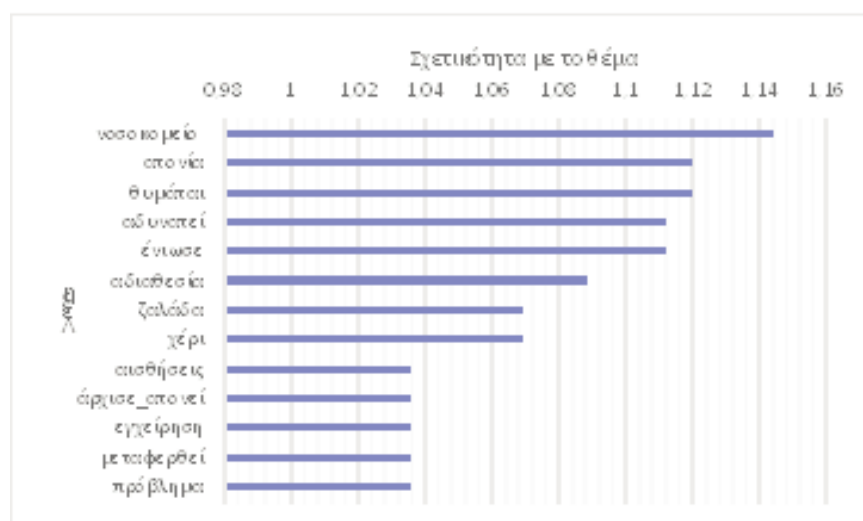


Fig 5 Θέμα 3 [Γενικά θέματα σχετικά με το εγκεφαλικό]