



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΦΙΛΟΣΟΦΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΦΙΛΟΛΟΓΙΑΣ
ΤΟΜΕΑΣ ΓΛΩΣΣΟΛΟΓΙΑΣ**



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ
ΡΟΜΠΟΤΙΚΗΣ**

**ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟΓΛΩΣΣΙΑ VII»**

Αυτόματη ανάλυση κατευθυντικών λεκτικών πράξεων

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΚΩΝΣΤΑΝΤΙΝΑ ΠΑΠΑΝΙΚΟΛΑΟΥ**

**Επιβλέπων: Χάρης Παπαγεωργίου
Ερευνητής Β'
Ινστιτούτο Επεξεργασίας του Λόγου**

Επιτροπή:

**Νίκος Γλάρος
Ειδικός Λειτουργικός Επιστήμονας Α'
Ινστιτούτο Επεξεργασίας του Λόγου**

**Ελένη Παναρέτου
Αναπληρώτρια Καθηγήτρια
Ε.Κ.Π.Α.**

**ΑΘΗΝΑ
ΙΟΥΝΙΟΣ 2014**



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΛΙΣΤΡΙΑΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΦΙΛΟΣΟΦΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΦΙΛΟΛΟΓΙΑΣ
ΤΟΜΕΑΣ ΓΛΩΣΣΟΛΟΓΙΑΣ**



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ
ΡΟΜΠΟΤΙΚΗΣ**

**ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ «ΤΕΧΝΟΓΛΩΣΣΙΑ VII»**

Αυτόματη ανάλυση κατευθυντικών λεκτικών πράξεων

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΚΩΝΣΤΑΝΤΙΝΑ ΠΑΠΑΝΙΚΟΛΑΟΥ**

**Επιβλέπων: Χάρης Παπαγεωργίου
Ερευνητής Β'
Ινστιτούτο Επεξεργασίας του Λόγου**

Επιτροπή:

**Νίκος Γλάρος
Ειδικός Λειτουργικός Επιστήμονας Α'
Ινστιτούτο Επεξεργασίας του Λόγου**

**Ελένη Παναρέτου
Αναπληρώτρια
Καθηγήτρια Ε.Κ.Π.Α.**

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18^η Ιουνίου 2014.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....

ΑΘΗΝΑ, ΙΟΥΝΙΟΣ 2014

Στον μπαμπά μου

Ευχαριστίες

Θεωρώ σημαντικό να ευχαριστήσω όλους εκείνους που συνέβαλαν με οποιονδήποτε τρόπο στην ολοκλήρωση της παρούσας διπλωματικής εργασίας. Θερμές ευχαριστίες οφείλω στον κ. Χάρη Παπαγεωργίου, επιβλέποντα της εργασίας μου, για τους ορίζοντες που μου άνοιξε φέρνοντάς με σε επαφή με ένα συναρπαστικό γνωστικό αντικείμενο. Η υποστήριξή του, επιστημονική και ηθική, κατά τη διάρκεια της προσπάθειας ήταν καθοριστική.

Ιδιαίτερες ευχαριστίες θα ήθελα να απευθύνω και στην κ. Μαρία Ποντίκη, για την αμέριστη συμπαράστασή της σε όλη τη διαδικασία εκπόνησης της διπλωματικής μου εργασίας. Οι πολύτιμες συμβουλές αλλά και οι εύστοχες παρατηρήσεις της έπαιξαν αποφασιστικό ρόλο στην επιτυχή διεκπεραίωση της μελέτης αυτής. Εκτός όμως από την καθοδήγησή της σε επιστημονικό επίπεδο, η υπομονή και η συνεχής ενθάρρυνσή της σε όλη την πορεία της προσπάθειάς μου ήταν εξαιρετικά πολύτιμες.

Δεν θα μπορούσα να μην εκφράσω τη βαθιά μου ευγνωμοσύνη στους γονείς μου γιατί με την πίστη τους σε εμένα και την συνεχή υποστήριξή τους σε κάθε επίπεδο όλα αυτά τα χρόνια, συνέβαλαν στη διαμόρφωσή μου, ως ανθρώπου αλλά και ως επιστήμονα. Ιδιαίτερα θα ήθελα να ευχαριστήσω τον μπαμπά μου για όλα αυτά που αγόγγυστα έκανε για μένα και για όλα εκείνα που ήθελε ακόμα να κάνει. Τα φυλάω σαν ανεκτίμητη παρακαταθήκη.

Θα ήταν ασφαλώς άδικο να παραλείψω να ευχαριστήσω ειλικρινά τον σύζυγό μου, ο οποίος με στήριξε και με στηρίζει χωρίς όρους, επιδεικνύοντας παροιμιώδη υπομονή και μεταδίδοντάς μου την αισιοδοξία που τον χαρακτηρίζει ως άνθρωπο. Του οφείλω πολλά από όσα έχω καταφέρει, αφού χωρίς την ανιδιοτελή βοήθεια που μου προσφέρει απλόχερα τα χρόνια του κοινού μας βίου, δεν θα είχαν πραγματοποιηθεί.

Περίληψη

Στόχος της παρούσας εργασίας είναι η αυτόματη ανάλυση κατευθυντικών λεκτικών πράξεων, και πιο συγκεκριμένα ο εντοπισμός συστάσεων και αιτημάτων μέσα σε κείμενα παραγόμενα από χρήστες του διαδικτύου. Η εύρεση αυτών των σημασιολογικών κατηγοριών μέσα σε γλωσσικό υλικό ανεξαρτήτως θεματικού πεδίου, μπορεί να βρει εφαρμογή τόσο σε επίπεδο εμπορικό για τη βελτίωση στρατηγικών προώθησης και πωλήσεων προϊόντων, όσο και στην πολιτική για τη διαμόρφωση εκλογικών στρατηγικών, ενώ μπορεί να συμβάλει και στην εθνική ασφάλεια. Η θεωρία των λεκτικών πράξεων αποτελεί το θεωρητικό υπόβαθρο, ενώ η προσέγγιση στηρίζεται στη μελέτη γλωσσικών δεδομένων. Η μεθοδολογία που χρησιμοποιείται βασίζεται στη δημιουργία γλωσσικών πόρων και τη μοντελοποίησή τους μέσω κανόνων. Η αξιολόγηση έδειξε ότι η μέθοδος παρουσιάζει ικανοποιητικά αποτελέσματα σε επίπεδο ακρίβειας.

Abstract

The aim of this thesis is the automatic analysis of directive speech acts, and specifically, the detection of recommendations and requests within user generated content. These insights' detection in domain independent text, can be used for marketing and sales, but also by politicians to influence the electorate. Furthermore, it may contribute to improving national security. Speech acts theory consists the theoretical background, while the approach is data driven. The methodology used is linguistically driven and is based on creating language resources and modeling them with rules. The evaluation showed that the results of our method are quite efficient concerning the precision.

Περιεχόμενα

Ευχαριστίες	1
Περίληψη	2
Abstract	3
Περιεχόμενα.....	4
Εισαγωγή	6
1. Θεωρητικό και υπολογιστικό υπόβαθρο.....	9
1.1. Λεκτικές πράξεις	9
1.1.1. Κατευθυντικές λεκτικές πράξεις	12
1.2. Σχετική έρευνα.....	14
1.2.1. Λεκτικές πράξεις	14
1.2.2. Συστάσεις και αιτήματα.....	15
2. Γλωσσικά δεδομένα	17
2.1. Σώματα κειμένων	17
2.1.1. Γλωσσικά χαρακτηριστικά	22
2.1.2. Γλωσσικά μέσα έκφρασης κατευθυντικών πράξεων.....	25
3. Μεθοδολογία - Υλοποίηση	29
3.1. Στόχος - Μεθοδολογία	29
3.2. Σημασιολογικές κατηγορίες (insights).....	30
3.3. Λεξικά	31
3.4. Δομικά σχήματα (Patterns)	34
3.5. Κανόνες.....	38
4. Εφαρμογή – Αξιολόγηση	42
4.1. Εφαρμογή.....	42
4.2. Αξιολόγηση	45
4.3. Σχολιασμός αποτελεσμάτων – Ανάλυση λαθών	49
5. Συμπεράσματα – Περαιτέρω έρευνα	55
5.1. Συμπεράσματα	55
5.2. Περαιτέρω έρευνα – Εφαρμογές.....	56

ΒΙΒΛΙΟΓΡΑΦΙΑ	59
ΠΑΡΑΡΤΗΜΑ Ι – ΛΕΞΙΚΟ ΡΗΜΑΤΩΝ	62
ΠΑΡΑΡΤΗΜΑ ΙΙ – ΛΕΞΙΚΟ ΕΚΦΡΑΣΕΩΝ	66

Εισαγωγή

Το διαδίκτυο έχει αλλάξει τον τρόπο με τον οποίο οι άνθρωποι επικοινωνούν και ανταλλάσσουν απόψεις. Μέσω του διαδικτύου και ιδίως των διαφόρων μέσων κοινωνικής δικτύωσης (π.χ. Facebook, Twitter, online fora, blogs), οι χρήστες εκφράζουν απόψεις, επιθυμίες, συναισθήματα, μοιράζονται προσωπικά τους δεδομένα, δηλώνουν τις προθέσεις τους και καταθέτουν τις εμπειρίες τους και μάλιστα μέσα σε μια κοινότητα, έχοντας δηλαδή ένα ευρύ ακροατήριο. Αυτού του είδους το υλικό είναι ιδιαίτερα ενδιαφέρον για τη γλωσσική έρευνα, καθώς είναι αυθεντικό, αφού παράγεται από τους ίδιους τους χρήστες, ενώ προέρχεται από ένα πολυποίκιλο δείγμα, δεδομένου ότι το διαδίκτυο έχει μεγάλη απήχηση στους νέους αλλά και σε ανθρώπους όλων των ηλικιών, ανεξαρτήτως φύλου, μορφωτικού επιπέδου, κοινωνικής και οικονομικής κατάστασης. Υπάρχουν διάφοροι τρόποι αξιοποίησης αυτού του υλικού, καθώς οι τύποι πληροφορίας που μπορούν να εξαχθούν είναι ποικίλοι. Πέρα από τους γενικότερους, δηλαδή τις απόψεις και τα συναισθήματα, είναι δυνατό να εντοπιστούν στοιχεία όπως προθέσεις, συστάσεις, απαιτήσεις, ανάλογα με το πεδίο ενδιαφέροντος της εκάστοτε έρευνας.

Θέμα της παρούσας εργασίας είναι η αυτόματη ανάλυση των κατευθυντικών λεκτικών πράξεων (ΚΛΠ), δηλαδή εκφράσεων με τις οποίες ένας ομιλητής ή συγγραφέας ζητά από τον ακροατή ή αποδέκτη να κάνει κάτι, στο πλαίσιο της θεωρίας των λεκτικών πράξεων (Austin, 1962, Searle, 1976, βλ. 1.1.). Συγκεκριμένα, στόχος είναι η αυτόματη αναγνώριση συστάσεων (recommendations) και αιτημάτων (requests) στην αγγλική γλώσσα, μέσα σε υλικό που παράγεται από χρήστες σε μέσα κοινωνικής δικτύωσης (π.χ. Twitter) και fora με κριτικές προϊόντων.

Η αυτόματη ανάλυση συστάσεων και αιτημάτων έχει πολλαπλές και ενδιαφέρουσες εφαρμογές σε τομείς όπως η πρόβλεψη και η λήψη αποφάσεων σε διαφορετικά θεματικά πεδία. Οι εταιρείες μπορούν να χρησιμοποιήσουν αυτό το υλικό προκειμένου να εκτιμήσουν τον αντίκτυπο που έχουν τα προϊόντα και οι υπηρεσίες τους γενικότερα στην αγορά ή σε συγκεκριμένο κοινό, να χαράξουν ή να τροποποιήσουν τις στρατηγικές των πωλήσεων αλλά και της προώθησης των προϊόντων τους, με σκοπό να αυξήσουν τις πωλήσεις τους. Αντίστοιχα, κυβερνήσεις και πολιτικοί μπορούν να ελέγξουν την αποδοχή ή μη που έχουν οι ίδιοι και οι στρατηγικές τους από το εκλογικό σώμα. Όμως και οι ίδιοι οι χρήστες του διαδικτύου, όταν θέλουν να αγοράσουν ένα προϊόν, να βρουν ένα μέρος για διακοπές ή για

διασκέδαση, ολοένα και περισσότερο αποφασίζουν με βάση τις απόψεις άλλων που εκφράζονται στις διάφορες σελίδες.

Το υλικό που προέρχεται από ηλεκτρονικά μέσα έχει το πλεονέκτημα ότι είναι άμεσα και σχετικά εύκολα προσβάσιμο, συγκριτικά τουλάχιστον με παλαιότερα, όταν χρειαζόταν να δημιουργηθούν ερωτηματολόγια με συγκεκριμένους σκοπούς και η συλλογή του υλικού ήταν εξαιρετικά χρονοβόρα και δύσκολη. Επιπλέον, δεν υπάρχουν περιορισμοί ως προς τη θεματική, καθώς οι χρήστες του διαδικτύου εκφράζουν τη γνώμη τους επί παντός επιστητού. Επιπλέον, δεδομένου ότι η άποψή τους διατυπώνεται ελεύθερα, όπου και όπως οι ίδιοι επιλέγουν και για όποιο θέμα επιθυμούν, δεν έχουν προκαταλήψεις, πέρα από την υποκειμενικότητά τους. Όλα αυτά τα χαρακτηριστικά καθιστούν το υλικό αυτό, πολύτιμο εργαλείο για τομείς έρευνας όπως η εξαγωγή πληροφορίας και η ανάλυση συναισθήματος.

Η ανάλυση συναισθήματος (Sentiment Analysis), ή εξαγωγή άποψης (Opinion Mining) είναι το πεδίο της έρευνας που αναλύει τις γνώμες, τα συναισθήματα, τις κρίσεις, τις αξιολογήσεις των ανθρώπων απέναντι σε οντότητες όπως προϊόντα, υπηρεσίες, οργανισμούς, άτομα, γεγονότα, θέματα και τα χαρακτηριστικά τους. Αυτός ο τομέας της έρευνας κυρίως εστιάζει σε απόψεις οι οποίες εκφράζουν ή υπονοούν θετικά ή αρνητικά συναισθήματα. (Liu, 2012:7). Η ανάλυση συναισθήματος είναι τομέας της Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing – NLP), η οποία με τη σειρά της ανήκει στον κλάδο της Τεχνητής Νοημοσύνης (Artificial Intelligence – AI), ενώ συνδέεται άμεσα με έναν άλλο ερευνητικό τομέα, αυτόν της εξαγωγής πληροφορίας (Information Extraction). Ο γενικότερος στόχος αυτού του πεδίου είναι η εξαγωγή σημασιολογικής πληροφορίας μέσα από κείμενα τα οποία προέρχονται από ηλεκτρονικές πηγές. Πιο συγκεκριμένα, στοχεύει στη συλλογή συγκεκριμένων πληροφοριών και την οργάνωσή τους με βάση τις μεταξύ τους σχέσεις, σε μορφή τέτοια που να μπορεί να χρησιμοποιηθεί για περαιτέρω επεξεργασία.

Μέσα σε αυτά τα πλαίσια εντάσσεται και η παρούσα εργασία. Η προσέγγιση που ακολουθείται στηρίζεται στην ανάλυση γλωσσικών δεδομένων (data driven), ενώ προτεραιότητα αποτελεί η ακρίβεια των αποτελεσμάτων (precision oriented). Η μέθοδος που ακολουθείται είναι στη βάση της γλωσσολογική, καθώς για τον σκοπό της αυτόματης ανάλυσης ΚΛΠ δημιουργήθηκε ένα εργαλείο, το οποίο ενσωματώνει σημασιολογικά στοιχεία έκφρασης κατευθυντικότητας αλλά και συντακτικά σχήματα μέσα στα οποία εντάσσονται αυτά τα λεξικά μέσα. Οι κανόνες που υλοποιούν αυτά τα

σχήματα αξιοποιούν την επιφανειακή σύνταξη, αντιμετωπίζουν με άλλα λόγια το κείμενο ως γραμμική παράταξη λέξεων.

Η εργασία διαρθρώνεται σε πέντε κεφάλαια. Το θεωρητικό και υπολογιστικό υπόβαθρο στο οποίο στηρίζεται παρουσιάζεται σύντομα στο κεφάλαιο 1. Η έρευνα αναλύεται σε δύο στάδια. Στο πρώτο, παρουσιάζεται η ανάλυση των γλωσσικών δεδομένων και η μελέτη των λεκτικών μέσων και των συντακτικών δομών μέσω των οποίων εκφράζονται ΚΛΠ στην αγγλική γλώσσα (Κεφάλαιο 2). Στο δεύτερο στάδιο στόχος είναι η δημιουργία λεξικών πόρων και η μοντελοποίηση των δομών που εντοπίστηκαν σε επίπεδο επιφανειακής σύνταξης για την αυτόματη αναγνώρισή τους μέσω ενός συνόλου κανόνων που αποτελούν τη γραμματική (Κεφάλαιο 3). Τα αποτελέσματα από την αξιολόγηση του εργαλείου σε ένα νέο σύνολο δεδομένων παρουσιάζονται στο Κεφάλαιο 4, ενώ στο Κεφάλαιο 5 παρατίθενται τα συνολικά συμπεράσματα της μελέτης καθώς και οι δυνατότητες για περαιτέρω έρευνα.

1. Θεωρητικό και υπολογιστικό υπόβαθρο

1.1. Λεκτικές πράξεις

Βασικό αξίωμα της θεωρίας των λεκτικών πράξεων¹ είναι ότι οι άνθρωποι δε χρησιμοποιούν τη γλώσσα μόνο για να πουν κάτι, να κάνουν μια δήλωση (statement) αλλά για να κάνουν κάτι, να επιτελέσουν πράξεις. Συνεπώς, κάποια εκφωνήματα δεν περιγράφουν απλώς μια πράξη, την επιτελούν (Austin, 1962:6). Ο Austin διέκρινε τα εκφωνήματα σε: α) διαπιστωτικά (constative), τα οποία μπορούν να χαρακτηριστούν ως αληθή ή ψευδή (true or false). Τέτοιο παράδειγμα αποτελεί το εκφώνημα «*Το ποτήρι είναι στο τραπέζι*». Και β) επιτελεστικά (performative), τα οποία είναι δυνατό να χαρακτηριστούν ως επιτυχή ή ανεπιτυχή (happy or unhappy) με βάση κάποιες συνθήκες επιτυχίας (felicity conditions²). Π.χ. «*Υπόσχομαι ότι έρθω αύριο*».

Ο Austin υποστήριξε ότι στα επιτελεστικά εκφωνήματα, εμφανίζεται ένα ρήμα το οποίο περιγράφει ρητά τη λεκτική πράξη και για τον λόγο αυτόν τα ονόμασε ρητά επιτελεστικά εκφωνήματα (explicit performatives) και τα ρήματα αντίστοιχα ρητά επιτελεστικά ρήματα. Αυτά τα αντιπαραθέτει με τα μη ρητά (implicit) επιτελεστικά (Austin, 1962:32) στα οποία δεν εμφανίζεται κάποιο συγκεκριμένο λεξικό μέσο, ο αποδέκτης ωστόσο του μηνύματος αντιλαμβάνεται την προσλεκτική ισχύ (illocutionary force) του εκφωνήματος, κατανοεί δηλαδή τι σημαίνει και ποια λεκτική πράξη αντιπροσωπεύει. Για παράδειγμα, το εκφώνημα «*Φύγε*» μπορεί να αντικαθιστά τη φράση «*Σε διατάζω να φύγεις*» επιτελώντας την ίδια λεκτική πράξη της διαταγής. Μάλιστα σε αρκετές περιπτώσεις η χρήση του ρητού επιτελεστικού ρήματος είναι εντελώς ασυνήθιστη ή περιορίζεται σε πολύ συγκεκριμένα περιβάλλοντα. Έτσι, αν θέλουμε να προσκαλέσουμε κάποιον στο σπίτι μας μπορούμε να πούμε: «*Έλα στο σπίτι μου.*» ή «*Θέλεις να έρθεις στο σπίτι μου;*», ενώ μόνο σε εξαιρετικά επίσημες περιστάσεις θα συναντήσουμε το εκφώνημα «*Σε προσκαλώ στο σπίτι μου.*».

¹ Αν και στη βιβλιογραφία χρησιμοποιούνται τόσο ο όρος γλωσσικές πράξεις (βλ. Γεωργακοπούλου-Γούτσος 1999:22) όσο και ο όρος λεκτικές πράξεις (βλ. Φιλιππάκη-Warburton, 1992:299), εδώ υιοθετείται ο δεύτερος καθώς φαίνεται να επικρατεί. Αλλωστε όπως το έθεσε και ο Verschueren: «Ο Austin έδωσε ένα όνομα στα “πράγματα που κάνουν οι άνθρωποι με τις λέξεις”: τα ονόμασε λεκτικές πράξεις” (speech acts) (ίσως ο όρος “γλωσσικές πράξεις” (linguistic acts) να ήταν πιο κατάλληλος, εφόσον περιλαμβάνονται όχι μόνο οι προφορικές αλλά και οι γραπτές λεκτικές πράξεις, παρόλα αυτά υποθέτουμε ότι ο αναγνώστης -όπως κι εγώ- θα είναι πρόθυμος να θυσιάσει κάπως τη σαφήνεια για χάρη της συνέχειας και της παράδοσης)» (Verschueren 1980:3)

² Για περισσότερα σχετικά με τις συνθήκες επιτυχίας, βλ. Austin, 1962:14-15.

Σε αυτά τα πλαίσια, ο Austin υποστήριξε ότι κάθε εκφώνημα εκτελεί τρία είδη πράξεων: α) την εκφωνητική πράξη (locutionary act), την πράξη δηλαδή της εκφώνησης μιας πρότασης με ορισμένη σημασία και αναφορά, β) την προσλεκτική πράξη (illocutionary act). Κάθε πρόταση που εκφωνείται από έναν ομιλητή σε δεδομένη περίπτωση, έχει μια συγκεκριμένη προσλεκτική ισχύ, όπως υπόσχεση, διαταγή, παράκληση. Ανάλογα με αυτή, ο ομιλητής κάνει αντίστοιχα την προσλεκτική πράξη της υπόσχεσης, της διαταγής, της παράκλησης. Και γ) τη δια-λεκτική πράξη (perlocutionary act)³, με άλλα λόγια τα αποτελέσματα ή τις αντιδράσεις του ακροατή που αναμένει ο ομιλητής με την πρόταση που εκφωνεί. Για παράδειγμα, στο εκφώνημα «Μου δίνεις αυτό το βιβλίο;», η εκφωνητική πράξη είναι το ότι σε δεδομένη περίπτωση ο ομιλητής επέλεξε να εκφωνήσει την πρόταση αυτή με τη συγκεκριμένη δομή. Η προσλεκτική πράξη είναι μια παράκληση που θα μπορούσε ρητά να εκφραστεί ως «Σε παρακαλώ να μου δώσεις αυτό το βιβλίο.». Τέλος, η επιδιωκόμενη από τον ομιλητή δια-λεκτική πράξη είναι να του δώσει ο ακροατής το βιβλίο.

Ανάλογα με την προσλεκτική τους ισχύ, τα εκφωνήματα κατατάσσονται στις εξής κατηγορίες (Austin, 1962:150): α) Ετυμηγορικά (verdictives), με τα οποία δίνεται μια ετυμηγορία, μια κρίση (π.χ. καταδικάζω), β) Επιβάλλοντα - εξασκητικά (exercitives), που δηλώνουν άσκηση εξουσίας, δικαιωμάτων ή επιρροής (π.χ. διατάζω), γ) Δεσμευτικά (commissives), τα οποία προϋποθέτουν κάποια δέσμευση του ομιλητή είτε για την ανάληψη κάποιας υποχρέωσης είτε για την πρόθεσή του (π.χ. υπόσχομαι), δ) Συμπεριφορικά (habbitives), τα οποία σχετίζονται με στάσεις και κοινωνικές συμπεριφορές (π.χ. απολογούμαι) και ε) Επεξηγηματικά (expositives), με τα οποία γίνεται διασαφήνιση λόγων ή επιχειρημάτων (π.χ. συνάγω). Αυτή η τελευταία κατηγορία δηλώνει και ο ίδιος ο Austin ότι είναι δύσκολο να οριστεί, ενώ γενικότερα παρατηρεί ότι δεν είναι πάντα εύκολη η κατάταξη ενός εκφωνήματος σε μια και μόνη κατηγορία και μπορούν να υπάρχουν αλληλοεπικαλύψεις.

Ωστόσο, η πιο διαδεδομένη και καθιερωμένη ταξινόμηση των λεκτικών πράξεων είναι αυτή του J. Searle (Searle, 1976, 1979), μαθητή του Austin. Σύμφωνα με τον Searle, υπάρχουν πέντε βασικά είδη λεκτικών πράξεων (Searle, 1976:1)⁴:

³ Για τη μετάφραση των όρων locutionary, illocutionary και perlocutionary acts, καθώς δεν υπάρχει απόλυτη συμφωνία στη βιβλιογραφία, υιοθετούνται στην παρούσα εργασία οι ελληνικοί όροι εκφωνητικές, προσλεκτικές και δια-λεκτικές πράξεις, δηλαδή η μετάφραση της Φιλιππάκη (Φιλιππάκη-Warburton, 1992:301), επειδή θεωρώ ότι είναι πιο αντιπροσωπευτικές των αγγλικών όρων.

⁴ Ο Searle κάνει λόγο για τις κατηγορίες των προσλεκτικών πράξεων, ωστόσο στη βιβλιογραφία αυτές αναφέρονται ως (κυρίως) λεκτικές πράξεις. Όπως το θέτει η Φιλιππάκη: «Από τις τρεις αυτές πράξεις εκείνη που έχει προσελκύσει το μεγαλύτερο ενδιαφέρον των μελετητών της Πραγματολογίας είναι η

α) Αποφαντικές ή βεβαιωτικές (representatives or assertives). Στην κατηγορία αυτή ανήκουν οι λεκτικές πράξεις που δεσμεύουν σε κάποιο βαθμό τον ομιλητή ως προς την αλήθεια του περιεχομένου του εκφωνήματός του. Το λογικό περιεχόμενο των προτάσεων αυτών μπορεί να ελεγχθεί ως προς την αλήθεια ή το ψεύδος του. Π.χ. “*I predict he will come*” και «(Βεβαιώνω ότι) η κυρία Παπαδοπούλου βρίσκεται στο γραφείο της.»

β) Κατευθυντικές (directives). Με αυτές ο ομιλητής προσπαθεί να πείσει ή να κατευθύνει τον ακροατή να κάνει κάτι. Για παράδειγμα, ο ομιλητής μπορεί να κάνει μια ερώτηση, μια αίτηση ή μια πρόσκληση και να περιμένει την ανάλογη αντίδραση από τον ακροατή (Qadir & Riloff, 2011:3). Π.χ. “*I order you to leave*” και «(Σου ζητάω να) μη μου μιλάς έτσι.»

γ) Δεσμευτικές (commissives). Ο ομιλητής δεσμεύεται για μια μελλοντική δράση ή ενέργεια, δεσμεύεται – και πάλι σε κάποιον βαθμό – να πραγματοποιήσει αυτό που λέει. Π.χ. “*I promise to pay you the money*” και «Υπόσχομαι να έρθω αύριο.»

δ) Εκφραστικές (expressives). Αυτές εκφράζουν την ψυχολογική κατάσταση ή στάση του ομιλητή απέναντι στο περιεχόμενο του εκφωνήματος. Τυπικές περιπτώσεις εκφραστικών λεκτικών πράξεων είναι όταν ο ομιλητής ευχαριστεί, απολογείται, συγχαίρει ή καλωσορίζει τον ακροατή. Π.χ. “*I apologize for behaving badly*” και «Σε ευχαριστώ για τη βοήθεια.»

ε) Διακηρυκτικές (declarations). Πρόκειται για εκφωνήματα τα οποία επιφέρουν μια άμεση αλλαγή στο περιβάλλον. Με αυτά ο ομιλητής επιτελεί μια συγκεκριμένη πράξη. Π.χ. “*I now pronounce you man and wife*” και «Σε καταδικάζω σε τρία χρόνια φυλάκιση.» Χαρακτηριστικό αυτών των λεκτικών πράξεων είναι ότι προκειμένου να ισχύουν πρέπει να πληρούνται ορισμένες προϋποθέσεις. Στο παραπάνω παράδειγμα, θα πρέπει αυτός που το εκφωνεί να είναι σε θέση να καταδικάσει κάποιον, να είναι λόγου χάρη δικαστής.

Η ταξινόμηση αυτή δεν είναι απόλυτη, αφού συνήθως η προσλεκτική ισχύς μιας πρότασης γίνεται κατανοητή σε συσχέτιση με τις περιστάσεις εκφοράς της. Η χρήση ρημάτων χαρακτηριστικών κάθε κατηγορίας (αδρομερώς αυτά που ο Austin ονόμαζε επιτελεστικά) μπορεί να αποτελεί ένα κριτήριο. Για παράδειγμα το ρήμα «υπόσχομαι», εκφράζει υπόσχεση και άρα δεσμευτική λεκτική πράξη. Στις υπόλοιπες περιπτώσεις

δεύτερη, δηλαδή η προσλεκτική ή λεκτική πράξη, ίσως επειδή προσφέρεται περισσότερο για αντικειμενική περιγραφή και επειδή η προσλεκτική και η διαλεκτική πράξη δεν ξεχωρίζουν πάντα καθαρά» (Φιλίππακη-Warburton, 1992:301).

όμως, τα πράγματα δεν είναι τόσο ξεκάθαρα. Λόγου χάρη, το εκφώνημα «Θα είμαι στο αεροδρόμιο στις 7» είναι δεσμευτικό αφού ο ομιλητής αναλαμβάνει την υποχρέωση να είναι τη συγκεκριμένη ώρα στο αεροδρόμιο για να παραλάβει κάποιον, είναι ωστόσο και βεβαιωτικό καθώς το λογικό του περιεχόμενο μπορεί να χαρακτηριστεί ως αληθές ή ψευδές.

Συνεπώς, υπάρχουν περιπτώσεις στις οποίες ο ομιλητής εκφωνεί μια πρόταση με την οποία εννοεί κάτι άλλο από αυτό που λέει (π.χ. όταν χρησιμοποιεί ειρωνεία ή μεταφορά) ή ακόμα μπορεί να εννοεί αυτό που λέει αλλά και κάτι άλλο (Searle, 1975b: 59). Λόγου χάρη, με το εκφώνημα «*Κάνει πολύ ζέστη εδώ μέσα*», ο ομιλητής μπορεί απλώς να αποφαίνεται, να βεβαιώνει κάτι, ταυτόχρονα όμως να προσπαθεί να πείσει τον ακροατή να ανοίξει το παράθυρο, άρα να πρόκειται για μια ΚΛΠ. Αυτού του είδους τις λεκτικές πράξεις ο Searle τις ονόμασε έμμεσες λεκτικές πράξεις. Έχουν διατυπωθεί διάφορες θεωρίες σχετικά με το πώς αντιλαμβάνονται οι ακροατές τις έμμεσες λεκτικές πράξεις⁵. Καθώς η στόχευση της εργασίας είναι να εντοπιστούν συγκεκριμένοι τύποι λεκτικών πράξεων με βάση κυρίως την επιφανειακή συντακτική τους δομή, το ενδιαφέρον εστιάζεται μόνο σε παγιωμένους τρόπους έκφρασης έμμεσων λεκτικών πράξεων, όπως είναι οι ερωτήσεις του τύπου «*Μπορείς να μου δώσεις το αλάτι;*» (“*Can you pass the salt?*”).

1.1.1. Κατευθυντικές λεκτικές πράξεις

Ως κατευθυντικές ορίζονται οι λεκτικές πράξεις μέσω των οποίων ο ομιλητής κατευθύνει ή επηρεάζει την επακόλουθη συμπεριφορά του ακροατή. Με άλλα λόγια, ο ομιλητής επιδιώκει να κάνει κάτι ο ακροατής. Όπως προκύπτει από τον ορισμό αυτόν ότι οι ΚΛΠ ενέχουν μια εξουσία ή επιβολή από την πλευρά του ομιλητή η οποία φανερώνεται σε μικρότερο ή μεγαλύτερο βαθμό, ανάλογα και με τα γλωσσικά μέσα που κάθε φορά χρησιμοποιούνται. Παρόλα αυτά οι ΚΛΠ είναι καθολικές πράξεις επικοινωνίας, ενώ ποικίλλουν τα μέσα με τα οποία εκφράζονται, από άμεσους τύπους, π.χ. «*Σε διατάζω να φύγεις*», ως τους πλέον έμμεσους, όπως π.χ. «*Κάνει κρύο εδώ μέσα*». Επιπλέον, υπάρχουν και πολλοί ενδιάμεσοι τρόποι, ανάλογα με τα πραγματολογικά στοιχεία που εμπλέκονται στις περιστάσεις επικοινωνίας, δηλαδή ευγένεια, φιλικότητα κ.ά.

⁵ Ενδεικτικά βλ. Searle 1975b, Morgan 1978, Bach & Harnish 1979.

Σε σημασιολογικό επίπεδο, οι επικοινωνιακές λειτουργίες που περιλαμβάνονται στην κατηγορία των ΚΛΠ είναι ποικίλες. Από τις πιο χαρακτηριστικές είναι οι ακόλουθες:

- εντολές, διαταγές, π.χ. «*Βγες έξω!*», «*Παρουσιάστε!*».
- οδηγίες, π.χ. «*Βάλε τους φακέλους στο πάνω ράφι.*», «*Πριν την πρώτη χρήση ανοίξτε προσεκτικά το καπάκι*».
- παρακλήσεις, αιτήματα, π.χ. «*Μπορείς να μου βάλεις λίγο καφέ;*».
- προτροπές, συστάσεις, π.χ. «*Σου συνιστώ ανεπιφύλακτα αυτό το ξενοδοχείο.*», «*Διάβασε αυτό το βιβλίο. Θα σου αρέσει πολύ*».
- συμβουλές, π.χ. «*Θα σε συμβούλευα να μη μιλάς τόσο πολύ*», «*Θα ήταν καλό να ξεκουραστείς λίγες μέρες*».

Οι παραπάνω έννοιες δεν υπόκεινται σε διαγλωσσικούς φραγμούς, αλλά εμφανίζονται σε κάθε γλώσσα. Διαφέρουν βέβαια, όπως είναι αναμενόμενο, οι τρόποι με τους οποίους εκφράζονται σε κάθε γλώσσα. Απ' την άλλη μεριά, μολονότι αναγνωρίζονται από όλους τους ομιλητές, δεν είναι πάντα ξεκάθαρα τα μεταξύ τους όρια.

Οποιαδήποτε απόπειρα ταξινόμησης, θα έπρεπε να λάβει υπόψη ορισμένα κριτήρια, όπως π.χ. α) το ποιος ωφελείται από την πράξη, αν δηλαδή αυτό που ζητά ο ομιλητής είναι επ' ωφελεία δική του ή του δράστη της ενέργειας και β) κατά πόσο αυτό που ζητείται είναι προαιρετικό ή κατά κάποιον τρόπο υποχρεωτικό. Αυτό το κριτήριο προφανώς σχετίζεται και με τη θέση εξουσίας που μπορεί να κατέχει ο ομιλητής απέναντι στον ομιλητή.

Όσον αφορά στη σύνταξη, σύμφωνα με τον Searle (Searle, 1976:17), το γενικό συντακτικό σχήμα των ΚΛΠ στη βαθεία προτασιακή δομή είναι το εξής:

Εγώ ρήμα εσύ + εσύ Μελ. Ρήμα (ΟΦ) (Επίρ.)

Έτσι, π.χ. η πρόταση “*I order you to leave*” είναι η πραγμάτωση του παραπάνω σχήματος στην επιφανειακή δομή. Με βάση τη δομή αυτή πραγματώνονται κυρίως οι άμεσες κατευθυντικές πράξεις. Το βασικότερο χαρακτηριστικό ωστόσο είναι ότι στο εκφώνημα που περιέχει το αντικείμενο της κατευθυντικής πράξης, ο δράστης, το πρόσωπο το οποίο υλοποιεί την πράξη, μπορεί να είναι ο ακροατής ή/ και κάποιο τρίτο άτομο, διαφέρει πάντως από τον ομιλητή.

1.2. Σχετική έρευνα

Η μεγάλη εξάπλωση του διαδικτύου έχει αναγάγει την εξαγωγή άποψης σε κεντρικό και ουσιώδη τομέα έρευνας στα πλαίσια της επεξεργασίας φυσικής γλώσσας. Έτσι, οι Dave et al. (2003) και οι Hu & Liu (2004) ενδιαφέρονται για την εξαγωγή γνώμης μέσα από κριτικές προϊόντων που κάνουν οι χρήστες του διαδικτύου. Και οι δύο εστιάζουν αποκλειστικά σε product reviews, ενώ η δεύτερη επιχειρεί και τη δημιουργία περίληψης (summarization). Επίσης στα πλαίσια της ανάλυσης συναισθήματος, οι Pang & Lee (2004) αξιοποιούν το στοιχείο της υποκειμενικότητας προκειμένου να δημιουργήσουν ένα εργαλείο που θα κατατάσσει αυτόματα κριτικές ταινιών με βάση την πολικότητά τους, δηλαδή σε θετικές ή αρνητικές.

Μεθοδολογικά, η προσέγγιση που ακολουθείται εδώ για τη δημιουργία του εργαλείου στηρίχθηκε σε λεξικά, όπως και των Ding et al. (2008), με τη διαφορά ότι το λεξικό που παρουσιάζεται εκεί είναι γενικότερα για την εξαγωγή γνώμης και όχι συγκεκριμένα για τις ΚΛΠ. Επίσης, παρόμοια μεθοδολογία, δηλαδή δημιουργία λεξικών και κανόνων, ακολουθείται από τους Ramanand et al. (2010) και Pontiki & Papageorgiou (2014), με τη διαφορά ότι ασχολούνται με διαφορετικό σημασιολογικό πεδίο. Οι πρώτοι εστιάζουν στις προτάσεις για βελτίωση και τις ευχές μέσα σε κριτικές προϊόντων, ενώ οι δεύτεροι στην ανάλυση της πρόθεσης (Intention Analysis).

1.2.1. Λεκτικές πράξεις

Με την ανάλυση των λεκτικών πράξεων εν γένει ασχολούνται οι Qadir & Riloff (2011). Η συγκεκριμένη έρευνα επιχειρεί την ταξινόμηση προτάσεων από δημοσιεύσεις σε πίνακες μηνυμάτων (message board posts) στις κατηγορίες των λεκτικών πράξεων που όρισε ο Searle (1976). Τα λεξικά μέσα που χρησιμοποιούν είναι εν πολλοίς κοινά με αυτά που συμπεριλαμβάνονται και στα λεξικά για τις ΚΛΠ που περιγράφηκαν εδώ. Ωστόσο, η μεθοδολογία βασίζεται μόνο σε μεθόδους μηχανικής μάθησης. Παρόμοια δουλειά γίνεται και από τους Cohen et al. (2004) οι οποίοι επίσης χρησιμοποιούν μηχανική μάθηση προκειμένου να κατατάξουν email σε λεκτικές πράξεις, ορίζουν όμως μια δική τους οντολογία ως προς αυτές, δεν ακολουθούν δηλαδή την ταξινόμηση του Searle. Την κατηγοριοποίηση αυτή υιοθετούν και οι Applling et al. (2013) οι οποίοι κάνουν μια απόπειρα να εντοπίσουν λεκτικές πράξεις σε κείμενα από κοινωνικά δίκτυα χρησιμοποιώντας λεξικά και εν μέρει συντακτικά μέσα, στοχεύουν όμως κυρίως στην

αυτοματοποιημένη αναγνώριση χαρακτηριστικών προσωπικότητας. Τη θεωρία του Searle αξιοποιούν και οι Carlos & Yalamanchi (2012), χωρίς όμως να υιοθετούν τη διάκρισή του, αφού η ταξινόμηση που παρουσιάζουν προσανατολίζεται στην ανάλυση της πρόθεσης των ομιλητών.

Γενικότερα, τα ηλεκτρονικά μηνύματα απασχολούν αρκετά την έρευνα, αφού εκτός από τους Cohen et al. (2004) ασχολούνται με αυτά και οι Lampert et al. (2008a, 2008b, 2010). Οι τελευταίοι, ειδικότερα, εστιάζουν τις μελέτες τους στα αιτήματα (requests) και τις δεσμεύσεις (commitments). Αξιοποιώντας τη θεωρία των λεκτικών πράξεων, καταγράφουν τα φαινόμενα που μπορούν να δυσκολέψουν τον εντοπισμό των αιτημάτων και των δεσμεύσεων και προτείνουν λύσεις για αυτά, χρησιμοποιώντας κυρίως τεχνικές μηχανικής μάθησης.

1.2.2. Συστάσεις και αιτήματα

Το ζήτημα των συστάσεων (recommendations) έχει αποτελέσει επίσης αντικείμενο έρευνας. Τόσο οι Terveen et al. (1997) όσο και οι Balabanović & Shoham (1997) παρουσιάζουν συστήματα τα οποία χρησιμοποιούν διάφορες τεχνικές προκειμένου να προτείνουν σε χρήστες ιστοσελίδες με βάση τις προτιμήσεις τους. Με συστήματα συστάσεων (recommender systems) ασχολούνται και οι Mooney et al. (1998) οι οποίοι προτείνουν ένα εργαλείο το οποίο συστήνει σε χρήστες βιβλία. Βασική διαφοροποίηση της παρούσας εργασίας είναι ότι στηρίζεται σε υλικό που παράγουν οι ίδιοι οι χρήστες (user generated content) και επιδιώκει σε αυτό να εντοπίσει συστάσεις και αιτήματα. Αντίθετα, στις εργασίες που προαναφέρθηκαν ο στόχος είναι να δημιουργηθούν συστάσεις για τους χρήστες.

Επίσης, οι Natt och Dag et al. (2005) ακολουθούν μια προσέγγιση στηριγμένη τόσο σε γλωσσολογικά μέσα όσο και σε μηχανική μάθηση προκειμένου να εντοπίζουν επιθυμίες και ανάγκες χρηστών για προϊόντα. Η μελέτη είναι προσανατολισμένη σε εμπορική χρήση του εργαλείου από εταιρείες που θέλουν να ανακαλύψουν ευκαιρίες για πωλήσεις. Και οι Carlos & Yalamanchi (2012) εστιάζουν στην αξιοποίηση ειδικότερα της ανάλυσης πρόθεσης (intention analysis) για εμπορικούς σκοπούς. Στις κατηγορίες πρόθεσης που εξετάζουν συμπεριλαμβάνουν την πρόθεση των χρηστών να κατευθύνουν (direct) και να ρωτήσουν (inquire), εμπλέκοντας έμμεσα και τις συστάσεις και τα αιτήματα.

Τέλος, και οι Lampert et al. (2008a, 2008b, 2010) στις εργασίες τους ασχολούνται με τα αιτήματα (requests), περιορίζονται ωστόσο στην ανάλυση δεδομένων από ηλεκτρονικά μηνύματα και στοχεύουν στον εντοπισμό σε αυτά αιτημάτων για δράση.

2. Γλωσσικά δεδομένα

2.1. Σώματα κειμένων

Στα πλαίσια της παρούσας εργασίας, χρησιμοποιήθηκαν τρία σώματα κειμένων, καθένα από τα οποία καλύπτει διαφορετική θεματική περιοχή. Πιο συγκεκριμένα:

- **Boston:** σχετίζεται με τον μαραθώνιο που διεξήχθη στη Βοστώνη την άνοιξη του 2013 καθώς και με το δραματικό γεγονός των εκρήξεων που σημειώθηκαν εκεί. Αποτελείται από 776 προτάσεις και προέρχεται από το Twitter.
- **Product reviews:** περιλαμβάνει αξιολογήσεις και σχόλια για διάφορα προϊόντα. Περιέχει 1235 προτάσεις και προέρχεται από αξιολογήσεις καταναλωτών για διάφορα προϊόντα από τις ιστοσελίδες amazon.com και cnet.com.
- **Politics:** είναι σχετικό με την πολιτική και περιέχει κυρίως απόψεις για διάφορες παρατάξεις, αλλά και κριτική πολιτικών αποφάσεων και στρατηγικών κυβερνήσεων και μεμονωμένων πολιτικών προσώπων. Προέρχεται από δημοσιεύσεις αναγνωστών στην ιστοσελίδα politics.com. Το σώμα κειμένων ήταν αρκετά μεγάλο (6379 προτάσεις) και εξήχθη ένα μέρος από αυτό με βάση κάποιες λέξεις-κλειδιά που χρησιμοποιήθηκαν. Αυτές ήταν επίθετα που εκφράζουν κάποια άποψη ή συναίσθημα, ενώ συμπεριελήφθησαν και τα αντίστοιχα επιρρήματα. Το σώμα που τελικά προέκυψε συνίσταται από 1752 προτάσεις.

Οι δύο τελευταίες συλλογές κειμένων προέρχονται από το **WISH corpus** (Goldberg et al., 2009). Πρόκειται για ένα σώμα κειμένων το οποίο χρησιμοποιήθηκε προκειμένου να δοκιμαστεί σε διαφορετικά θεματικά πεδία ένα εργαλείο για τον εντοπισμό των ευχών.

Ως προς τα βασικά χαρακτηριστικά των συλλογών, και οι τρεις αποτελούνται από κείμενα και όχι από μεμονωμένες προτάσεις. Στο πρώτο, επειδή στο Twitter υπάρχει περιορισμός στον αριθμό των χαρακτήρων που μπορεί να δημοσιεύσει ο χρήστης, είναι και τα κείμενα σχετικά μικρά σε έκταση, πολλά ωστόσο από αυτά αποτελούνται από περισσότερες της μιας προτάσεις. Στα άλλα δύο δεν υπάρχει κανένας περιορισμός ως προς το μέγεθος και έτσι ποικίλλουν από μερικές λέξεις ως αρκετές προτάσεις. Στην περιγραφή που ακολουθεί, οι όροι κείμενο και πρόταση

χρησιμοποιούνται χωρίς διαφοροποίηση προκειμένου να δηλώσουν κάθε ξεχωριστή δημοσίευση ή αξιολόγηση. Επιπλέον, το πρώτο σώμα δεν έχει υποστεί καμιά επεξεργασία, ενώ στα άλλα δύο έχει γίνει μια προ-επεξεργασία κατά την οποία τα κεφαλαία γράμματα έχουν μετατραπεί σε μικρά και έχει αφαιρεθεί η στίξη.

Η αρχική επισημείωση έγινε σε επίπεδο κειμένου και όχι πρότασης. Κάθε αξιολόγηση (review) ή δημοσίευση (post) χαρακτηρίστηκε ως σύνολο ως προς συγκεκριμένες κατηγορίες που περιγράφονται αμέσως παρακάτω. Επίσης, ο τρόπος με τον οποίο έγινε η επισημείωση είναι δυαδικός. Σημειώνεται δηλαδή η τιμή ένα (1) για τις περιπτώσεις που υπάρχει το χαρακτηριστικό ως προς το οποίο εξετάζεται το κείμενο και η τιμή μηδέν (0) όταν δεν υπάρχει.

Τα δεδομένα εξετάστηκαν σε τέσσερις βασικούς άξονες, ο καθένας από τους οποίους διακρίνεται περαιτέρω σε ένα δίπολο. Εδώ παρουσιάζονται ξεχωριστά, καθαρά ως κλάσεις στα πλαίσια της επισημείωσης. Παρακάτω παρουσιάζονται αναλυτικότερα οι τέσσερις άξονες με τις επιπλέον διακρίσεις τους.

- Ο πρώτος άξονας είναι αυτός του **SUBJECTIVITY**, με άλλα λόγια της υποκειμενικότητας ή αντικειμενικότητας του συγγραφέα. Διακρίνεται οπότε περαιτέρω σε:
 - **Subjectivity**, κατηγορία που παίρνει τιμή 1 εφόσον ο συγγραφέας εκφράζει κάποια προσωπική του άποψη και τιμή 0 εάν πρόκειται για παράθεση αντικειμενικών γεγονότων.
 - **Objectivity**, επίσης με τιμές 0 ή 1, αντίστροφα με την προηγούμενη.
- Ο δεύτερος άξονας είναι το **POLARITY** ή πολικότητα, αν δηλαδή κάθε πρόταση φανερώνει θετική ή αρνητική στάση του συγγραφέα ή βγάζει γενικότερα κάποιο θετικό ή αρνητικό συναίσθημα. Συνεπώς οι δύο υποκατηγορίες είναι:
 - **Positive**, με τιμή 1 για θετική στάση ή συναίσθημα και 0 για αρνητική και
 - **Negative**, με τιμή 1 για αρνητική στάση ή συναίσθημα και 0 για θετική.
- Ο τρίτος άξονας είναι αυτός του **ATTITUDE**, της στάσης του συγγραφέα, της κρίσης του απέναντι σε πρόσωπα (judgment) ή πράγματα (appreciation). Αυτός διακρίνεται επιπλέον σε:
 - **Praise**, με τιμή 1 για θετική κρίση και 0 για αρνητική και
 - **Criticism**, με τιμή 1 για αρνητική στάση και 0 για θετική.

- Ο τελευταίος άξονας επισημείωσης είναι αυτός του **DIRECT**, κατά πόσο δηλαδή ο συγγραφέας ζητά από κάποιον να κάνει κάτι. Οι δύο υποκατηγορίες είναι:
 - **Direct**, που παίρνει τιμή 1 όταν πρόκειται για παρακίνηση σε δράση και 0 σε κάθε άλλη περίπτωση και
 - **Negated**, που παίρνει τιμή 1 όταν ο συγγραφέας προσπαθεί να αποτρέψει κάποιον από το να κάνει κάτι και 0 σε κάθε άλλη περίπτωση.

Αφού επισημειώθηκαν και τα τρία σώματα κειμένων με βάση το παραπάνω σχήμα κατηγοριοποίησης, στη συνέχεια ενώθηκαν σε ένα μεγαλύτερο, ενιαίο corpus με 3759 προτάσεις οι οποίες με βάση τις κατηγορίες κατανέμονται όπως φαίνεται στον παρακάτω πίνακα.

		C01_Boston	C02_Pr_Rev	C03_Politics
SUBJECTIVITY	SUBJECTIVE	585	1231	1625
	OBJECTIVE	189	2	127
POLARITY	POSITIVE	335	640	514
	NEGATIVE	260	451	804
ATTITUDE	PRAISE	43	452	37
	CRITICISM	19	258	73
DIRECT	DIRECT	107	176	245
	NEGATED	6	9	49

Όπως προκύπτει από τον πίνακα, ΚΛΠ εμφανίζονται και στα τρία θεματικά πεδία που μελετήθηκαν. Αυτό αποτελεί ένδειξη ότι τα γλωσσικά μέσα με τα οποία εκφράζονται αυτού του είδους οι λεκτικές πράξεις είναι κοινά και κατά συνέπεια, το εργαλείο που δημιουργήθηκε μπορεί να δοκιμαστεί σε διαφορετικά domain. Επιπλέον, το ποσοστό εμφάνισης των ΚΛΠ κυμαίνεται γύρω στο 15-16%. Αυτό δείχνει ότι δεν είναι σπάνιες μέσα στον λόγο, και άρα αξίζει να διερευνηθούν.

Οι προτάσεις αρχικά επισημειώθηκαν ως προς το αν εκφράζουν υποκειμενικότητα ή αντικειμενικότητα. Όπως είναι αναμενόμενο, η πλειοψηφία των προτάσεων εκφράζουν την προσωπική άποψη ή συναίσθημα του συγγραφέα τους. Το μικρό ποσοστό αντικειμενικών προτάσεων είναι αυτές που μεταφέρουν την άποψη

κάποιου τρίτου ή πρόκειται για τίτλους ειδήσεων. Ουσιαστικά, με τις ΚΛΠ συνδέονται κατ' αποκλειστικότητα απόψεις προσώπων, και κατά συνέπεια μόνο εκείνες που είναι υποκειμενικές. Αυτό όμως συνδέεται άμεσα με το άτομο που εκφράζει την πράξη, αν δηλαδή ο ίδιος ο συγγραφέας παρακινεί σε δράση ή μεταφέρει την προτροπή κάποιου άλλου. Σε αυτή τη δεύτερη περίπτωση μπορεί να συνδυάζεται η αντικειμενικότητα με την κατευθυντικότητα, όπως στην πρόταση “*Venezuela opposition candidate demands recount after Chavez's heir win*”.

Ως προς την πολικότητα, είναι ουσιαστικά η κατηγορία που δηλώνει την άποψη και μπορεί να είναι θετική, αρνητική ή ουδέτερη. Κατά πρώτο λόγο, η ουδέτερη στάση δεν επισημειώνεται, υπονοείται όμως όταν και η θετική και η αρνητική πολικότητα παίρνουν την τιμή 0. Συνήθως, συναίσθημα εκφράζουν τα επίθετα και τα επιρρήματα, ενώ σε κάποιες περιπτώσεις και τα ρήματα μπορούν να δηλώνουν την πολικότητα μιας πρότασης. Καθώς η ανθρώπινη γλώσσα μπορεί να εκφράζει κάθε είδους τροπικότητα και κατεξοχήν εμπλέκεται η υποκειμενικότητα του ομιλητή – ή συγγραφέα – υπήρξαν και κάποιες αμφισημίες. Έπειτα, αν και η κατηγορία της πολικότητας δεν αποτελεί στόχο της παρούσας μελέτης, συνήθως, όταν ο ομιλητής παρακινεί τον ακροατή να κάνει κάτι, εκφράζει και κάτι θετικό ως προς αυτό, ενώ στην αντίθετη περίπτωση η άποψη που αποτυπώνεται είναι αρνητική. Για παράδειγμα στις προτάσεις “*i highly recommend this camera to anyone looking for a good digital camera that takes great pictures yet does n t take weeks to figure out how to operate*” και “*whatever you do do n t buy this product*”. Ωστόσο, αυτό ισχύει κυρίως όταν το αντικείμενο της σύστασης είναι κάτι συγκεκριμένο, ένα προϊόν, ένα πρόσωπο και όχι σε άλλες περιπτώσεις, π.χ. “*if you get nis be sure you want all of them installed and running on your machine*”.

Τέλος, ως προς το αν ο συγγραφέας εκφράζει κάποια θετική ή αρνητική στάση (praise ή criticism), όπως προκύπτει και από τον πίνακα, εξαρτάται από τη θεματική περιοχή. Στις αξιολογήσεις προϊόντων είναι αρκετά συχνές οι περιπτώσεις που δηλώνεται ρητά κάποια εκτίμηση, ενώ στα άλλα δύο πεδία είναι αισθητά λιγότερες. Και ως προς τη σχέση της κατηγορίας αυτής με την κατευθυντικότητα που μπορεί να εκφράζεται από μια πρόταση, φαίνεται να ισχύει το ίδιο. Δηλαδή, όταν το αντικείμενο της σύστασης ή αποτροπής είναι ένα προϊόν, υπάρχει αντίστοιχα και θετική ή αρνητική στάση. Π.χ. “*even with that i highly recommend this router outstanding performer*” και “*you re going to need to buy new headphones the stock ones suck*”. Όπως παρατηρείται από τα αποτελέσματα στον πίνακα παραπάνω, ότι στις αξιολογήσεις προϊόντων οι ομιλητές φαίνεται περισσότερο να εκφράζουν κάτι θετικό, αφού τόσο οι προτάσεις με

θετική πολικότητα όσο και εκείνες που περιέχουν θετική κριτική για κάτι είναι περισσότερες από τις αρνητικές. Το ακριβώς αντίθετο φαίνεται να ισχύει στις απόψεις περί πολιτικής, όπου η αρνητική στάση είναι αυτή που επικρατεί. Και στα tweets δείχνουν να υπερισχύουν οι θετικές απόψεις, αλλά οι διαφορές δεν είναι ιδιαίτερα μεγάλες και θα χρειαζόταν ίσως ένα μεγαλύτερο σώμα κειμένων προκειμένου να εξαχθεί ασφαλές συμπέρασμα.

Αφού ολοκληρώθηκε η επισημείωση, αυτό το corpus χωρίστηκε σε δύο μικρότερα με βάση την τιμή της κατηγορίας DIRECT. Έτσι, το ένα περιλαμβάνει όσες προτάσεις από τα επισημειωμένα σώματα έχουν τιμή 1 στο συγκεκριμένο πεδίο (581 προτάσεις), εκφράζουν επομένως κάποια ΚΛΠ, ενώ το άλλο όσες προτάσεις έχουν τιμή 0 (3178 προτάσεις). Επίσης, αποφασίστηκε να μην συμπεριληφθούν στην ανάλυση ΚΛΠ που το πρόσωπο που τις εκφράζει δεν είναι ο ίδιος ο ομιλητής και γι' αυτό από το πρώτο σώμα αφαιρέθηκαν οι προτάσεις των οποίων το υποκείμενο είναι κάποιος τρίτος.

Στη συνέχεια, για λόγους πληρότητας, το σώμα κειμένων επεκτάθηκε και με παραδείγματα από λεκτικά μέσα έκφρασης ΚΛΠ τα οποία δεν εμφανίστηκαν. Προκειμένου να γίνει αυτό, χρησιμοποιήθηκε κυρίως το FrameNet⁶, όπου, για κάθε λεξική μονάδα έκφρασης ΚΛΠ που εντοπίστηκε μέσα στα υπάρχοντα κείμενα, βρέθηκαν τα σημασιολογικά πεδία στα οποία εντάσσεται και αντίστοιχες λεξικές μονάδες που μοιράζονται τα ίδια σημασιολογικά χαρακτηριστικά. Ακόμα, χρησιμοποιήθηκαν και λίστες συνωνύμων από τις ηλεκτρονικές εκδοχές των λεξικών Oxford⁷ και Cambridge⁸. Έπειτα, με βάση αυτά τα γλωσσικά δεδομένα, αναζητήθηκαν παραδείγματα χρήσης μέσα από συλλογές σωμάτων κειμένων, και συγκεκριμένα το Corpus of Contemporary American English (COCA)⁹, το British National Corpus (BNC)¹⁰ και το Sketch Engine¹¹. Τα παραδείγματα αυτά χρησιμοποιήθηκαν για να εμπλουτίσουν το σώμα κειμένων στο οποίο στηρίχθηκε η ανάπτυξη του εργαλείου για τον εντοπισμό των ΚΛΠ. Το τελικό σώμα κειμένων που προέκυψε με αυτόν τον τρόπο αποτελείται από 1063 προτάσεις.

⁶ <https://framenet.icsi.berkeley.edu/fndrupal/>

⁷ <http://www.oxfordlearnersdictionaries.com/>

⁸ <http://dictionary.cambridge.org/>

⁹ <http://corpus.byu.edu/coca/>

¹⁰ <http://www.natcorp.ox.ac.uk/>

¹¹ <https://the.sketchengine.co.uk/>

Οι προτάσεις που προστέθηκαν σε αυτό το corpus, επισημειώθηκαν με βάση το σχήμα που εφαρμόστηκε στα αρχικά σώματα κειμένων. Τα αποτελέσματα που προέκυψαν φαίνονται στον ακόλουθο πίνακα.

SUBJECTIVITY	SUBJECTIVE	1037
	OBJECTIVE	27
POLARITY	POSITIVE	269
	NEGATIVE	339
ATTITUDE	PRAISE	70
	CRITISISM	49

Όπως προκύπτει από τον παραπάνω πίνακα, η πλειοψηφία των προτάσεων που εκφράζουν ΚΛΠ, είναι υποκειμενικές δηλώσεις. Σημαντικό είναι επίσης ότι δεν εκφράζουν πάντα θετική ή αρνητική πολικότητα. Τέλος, ως προς τη στάση του ομιλητή, δε φαίνεται να είναι πολλές οι προτάσεις στις οποίες εκφράζεται θετική ή αρνητική κριτική. Αυτό βέβαια μπορεί να σχετίζεται με το θεματικό πεδίο, οπότε το συγκεκριμένο θέμα θα μπορούσε να αποτελέσει αντικείμενο ιδιαίτερης έρευνας σε εκτενέστερα σώματα κειμένων.

2.1.1. Γλωσσικά χαρακτηριστικά

Με την ευρεία χρήση του διαδικτύου από όλους, χωρίς περιορισμούς ηλικίας, φύλου, εθνικότητας, κοινωνικής ή οικονομικής κατάστασης – τουλάχιστον για τις περισσότερες αναπτυγμένες κοινωνίες – προέκυψε η ανάγκη για επικοινωνία μεταξύ ετερόκλητων ομάδων ανθρώπων. Έτσι, δημιουργήθηκε μια «διάλεκτος» η οποία διαφοροποιείται ανάλογα με τις ανάγκες των χρηστών. Εδώ καταγράφονται κάποια χαρακτηριστικά της γλώσσας που χρησιμοποιείται στο διαδίκτυο, με βάση τις παρατηρήσεις που προέκυψαν από τα σώματα κειμένων που εξετάστηκαν.

Το βασικότερο χαρακτηριστικό του λόγου που χρησιμοποιείται στα κοινωνικά δίκτυα είναι ότι, παρόλο που τυπικά είναι γραπτός, το ύφος του κινείται ανάμεσα σε γραπτό και προφορικό, προσιδιάζει όμως περισσότερο στον προφορικό, καθημερινό και κυρίως νεανικό τρόπο ομιλίας. Οι χρήστες φαίνεται να αισθάνονται εξαιρετικά

οικεία και έτσι γράφουν σε φιλικό τόνο, όπως θα μιλούσαν σε μια παρέα με φίλους. Γι' αυτό και δε χρησιμοποιούν τα τυπικά χαρακτηριστικά του γραπτού λόγου.

Αρχικά, σε επίπεδο λεξιλογίου είναι πολύ αντιπροσωπευτική η χρήση ιδιότυπων εκφράσεων για χαιρετισμούς και προσφωνήσεις, επιδοκιμασία ή αποδοκιμασία, η βωμολοχία, χρησιμοποίηση δηλαδή υβριστικών λέξεων και φράσεων ακόμα και χωρίς διάθεση προσβολής ή κριτικής. Επίσης, εντοπίζονται διάφοροι νεολογισμοί, καινούριες χρήσεις υπαρχόντων γλωσσικών στοιχείων ή και δημιουργία καινούριων. Μερικά παραδείγματα:

*“#PrayforBoston so horrible ?? what a **fucked up** thing to do..”*

*“A3: **dam run and oohing and aahing** at the event #runchat #bostonmarathon @RunningBecause”*

*“sleep well you win you cold **bastard**”.*

Σε σημασιολογικό επίπεδο, το πιο χαρακτηριστικό γνώρισμα αυτού του είδους λόγου είναι οι συντετμημένες μορφές λέξεων ή φράσεων, τα ακρόνυμα, καθώς και η χρήση διαφόρων συμβόλων τα οποία κατά κύριο λόγο χρησιμοποιούνται για να εκφράσουν κάποιο συναίσθημα, τα γνωστά ως emoticons. Αυτοί οι τρόποι επικοινωνίας χρησιμοποιήθηκαν αρχικά για λόγους οικονομίας χρόνου αλλά κυρίως χώρου, καθώς στο Twitter υπάρχει περιορισμός στους χαρακτήρες που μπορεί ο χρήστης να πληκτρολογήσει σε κάθε μήνυμα. Επιπλέον, πολλά από τα ευρέως χρησιμοποιούμενα ακρόνυμα δημιουργήθηκαν από νεαρούς χρήστες του διαδικτύου προκειμένου να «κωδικοποιούν» την επικοινωνία τους και να μην γίνεται κατανοητή από τους γονείς. Όλα αυτά αποτελούν ένα ευφάνταστο, ιδιότυπο λεξικό, το οποίο συνεχώς εμπλουτίζεται με νέα στοιχεία, ιδίως εικόνες αλλά και σκίτσάκια με τη χρήση γραμμάτων, αριθμών και σημείων στίξεως. Τέτοια σύμβολα εμφανίζονται μόνο στο πρώτο σώμα κειμένων που χρησιμοποιήθηκε από την παρούσα εργασία, το οποίο προέρχεται από το Twitter, αφού από τα άλλα δύο έχει αφαιρεθεί κάθε στίξη. Παραδείγματα μέσα από τα κείμενα:

*“**lol** that pawlenty guy must be smokin pawlenty of some good weed smiley o ordm deg acute uml middot l macr middot cedil middot acute macr middot shy gt happy”*

*“Surgery isn't going well. If you pray, please say one for my **MIL**. #prayers #cancer”*

*“@LogicallyLoud right back at you #MIT **:D :D.** thanks thala :)”*

“@spamtan @trishuson @noidauson I hope you find her :(#pray”

Ένα άλλο χαρακτηριστικό είναι τα ορθογραφικά λάθη και οι αναγραμματισμοί λέξεων που πολλές φορές οφείλονται στην πληκτρολόγηση και καθιστούν δύσκολη την ανάγνωση και κατανόηση του κειμένου, πόσο μάλλον την αυτόματη αναγνώρισή του από εργαλεία. Τέτοια παραδείγματα υπάρχουν πολλά στα κείμενα:

*“why would you want to do that to someone its **imhumane** sick and barbaric to force this woman to stay alive for 50 or 60 years in this state hey fuck you why you got ta resort to calling me a moron you dumb hillbilly you shouldnt be forcing anyone to live anyway its not your damn problem to be forcing people to live”*

*“i would **recogmend** this player to anyone i know and you will not be disappointed”*

*“ #MSM #CBS #PBS #CNN #ABC grow a set and everytime your station talks about #Newtown **folllw** it up with a story on the #Gosnell trial.”*

Τέλος, και σε συντακτικό επίπεδο η γλώσσα που χρησιμοποιείται στα μέσα κοινωνικής δικτύωσης έχει γνωρίσματα του προφορικού λόγου, καθώς εμφανίζονται ασυνταξίες, επαναλήψεις, παραλείψεις, μακροπερίοδος λόγος ή, αντίθετα, εξαιρετικά μικρές προτάσεις. Ειδικότερα, στα κείμενα που εξετάστηκαν αποτελεί συχνό φαινόμενο η παράλειψη του αντωνυμικού υποκειμένου, παρόλο που η αγγλική ανήκει στις γλώσσες στις οποίες είναι υποχρεωτική η δήλωσή του. Σε συνδυασμό με τη φτωχή μορφολογία της, η έλλειψη του υποκειμένου μπορεί να δημιουργήσει προβλήματα. Μερικά παραδείγματα τέτοιου είδους:

“(you) better find a damn good reason to let him go and his book free”

“Awake. (I) Can't sleep. (I am) Waiting for #BostonMarathon to start.”

Κλείνοντας, υπάρχουν και κάποια στοιχεία στον λόγο τα οποία δυσκολεύουν πολύ το έργο της επισημείωσης, τα οποία δεν σχετίζονται αποκλειστικά με το συγκεκριμένο είδος κειμένων αλλά είναι γενικότερα χαρακτηριστικά τόσο του γραπτού όσο και του προφορικού λόγου. Καταρχάς, σε σημασιολογικό επίπεδο υπάρχουν περιπτώσεις στις οποίες η δυσκολία προκύπτει από την εγγενή αμφισημία της γλώσσας, σε εκφράσεις δηλαδή που μπορούν να είναι και θετικές και αρνητικές και δεν μπορούμε να είμαστε σίγουροι για το τι πιστεύει ο συγγραφέας. Π.χ.

*“@FoxNews: #BREAKING: Hugo Chavez's chosen heir Nicolas #Maduro wins Venezuelan presidency in tight race”. **SURPRISE!**”*

*“A3: 6am run and **oohing** and **aahing** at the event #runchat #bostonmarathon @RunningBecause”*

“oh my goodness”

Σε συντακτικό επίπεδο απ' την άλλη, οι υποθετικές προτάσεις¹² έχουν αποτελέσει ιδιαίτερο πεδίο στη γλωσσική έρευνα αφού μπορούν να εμπλέκουν θετικές και αρνητικές απόψεις, χωρίς να είναι πάντα εύκολο να ξεχωρίσουν. Π.χ.

*“Zero Dark Thirty is a really **good** movie **if you're into killing** Osama Bin Laden #America”*

Σε πραγματολογικό επίπεδο, τέλος, η ειρωνεία είναι ένα φαινόμενο που μπορεί να δημιουργήσει αρκετά προβλήματα, αφού είναι δυνατό να αλλάξει εντελώς το νόημα μιας πρότασης, ενώ δεν είναι πάντα εύκολο να εντοπιστεί. Π.χ.

*“well actually yeah you are right i am crying cause i ca n t stop myself from **laughing at your retarded ass**”.*

2.1.2. Γλωσσικά μέσα έκφρασης κατευθυντικών πράξεων

Υπάρχουν διάφοροι τρόποι με τους οποίους εκφράζονται οι ΚΛΠ στην αγγλική γλώσσα, είτε πρόκειται για άμεσες, ρητές πράξεις είτε για έμμεσες. Το πιο χαρακτηριστικό λεκτικό μέσο είναι τα λεξιλογικά στοιχεία που ρητά εκφράζουν κάποια ΚΛΠ. Αυτά μπορεί να είναι ρήματα, όπως τα *“demand”, “recommend”, “request”, “advise”*. Υπάρχουν επίσης αντίστοιχα ουσιαστικά, τα οποία εντάσσονται σε μια ευρύτερη φράση, όπως *“my advice/recommendation/suggestion is”*.

Από γραμματικά μέσα, το πιο ξεκάθαρο είναι η προστακτική, η οποία διαγλωσσικά χρησιμοποιείται προκειμένου να εκφράσει όλες εκείνες τις επικοινωνιακές λειτουργίες οι οποίες εντάσσονται στις ΚΛΠ, όπως η προσταγή, η προτροπή, η παράκληση, η απαγόρευση, η ευχή. Οι διαφοροποιήσεις που εμφανίζονται ανάμεσα στις γλώσσες εντοπίζονται στη μορφή, τη λεκτική τους πραγμάτωση.

Στην αγγλική γλώσσα, ωστόσο, δεν υπάρχουν ξεχωριστοί μορφολογικά τύποι της προστακτικής. Γενικότερα η αγγλική δεν είναι μια πλούσια μορφολογικά γλώσσα, γεγονός που ισχύει και για την προστακτική, για την έκφραση της οποίας χρησιμοποιείται ο βασικός τύπος του ρήματος, αυτό που αποκαλείται απαρέμφατο. Π.χ. *“Come here!”*, *“Go out!”*. Το γεγονός αυτό δυσκολεύει τον αυτόματο εντοπισμό των συγκεκριμένων γλωσσικών τύπων.

Η ερώτηση ως προτασιακός τύπος εντάσσεται επίσης στις ΚΛΠ, με τη λογική ότι ο ομιλητής περιμένει από τον ακροατή κάποια απόκριση έστω κι αν αυτή είναι

¹² Βλ. για παράδειγμα Jindal & Liu (2006), Ganapathibhotla (2008), Narayanan (2009).

απλώς μια λέξη, μια απάντηση. Π.χ. “*What should I do about situation?*”. Βέβαια, μερικές ερωτήσεις μπορεί να είναι ρητορικές και να μην επιτελούν ΚΛΠ, όπως π.χ. “*Can you believe that?*”. Από τους πιο αντιπροσωπευτικούς τύπους ερωτήσεων που δηλώνουν κατευθυντικότητα είναι οι ερωτηματικές δομές σε μέλλοντα, του τύπου δηλαδή

“*Will/Would you+VB*”

“*Are you going to+VB*”

καθώς και οι αντίστοιχες αρνητικές μορφές τους,

“*Won't you+VB*”

“*Aren't you going to+VB*”

Εξίσου χαρακτηριστική είναι και η ερωτηματική δομή του τύπου

“*Can/Could you+VB*”

όπως η περίπτωση “*Can you pass the salt?*” που θεωρούνται σε γενικές γραμμές ευγενικά αιτήματα. Ωστόσο, το ρήμα *can*, εκφράζει τόσο την ικανότητα όσο και τη δυνατότητα, οπότε, ειδικά σε ορισμένα περιβάλλοντα, οι ερωτήσεις αυτού του τύπου μπορούν να αφορούν στην ικανότητα του ατόμου να κάνει κάτι και να μην δηλώνουν κατευθυντικότητα. Π.χ. η ερώτηση “*Can you reach my glasses?*” όταν απευθύνεται από έναν ορθοπεδικό σε έναν ασθενή είναι πιθανό να σχετίζεται με τη φυσική κατάσταση του δεύτερου και να μην είναι ευγενικός τρόπος έκφρασης αιτήματος. Χρειάζεται επομένως να γίνει έλεγχος του ρήματος που ακολουθεί και να περιοριστεί το σχήμα σε συγκεκριμένες μόνο κατηγορίες ρημάτων, ή έστω να αποκλειστούν κάποια ρήματα με βάση τη σημασία τους.

Κατευθυντικές λεκτικές πράξεις σηματοδοτούν επιπλέον κάποιες στερεότυπες ερωτηματικές δομές, οι οποίες χρησιμοποιούνται προκειμένου να μετριάσουν την ένταση της κατευθυντικότητας, πρωτίστως δηλαδή για λόγους που σχετίζονται με την ευγένεια. Τέτοιες δομές είναι οι ακόλουθες:

“*Why don't you+VB?*”

“*Why not+V?*”

“*How about+V?*”

“*What about+V?*”

“*Would you mind+V?*”

“*Would you like to+V?*”.

Ακόμα, στην αγγλική γλώσσα υπάρχει μια ιδιαίτερη κατηγορία ρημάτων, αυτά που ονομάζονται τροπικά (modal verbs). Είναι τα εξής: *can, could, may, might, will*,

would, must, shall, should, ought to. Αυτά χρησιμοποιούνται προκειμένου να εκφραστούν γλωσσικά κάποιες τροπικότητες, μεταξύ των οποίων και η κατευθυντικότητα. Η χρήση αυτή των τροπικών ρημάτων της αγγλικής είναι πολύ συχνή, όπως παρατηρείται και στα σώματα κειμένων που χρησιμοποιήθηκαν στην παρούσα εργασία, από τα οποία προέρχονται τα παραδείγματα.

*“you **can** buy a new battery for 49 99”*

*“by the way you **might** as well get rid of your fire and theft insurance too if that s the way you want to live your life”*

*“so **would** you please kindly go away?”*

*“that is wonderful harvy but please you **must** let us in on whether you agree or disagree otherwise where is the fun of it”*

*“however you **should** still buy an ipod”*

*“you **ought to** pick your battles better my friend it”*

Με την ίδια λογική και προκειμένου να εκφράσουν κάποια ΚΛΠ χρησιμοποιούνται και οι περιφράσεις *have to* και *need to*:

*“you just **have to** buy a new player”*

*“you **need to** test your logic more conclusively before you espouse if i may be so bold”.*

Στα λεκτικά μέσα με τα οποία εκφράζεται κάποια ΚΛΠ, μπορούν επίσης να ενταχθούν οι φράσεις *make sure* ή *be sure*, όπως φαίνεται μέσα από παραδείγματα όπως:

*“**make sure** you read the documentation clearly if you dont know what you re doing”*

*“btw **be sure** to have your mp3 tags labled correctly as this will ensure a great organization”*

Το επίρρημα *better* είναι άλλος ένας τρόπος που φανερώνει κατευθυντικότητα, μόνο όμως στις περιπτώσεις που προηγείται ρήματος, καθώς πρόκειται για συντομευμένη μορφή της περίφρασης *had better* η οποία επίσης εμφανίζεται. Π.χ.

*“well desh you **better** go and talk to your buddy mrs clinton then because as i see right now she is trying to get every republican on her side”.*

Όλα τα λεκτικά μέσα έκφρασης της κατευθυντικότητας που αναφέρθηκαν παραπάνω, χρησιμοποιούνται κατά κύριο λόγο με προσωπικές αντωνυμίες, εκτός του πρώτου ενικού, αλλά και με άλλα είδη αντωνυμιών όπως οι αόριστες, οι δεικτικές και οι αναφορικές. Καθώς όμως το τρίτο πρόσωπο της αντωνυμίας καθορίζει ελάχιστα το

πρόσωπο, πολλές φορές προσδιορίζεται από κάποια ονοματική φράση που λειτουργεί ως υποκείμενο ώστε να προσδιοριστεί ο δράστης της ΚΛΠ. Για παράδειγμα:

“all the young right wing extremists should just go to bob jones university that way they cant complain about the liberal bias”.

Από τα γραμματικά και λεξιλογικά μέσα με τα οποία εκφράζονται οι ΚΛΠ, η προστακτική ως έγκλιση και τα γλωσσικά μέσα που φέρουν στη σημασία τους ρητά την κατευθυντικότητα, πραγματώνουν τις άμεσες κατευθυντικές πράξεις, ενώ τα υπόλοιπα είναι, λιγότερο ή περισσότερο έμμεσοι τρόποι δήλωσης. Υπάρχουν βέβαια και περιπτώσεις έμμεσων κατευθυντικών πράξεων που είναι εξαιρετικά δύσκολο να εντοπιστούν. Είναι προτάσεις τις οποίες ο ίδιος ο άνθρωπος, προκειμένου να τις αναγνωρίσει, χρειάζεται το περικείμενο αλλά και γνώσεις για τις επικοινωνιακές περιστάσεις. Για παράδειγμα το εκφώνημα *“You are standing on my foot”*, υπό ορισμένες συνθήκες θα μπορούσαν να είναι ΚΛΠ και ο ομιλητής να περιμένει από τον ακροατή να κάνει κάτι, να βάλει καφέ ή να μετακινήσει το πόδι του αντίστοιχα.

3. Μεθοδολογία - Υλοποίηση

3.1. Στόχος - Μεθοδολογία

Η παρούσα διπλωματική εργασία έχει ως κύριο στόχο την αυτόματη ανάλυση των ΚΛΠ μέσα σε κείμενα προερχόμενα από ηλεκτρονικές πηγές και τα οποία έχουν δημιουργηθεί από χρήστες (user-generated content). Για τον σκοπό αυτόν δημιουργήθηκε μια γραμματική που συνίσταται από ένα λεξικό, αποτελούμενο από λέξεις και φράσεις που δηλώνουν κατευθυντικότητα, το οποίο συνοδεύεται από μια σειρά κανόνων. Οι κανόνες αυτοί στοχεύουν στην εύρεση τόσο των κατευθυντικών εκφράσεων, όσο και του αντικειμένου αυτών. Με άλλα λόγια επιδιώκεται να εντοπιστεί η πράξη την οποία ο ομιλητής ή συγγραφέας¹³ ζητά από τον ακροατή να πραγματοποιήσει ή να μην πραγματοποιήσει, αλλά και ο τρόπος με τον οποίο το ζητάει.

Η μεθοδολογία που ακολουθήθηκε, στηρίχθηκε σε σώματα κειμένων, μέσα από τα οποία προέκυψε το γλωσσικό υλικό. Τα κείμενα που χρησιμοποιήθηκαν αντλήθηκαν κατά αποκλειστικότητα από ηλεκτρονικές πηγές. Σε πρώτη φάση, μετά την επισημείωση των συλλογών κειμένων, έγινε η επιλογή των σημασιολογικών κατηγοριών (insights) στις οποίες διακρίνονται οι ΚΛΠ. Η διάκριση που επιχειρείται εδώ είναι αρκετά γενικευτική. Ο λόγος γι' αυτό είναι ότι η παρούσα εργασία δεν εστιάζει σε ένα συγκεκριμένο γνωστικό πεδίο (domain). Αντιθέτως, επιδιώκεται το εργαλείο που δημιουργήθηκε να έχει αποτελέσματα, στον βαθμό που αυτό είναι εφικτό, σε διαφορετικά πεδία.

Στη συνέχεια, και αφού εντοπίστηκαν τα λεκτικά μέσα με τα οποία δηλώνονται ΚΛΠ, αυτά εντάχθηκαν σε δύο λεξικά. Το ένα αποτελείται από ρήματα και το δεύτερο από εκφράσεις. Ακολουθώντας, με βάση τα λεξικά και σε επίπεδο επιφανειακής δομής, καθορίστηκαν τα δομικά σχήματα (patterns), δηλαδή οι τρόποι σύνταξης. Τα σχήματα αυτά υλοποιήθηκαν σε συγκεκριμένους κανόνες και με τον τρόπο αυτόν ολοκληρώθηκε η γραμματική. Η υλοποίηση και η αξιολόγηση του εργαλείου έγιναν μέσα από την πλατφόρμα επεξεργασίας φυσικής γλώσσας GATE. Ακολουθεί αναλυτική περιγραφή κάθε φάσης της μεθοδολογίας.

¹³ Οι όροι ομιλητής και συγγραφέας χρησιμοποιούνται χωρίς διαφοροποίηση για να δηλώσουν τον εκφωνητή της πρότασης που χαρακτηρίζεται ως κατευθυντική.

3.2. Σημαιολογικές κατηγορίες (insights)

Στα πλαίσια της παρούσας εργασίας, προτείνεται η διάκριση των ΚΛΠ σε δύο ευρύτερες κατηγορίες, τα αιτήματα (requests) και τις συστάσεις (recommendations). Το κριτήριο για τη διάκριση σε αιτήματα και συστάσεις, είναι η κατεύθυνση της ωφέλειας από την επιζητούμενη δράση, αν δηλαδή αυτός που ωφελείται είναι ο ομιλητής ή ο ακροατής.

Η πρώτη κατηγορία, τα αιτήματα, περιλαμβάνει όλες εκείνες τις ΚΛΠ στις οποίες ο ομιλητής ζητά από το πρόσωπο στο οποίο απευθύνεται μια πράξη η οποία είναι επ' ωφελεία του ίδιου του ομιλητή. Με άλλα λόγια, ο συγγραφέας εκφράζει μια επιθυμία του, μια χάρη την οποία ζητά να πραγματοποιήσει ο αποδέκτης. Εδώ συμπεριλαμβάνονται τόσο οι παρακλήσεις όσο και οι ικεσίες, όταν δηλαδή κάποιος ζητά κάτι με ευγενικό ή ικετευτικό τρόπο, παρακαλεί ή ικετεύει τον αποδέκτη για κάτι. Π.χ.

*“Goodness. A bombing at a marathon? What is wrong with people? If you can, **please give blood today.** #redcross #bostonmarathon”*

*“**I beg you help** me, for only you are strong enough”*

*“Surgery isn't going well. If you pray, **please say one for my MIL.** #prayers #cancer”*

*“so **can someone please tell** me how we can have such vivid satellite pictures of shri lanka and not have even more and better pictures of every inch of iraq for at least every day for the year preceding the war”.*

Η δεύτερη κατηγορία είναι αυτή της σύστασης. Με αυτού του είδους τις ΚΛΠ ο συγγραφέας προτρέπει τον αποδέκτη να κάνει κάτι, του συνιστά έναν τρόπο δράσης. Πρέπει να σημειωθεί ότι ο όρος σύσταση (recommendation) χρησιμοποιείται συμβατικά προκειμένου να περιγράψει την κατηγορία των ΚΛΠ με τις οποίες ο ομιλητής κατευθύνει, καθοδηγεί τον ακροατή σε μια πράξη που δεν αποσκοπεί όμως στην ωφέλεια του ομιλητή, τουλάχιστον άμεσα. Εντός της κατηγορίας αυτής υπάρχουν περαιτέρω υποκατηγορίες, κυρίως με βάση τον βαθμό προαιρετικότητας της επιδιωκόμενης πράξης και την επιβολή της από τον ομιλητή. Ουσιαστικά, η σύσταση αποτελεί μόνο ένα υποσύνολο αυτής της κατηγορίας, γι' αυτό και η χρήση του όρου για όλη την κατηγορία είναι συμβατική. Έτσι, στα πλαίσια της επισημείωσης της παρούσας εργασίας, εκλαμβάνεται ως σύσταση τόσο η διαταγή, όσο και η συμβουλή, η προτροπή, η παραχώρηση. Μερικά παραδείγματα αυτής της κατηγορίας: “i

recommend lots of water and rest you ll feel better tomorrow”, “well desh you better go and talk to your buddy mrs clinton then because as i see right now she is trying to get every republican on her side”, “as i said this is the last digital camera i ll buy and if you are in the market one my advice is to buy the best thing you can afford and keep it forever”.

Εκτός από τις δύο παραπάνω σημασιολογικές κατηγορίες, εντοπίζεται και η άρνηση αυτών (negated), καθώς μεταβάλλει εντελώς το νόημα, μετατρέποντας την πράξη σε αποτροπή, δηλαδή ο ομιλητής ζητά από τον ακροατή να μην κάνει κάτι. Αυτό το χαρακτηριστικό (negated) παίρνει θετική τιμή (true) όταν τα ρήματα και οι εκφράσεις των δύο προηγούμενων κατηγοριών συνοδεύονται από άρνηση, η οποία αντιστρέφει το νόημα, όπως στα παραδείγματα:

*“ So, no, I would **not advise taking** a supplement unless your doctor specifically recommends it.”*

*“i am **not suggesting** that we would **give** our consent to all of these nominees”*

*“**All I want is** for them **not to pay** me any mind.”*

Υπάρχουν επίσης ρήματα που φέρουν στη σημασία τους την άρνηση, εκφράζουν δηλαδή αποτροπή ή απαγόρευση, όπως τα **prohibit, deter, discourage**. Π.χ.

*“Also i **prevent** you **from consuming** items like deserts”*

*“Micah and Yoel weren’t feeling well and I **prohibited** them **from running** today”*

*“I **discourage** friends **from calling** me in my studio;”*

Οι κατηγορίες που ορίζονται εδώ είναι ευρείες και περιλαμβάνουν διάφορες υποκατηγορίες. Μέσα από τα σώματα κειμένων που εξετάστηκαν, προέκυψαν και περιπτώσεις προτάσεων στις οποίες ο ομιλητής εκφράζει μια ευχή για κάτι που θέλει να γίνει. Αυτή η κατηγορία δεν περιελήφθη στην ανάλυση, διότι πιστεύεται πως πρέπει να αποτελέσει ξεχωριστό αντικείμενο μελέτης. Κι αυτό γιατί, μολονότι για τη δήλωση των ευχών χρησιμοποιούνται συχνά γλωσσικά μέσα κοινά με εκείνα των ΚΛΠ – όπως λ.χ. η προστακτική – θεωρείται πιο περιφερειακή περίπτωση, κάπου στα όρια μεταξύ κατευθυντικών και εκφραστικών λεκτικών πράξεων.

3.3. Λεξικά

Μέσα από τα σώματα κειμένων που εξετάστηκαν, προέκυψαν τα λεκτικά μέσα με τα οποία ρητά εκφράζονται οι ΚΛΠ. Στη συνέχεια, χρησιμοποιήθηκε το FrameNet καθώς

και τα ηλεκτρονικά λεξικά Oxford και Cambridge προκειμένου να εντοπιστούν σημασιολογικά συγγενείς όροι, κατά κύριο λόγο συνώνυμα και αντώνυμα. Όλα τα λεξικά μέσα που επελέγησαν, εντάχθηκαν σε δύο λεξικά.

Το πρώτο λεξικό (**DirectiveLex**) περιλαμβάνει μόνο τα ρήματα τα οποία πραγματώνουν κάποια ΚΛΠ. Πρόκειται τόσο για απλά ρήματα, όσο και για τροπικά – τα modals της αγγλικής γλώσσας – ενώ επίσης συμπεριλαμβάνονται ορισμένα από τα περιφραστικά ρήματα (phrasal verbs). Συνολικά οι εγγραφές αυτού του λεξικού είναι 100.

Η ταξινόμηση των ρημάτων γίνεται σε δύο επίπεδα, ένα σημασιολογικό και ένα συντακτικό. Κατ' αρχάς, με βάση τα σημασιολογικά τους χαρακτηριστικά, χαρακτηρίζονται ως προς την κατηγορία κατευθυντικότητας που δηλώνουν. Οι κατηγορίες αυτές, είναι δύο, το αίτημα (request) και η σύσταση (recommendation). Κάθε ρήμα παίρνει μία από τις δύο αυτές τιμές. Αυτές οι δύο ομάδες είναι αρκετά ξεκάθαρες και χωρίς αλληλεπικάλυψη. Μόνη εξαίρεση αποτελούν κάποια από τα τροπικά ρήματα, στα οποία ο διαφορετικός τρόπος σύνταξης επιφέρει και αλλαγή σημασίας. Για παράδειγμα το ρήμα *could* στην πρόταση “*you could make it unlimited for all i care*”, εκφράζει σύσταση, ενώ αντίθετα στην περίπτωση “*sr could you please give me a quote from bush s speech where he said we will go to war with any nation that does not have free peoples*” δηλώνει αίτημα. Το ζήτημα αυτό αντιμετωπίζεται σε επίπεδο κανόνων, ώστε στο αποτέλεσμα του εργαλείου να φαίνεται πάντα η σωστή κατηγορία. Να σημειωθεί επίσης ότι τα ρήματα τα οποία φέρουν στη σημασία τους την άρνηση, ζητούν δηλαδή από τον ακροατή να μην κάνει κάτι, φανερώνουν αποτροπή ή απαγόρευση, επισημειώνονται με την κατηγορία στην οποία ανήκουν (request, recommendation) συν το χαρακτηριστικό negated.

Σε συντακτικό επίπεδο τα λήμματα του λεξικού χωρίζονται με βάση τα συμπληρώματα που χρειάζονται, ανάλογα δηλαδή με τον τρόπο με τον οποίο συμπεριφέρονται συντακτικά. Ο χωρισμός αυτός σχετίζεται άμεσα με τα δομικά σχήματα τα οποία ορίστηκαν και τα οποία περιγράφονται αναλυτικά στην επόμενη ενότητα. Εδώ ωστόσο πρέπει να αναφερθεί ότι τα βασικά συμπληρώματα είναι τα εξής: α) μια απαρεμφατική φράση, δηλαδή η έννοια του ρήματος να συμπληρώνεται από το απαρέμφοτο ενός άλλου ρήματος, όπως π.χ. “*i ask you to forget about bush for a moment and think about freedom and liberty*”, β) γερούνδιο, το οποίο μπορεί να βρίσκεται μόνο του ή να συνοδεύεται από κάποια πρόθεση, λ.χ. “*I recommend getting up a quarter of an hour earlier and consecrating that time to morning prayer*”, γ) μια

ονοματική πρόταση με το *that*, π.χ. “*I advise that we negotiate with both the British and this sorcerer.*” και δ) μια ονοματική φράση, η οποία επίσης μπορεί να συνοδεύεται από πρόθεση ή να συμπληρώνει μόνη της το ρήμα, π.χ. “*if you want great sound quality do n t want to spend an arm and a leg per gigabyte do n t mind being something of an iconoclast among many ipod users and are willing to sacrifice a few cubic centimeters of your personal space for a large discount then i would suggest the zen xtra for you*”. Κάθε ρήμα του λεξικού, ανάλογα με το αν συντάσσεται με έναν ή περισσότερους από τους παραπάνω τρόπους, εντάσσεται σε μια συντακτική κατηγορία. Τελικά οι ξεχωριστές κατηγορίες είναι 23.

Ο πίνακας των ρημάτων παρατίθεται ολόκληρος στο παράρτημα I, εδώ εκτίθεται ένα μικρό μόνο δείγμα αυτού ώστε να φανεί η δομή του.

Εγγραφή	Σημασιολογική κατηγορία	Συντακτική κατηγορία
advise	recommendation	VB2
call on	request	VB10
have to	recommendation	VB23
discourage	recommendation negated	VB8
beg	request	VB9

Η σημασιολογική κατηγορία που παίρνει κάθε ρήμα είναι μία από τις δύο που αναφέρθηκαν παραπάνω, δηλαδή recommendation ή request. Οι συντακτικές κατηγορίες κωδικοποιούν τους τρόπους με τους οποίους συντάσσεται κάθε ρήμα. Στην κατηγορία VB2 για παράδειγμα εντάσσονται ρήματα τα οποία μπορούν να πάρουν ως συμπλήρωμα είτε μια **απαρεμφατική φράση**, π.χ. “*I advise ye to rest your mind for a while. Do not grieve over your loss*”, είτε έναν **γερονδιακό τύπο**, π.χ. “*Although teenagers have climbed Everest, I would advise waiting till you're more developed, physically and mentally.*”, είτε μια **ονοματική πρόταση με το that**, π.χ. “*I advise that we negotiate with both the British and this sorcerer.*”. Η ονομασία VB2 είναι συμβατική και χρησιμοποιείται για να ομαδοποιηθούν ρήματα που συντάσσονται με τον ίδιο τρόπο.

Στο δεύτερο λεξικό (**DirectiveLexExpr**) εντάσσονται άλλα μέρη του λόγου – εκτός από ρήματα – καθώς και εκφράσεις που δηλώνουν κατευθυντικότητα. Πρόκειται για στερεότυπες εκφράσεις οι οποίες αποτελούνται από δύο ή και περισσότερες λέξεις,

όπως για παράδειγμα “*how about*”. Εδώ περιλαμβάνονται και ουσιαστικά που πραγματώνουν ΚΛΠ, όπως το “*recommendation*”, τα οποία αναγνωρίζονται ως ευρύτερη φράση με τις διάφορες παραλλαγές της, δηλαδή “(my) (only) recommendation is”. Τα στοιχεία που βρίσκονται σε παρένθεση δηλώνουν ότι η παρουσία τους είναι προαιρετική. Αυτά τα λεξικά μέσα θα μπορούσαν να εντάσσονται και στο πρώτο λεξικό και ο τρόπος σύνταξής τους να καθορίζεται από τους κανόνες της γραμματικής. Επιλέχθηκε ωστόσο να ενταχθούν στο λεξικό ως φράσεις, μιας και στα δεδομένα που προέκυψαν από τα σώματα κειμένων που εξετάστηκαν εμφανίζονται σχεδόν πάντα με τον ίδιο τρόπο.

Οι εγγραφές αυτού του λεξικού είναι 21 και οι ταξινόμηση γίνεται με τον ίδιο τρόπο που γίνεται και στο λεξικό των ρημάτων. Με βάση τη σημασιολογική τους κατηγορία χωρίζονται σε αιτήματα και συστάσεις, ενώ σύμφωνα με τον τρόπο σύνταξής τους διακρίνονται σε δέκα ομάδες. Ένα μικρό απόσπασμα της λίστας αυτής φαίνεται παρακάτω, ενώ στο σύνολό της παρουσιάζεται στο παράρτημα II.

Εγγραφή	Σημασιολογική κατηγορία	Συντακτική κατηγορία
would do better to	recommendation	Expr1
be/make sure	recommendation	Expr9
(my) (only) request is	request	Expr10

Όπως και στην περίπτωση των ρημάτων, έτσι και στις εκφράσεις οι συντακτικές κατηγορίες ομαδοποιούν εκφράσεις που συντάσσονται με τον ίδιο τρόπο. Για παράδειγμα η ονομασία Expr1 είναι ένα συμβατικό όνομα που δόθηκε σε όσες εκφράσεις συντάσσονται με *απαρεμφατικό* τύπο ρήματος, π.χ. “*You would do better to be more generous with your money than yourself.*”

3.4. Δομικά σχήματα (Patterns)

Οι τέσσερις βασικές ομάδες συντακτικών συμπληρωμάτων που αναφέρθηκαν, αναλύθηκαν λεπτομερέστερα και αποτέλεσαν τελικά 8 δομικά σχήματα με βάση τα οποία έγινε η τελική ταξινόμηση. Το εργαλείο που κατασκευάστηκε στηρίζεται στην επιφανειακή σύνταξη, δηλαδή λαμβάνει υπόψη τη γραμμική εμφάνιση των λεκτικών μονάδων (tokens). Αυτός ήταν και ο λόγος για τον οποίο η κάθε κύρια κατηγορία χωρίστηκε σε επιμέρους, προκειμένου τα αποτελέσματα να είναι πιο ακριβή.

Η θεμελιώδης δομή των κατευθυντικών προτάσεων είναι η ακόλουθη:

Υποκείμενο + Έκφραση κατευθυντικότητας + (Αποδέκτης) + Αντικείμενο

Ως **Υποκείμενο** εκλαμβάνεται ο ομιλητής ή συγγραφέας, δηλαδή το πρόσωπο που απευθύνει την ΚΛΠ και εκφράζεται με μια προσωπική αντωνυμία πρώτου προσώπου (I, we). **Έκφραση κατευθυντικότητας** είναι το συγκεκριμένο γλωσσικό μέσο με το οποίο κάθε φορά δηλώνεται η ΚΛΠ, επομένως κάποιο στοιχείο από τα όσα περιέχονται στα λεξικά. **Αποδέκτης** θεωρείται το πρόσωπο από το οποίο ζητά ο ομιλητής της πραγματοποίηση της πράξης, αυτός στον οποίο έγκειται η υποχρέωση της πραγμάτωσής της. Μπορεί να είναι και αυτό προσωπική αντωνυμία αλλά μόνο δεύτερου ή τρίτου προσώπου, ή να είναι μια ονοματική φράση. Αυτό το στοιχείο, αν και είναι πολύ σημαντικό, βρίσκεται σε παρένθεση διότι δεν αναφέρεται πάντα. Υπάρχουν περιπτώσεις δηλαδή στις οποίες δεν καθορίζεται ρητά ή υπονοείται. Τέλος, **Αντικείμενο** θεωρείται αυτό το οποίο ο ομιλητής ζητά από τον ακροατή ή αυτό το οποίο του συστήνει. Αυτό πραγματώνεται είτε με ένα ουσιαστικό ή ονοματική φράση, είτε με ένα ρήμα.

Με βάση αυτά, τα δομικά σχήματα όπως διαμορφώθηκαν καταγράφονται αναλυτικά παρακάτω. Τα τρία πρώτα συστατικά, δηλαδή Υποκείμενο, Κατευθυντικό (στοιχείο) και Αποδέκτης σημειώνονται έτσι και πραγματώνονται όπως προαναφέρθηκε. Τα υπόλοιπα στοιχεία από τα οποία αποτελείται το κάθε σχήμα δηλώνονται ανάλογα με το μέρος του λόγου στο οποίο ανήκουν. Αυτά αποτυπώνονται με τα ονόματα που παίρνουν από το αντίστοιχο εργαλείο του GATE (Part-of-Speech Tagger). Μερικές από αυτές τις ετικέτες, που χρησιμοποιούνται και στα δομικά σχήματα, είναι:

VB – ρήμα στη βασική του μορφή, για την αγγλική περιλαμβάνει την προστακτική, το απαρέμφατο και την υποτακτική

VBG – γερούνδιο ή ενεργητική μετοχή

TO – το to επισημαίνεται έτσι σε όλες τις περιπτώσεις, είτε είναι σύνδεσμος είτε πρόθεση

IN – με την ετικέτα αυτή σημειώνονται οι προθέσεις και οι υποτακτικοί σύνδεσμοι

NN – ουσιαστικό στον ενικό

DT – προσδιοριστές, κυρίως δηλαδή το οριστικό και το αόριστο άρθρο, καθώς και στοιχεία που μπορούν να λάβουν αυτή τη θέση όπως τα *any* και *some*

JJ – επίθετο

Να σημειωθεί ότι ως NP το εργαλείο του GATE επισημαίνει τα κύρια ονόματα στον ενικό αριθμό. Εδώ όμως χρησιμοποιείται ως ακρώνυμο για το Noun Phrase (NP), την Ονοματική Φράση (ΟΦ) δηλαδή, για λόγους οικονομίας. Ως ΟΦ ορίζουμε οποιονδήποτε συνδυασμό προσδιοριστών, επιθέτων και ουσιαστικών.

Τα δομικά σχήματα λοιπόν είναι:

1) Υποκείμενο + Κατευθυντικό + Αποδέκτης + to + VB

Στη δομή αυτή εμφανίζεται ο σύνδεσμος <to> υποχρεωτικά και συμπληρώνεται από ένα ρήμα σε απαρεμφατική μορφή.

π.χ. “*i urge you to join me...*”

2) Υποκείμενο + Κατευθυντικό + (Αποδέκτης) +VBG

Εδώ ο Αποδέκτης εμφανίζεται προαιρετικά και το Αντικείμενο αποτελεί ένα γερούνδιο.

π.χ. “*Now I advise walking...*”

3) Υποκείμενο + Κατευθυντικό + (Αποδέκτης) + (that) +(Αποδέκτης) + VB

Στο σχήμα αυτό το Αντικείμενο βρίσκεται μέσα σε μια ονοματική πρόταση που εισάγεται με το *that* – το οποίο σε πολλές περιπτώσεις παραλείπεται, γι’ αυτό βρίσκεται μέσα σε παρένθεση. Χαρακτηριστικό είναι ότι ο Αποδέκτης μπορεί να δηλώνεται κανονικά μετά τον σύνδεσμο με αντωνυμία ή ονοματική φράση, μπορεί όμως να εμφανίζεται και ανάμεσα σε αυτόν και το Υποκείμενο. Στην πρώτη θέση εκφράζεται με προσωπική αντωνυμία ή ονοματική φράση και είναι δυνατό να συνοδεύεται από μια πρόθεση (*of, for, with*).

π.χ. “*I advise the preservice teachers that they should build...*”

4) Υποκείμενο + Κατευθυντικό + IN + PRP/NP + to +VB

Καταρχάς, με το NP εννοείται μια ονοματική φράση που μπορεί να αποτελείται μόνο από προσδιοριστές, επίθετα και ουσιαστικά σε όλους τους δυνατούς συνδυασμούς. Το σύμπλεγμα IN + PRP/NP δηλώνει τον Αποδέκτη, ενώ το Αντικείμενο είναι και εδώ ένα

ρήμα σε απαρεμφατική μορφή. Ο λόγος για τον οποίο κρίθηκε απαραίτητο να χωριστεί αυτή η δομή από την πρώτη, είναι η ακρίβεια, γιατί όταν ο αντίστοιχος κανόνας εφαρμοζόταν σε ρήματα που συντάσσονται χωρίς την πρόθεση, προκαλούσε πολλά λάθος αποτελέσματα.

π.χ. *“I beg of you to stay...”*

5) Υποκείμενο + Κατευθυντικό + Αποδέκτης + IN + VBG/NP

Εδώ, η πρόθεση (IN) είναι μία εκ των from, against, out of, about. Το χαρακτηριστικό αυτών είναι ότι προσδίδουν στο κατευθυντικό ρήμα ή την αντίστοιχη έκφραση αρνητική πολικότητα. Το Αντικείμενο στο σχήμα αυτό μπορεί να είναι ένα γερούνδιο ή μια ΟΦ.

π.χ. *“I persuaded them out of groping...”*

6) Υποκείμενο + Κατευθυντικό + Αποδέκτης + into + VBG/NP

Αυτή η δομή είναι ακριβώς ίδια με την παραπάνω, με τη διαφορά ότι χρησιμοποιείται μόνο η πρόθεση into η οποία δεν αλλάζει την πολικότητα του Κατευθυντικού. Αυτός είναι και ο λόγος για τον οποίο θεωρήθηκε σωστό να γίνει αυτός ο διαχωρισμός, καθαρά σημασιολογικά.

π.χ. *“I persuaded her into doing...”*

7) Υποκείμενο + Κατευθυντικό + (Αποδέκτης) + NP

Σε αυτό το σχήμα η δήλωση του Αποδέκτη είναι προαιρετική, ενώ το Αντικείμενο είναι μια ΟΦ.

π.χ. *“I prescribed him a daily dosage...”*

8) Υποκείμενο + Κατευθυντικό + (Αποδέκτης) + IN + NP

Και εδώ ο Αποδέκτης δε δηλώνεται ρητά σε όλες τις περιπτώσεις και το Αντικείμενο της ΚΛΠ είναι μια ΟΦ, η οποία όμως είναι συμπλήρωμα μιας Προθετικής Φράσης. Η ακρίβεια ήταν και πάλι ο λόγος για τον οποίο διαχωρίστηκε το σχήμα αυτό από το παραπάνω, καθώς προκαλούσε λανθασμένα αποτελέσματα.

π.χ. *“I beg you for forgiveness.”*

3.5. Κανόνες

Τα δομικά σχήματα, υλοποιούνται σε κανόνες. Ουσιαστικά οι κανόνες αυτοί αποτελούν τη γραμματική των ΚΛΠ. Λειτουργούν σε επίπεδο επιφανειακής συντακτικής δομής, που σημαίνει ότι διατρέχουν το κείμενο στο οποίο εφαρμόζονται γραμμικά, λέξη προς λέξη και αναζητούν συγκεκριμένα στοιχεία τα οποία καθορίζουν την ΚΛΠ.

Η γραμματική υλοποιείται σε JAPE (Java Annotation Patterns Engine). Πρόκειται για έναν μετατροπέα πεπερασμένης κατάστασης (Finite State Transducer) ο οποίος επεξεργάζεται τις επισημειώσεις που παράγονται από άλλα εργαλεία (τον POS Tagger) χρησιμοποιώντας κανονικές εκφράσεις (regular expressions). Με απλά λόγια, πρόκειται για ένα εργαλείο δημιουργίας γραμματικών, το οποίο χρησιμοποιεί τα μεταδεδομένα που έχουν λάβει οι λέξεις ενός κειμένου ως προς το μέρος του λόγου στο οποίο ανήκουν και τα χαρακτηριστικά τους και δημιουργεί κανόνες γενικευτικούς, χωρίς δηλαδή να πρέπει να φτιάχνεται ένας κανόνας για κάθε λεξική πραγμάτωση. Όταν επομένως, το δομικό σχήμα που αποτυπώνεται σε κανόνα περιλαμβάνει ένα ουσιαστικό, δε χρειάζεται να καταγράφονται όλες οι πιθανές λέξεις – κάτι που θα ήταν σχεδόν αδύνατο – αλλά δηλώνεται απλώς η ετικέτα NN που εμπεριέχει όλες τις πιθανότητες. Κάθε γραμματική αυτού του είδους αποτελείται από μια σειρά φάσεων. Οι φάσεις της συγκεκριμένης γραμματικής είναι τα δύο λεξικά που εκτέθηκαν προηγούμενα και οι κανόνες χρησιμοποιούν τα σημασιολογικά και συντακτικά χαρακτηριστικά των λημμάτων που περιλαμβάνονται σε αυτά.

Η λειτουργία της γραμματικής είναι αρκετά απλή. Οι κανόνες εφαρμόζονται πάνω σε οποιοδήποτε κείμενο και έχουν σχεδιαστεί για να εντοπίζουν τις ΚΛΠ. Πιο συγκεκριμένα, τα στοιχεία που έχει οριστεί να επιστρέφουν οι κανόνες ως αποτέλεσμα είναι τρία: α) τη γλωσσική πραγμάτωση της ΚΛΠ, δηλαδή ένα ρήμα ή μια έκφραση από όσα περιλαμβάνονται στα δύο λεξικά. Αυτό το στοιχείο ορίζεται ως Κατευθυντικό (**Direct**) και παίρνει από το αντίστοιχο λεξικό τη σημασιολογική του κατηγορία (insight), β) το Αντικείμενο της ΚΛΠ (**Object_of_Direct**). Δύο είναι οι τύποι Αντικειμένου που μπορούν να εμφανίζονται, είτε ένας ρηματικός τύπος, δηλαδή απαρέμφατο, γερούνδιο, τύπος υποτακτικής, είτε μια ονοματική φράση. Στη δεύτερη αυτή περίπτωση αυτό που επισημαίνεται τελικά ως Αντικείμενο είναι το ουσιαστικό ή η αντωνυμία που αποτελεί το βασικό στοιχείο της ΟΦ, και γ) το στοιχείο της αρνητικής πολικότητας (**negated**). Αυτό ενεργοποιείται μόνο στις προτάσεις όπου

εντοπίζεται κάποια αρνητική λέξη οπότε παίρνει την τιμή true. Αυτό συμβαίνει με όσα ρήματα του λεξικού φέρουν εγγενώς αυτό το σημασιολογικό χαρακτηριστικό, καθώς και τα ρήματα που το αποκτούν λόγω της σύνταξής τους σύμφωνα με το δομικό σχήμα 5 (βλ. παραπάνω στο 3.4). Σε κάθε τύπο πληροφορίας που επιστρέφει η γραμματική, αναφέρεται και ο αριθμός του κανόνα, ώστε να εντοπίζεται άμεσα αν ο κανόνας που ενεργοποιήθηκε είναι ο σωστός.

Υπάρχουν δύο σύνολα κανόνων, αυτοί που αφορούν στα ρήματα και εκείνοι που βασίζονται στις εκφράσεις. Συνολικά οι κανόνες (**DirectiveRules**) που περιλαμβάνονται στη γραμματική είναι 22. Κάθε κανόνας πραγματώνει κάποιο από τα δομικά σχήματα που περιγράφηκαν, ενώ συμπεριλαμβάνει και κάποια άλλα στοιχεία τα οποία μπορεί να εμφανίζονται προαιρετικά, για παράδειγμα τροπικά ρήματα τα οποία είναι πιθανό να παρεμβάλλονται μεταξύ των διάφορων συστατικών του κανόνα, και κυρίως ανάμεσα στο Υποκείμενο και το Κατευθυντικό. Ένα άλλο στοιχείο που προβλέπεται από τους κανόνες είναι η περίπτωση κάποιο από τα ρήματα του λεξικού να βρίσκεται σε κάποιον από τους σύνθετους χρόνους της αγγλικής γλώσσας.

Υπάρχουν ορισμένα στοιχεία με πολλαπλές τιμές τα οποία χρειάζεται κάθε κανόνας να αναζητά. Λόγου χάρη, το Υποκείμενο της κατευθυντικής πράξης μπορεί να είναι μια αντωνυμία πρώτου προσώπου, δηλαδή να παίρνει μία από τις τιμές *I* και *we*. Επίσης οι σύνθετοι χρόνοι στα αγγλικά είναι αρκετοί και κάθε κανόνας πρέπει να ελέγχει το κατευθυντικό ρήμα για το αν ανήκει σε οποιονδήποτε από αυτούς. Προκειμένου οι έλεγχοι αυτοί να μη χρειάζεται να ξαναγράφονται με όλες τις πιθανές τιμές τους σε κάθε κανόνα, δημιουργήθηκαν μακροεντολές (macros). Ένα παράδειγμα τέτοιας εντολής είναι το ακόλουθο:

```
MACRO:COMP2
(
  ({Token, Token.root ==~"be"})
  ({Token, Token.root ==~"go"})
  ({Token, Token.root ==~"to", Token.category=~"TO"})
)
```

Εκτελώντας την παραπάνω εντολή πριν το ρήμα που εκφράζει την κατευθυντικότητα, η γραμματική ελέγχει αυτόματα αν αυτό εντάσσεται στη δομή του μέλλοντα “be going to”. Ακολουθούν δύο παραδείγματα ολόκληρων κανόνων, ένας για ρήματα από το

αντίστοιχο λεξικό και ένας που πραγματώνει τη συντακτική δομή έκφρασης από το δεύτερο λεξικό.

Rule 1

PRP1 + (MD) + (COMP1) + (COMP2) + (negation) + (RB) + VB2 + (RB) + (PRP) + VBG

Rule 2

Expr2 + (RB) + (negation) + VBG

Όλα τα στοιχεία που βρίσκονται σε παρένθεση είναι προαιρετικά, ενώ όσα είναι εκτός είναι απαραίτητα και αποτελούν τα βασικά συστατικά του κάθε κανόνα.

PRP1 είναι προσωπική αντωνυμία πρώτου προσώπου (I, we). Πρόκειται για μακροεντολή.

MD είναι οποιοδήποτε τροπικό ρήμα μπορεί να υπάρχει πριν το κατευθυντικό (would, should e.t.c.)

COMP1 είναι μακροεντολή που ελέγχει αν το ρήμα που δηλώνει την ΚΛΠ βρίσκεται σε κάποιον από τους σύνθετους χρόνους, παροντικούς, μελλοντικούς ή παρελθοντικούς, εκτός από τη μελλοντική δομή “be going to”.

COMP2 μακροεντολή που ελέγχει τη μελλοντική δομή “be going to”.

Negation μπορεί να είναι στοιχείο άρνησης του ρήματος (no, not). Υπενθυμίζεται ότι όταν υπάρχει αλλάζει την τιμή της άρνησης σε αληθή (negated=true)

RB οποιοδήποτε επίρρημα

VB2 ο συντακτικός τύπος του ρήματος όπως σημειώνεται στο λεξικό

PRP οποιαδήποτε προσωπική αντωνυμία

VBG γερουνδιακός τύπος

Expr2 ο συντακτικός τύπος της έκφρασης από το αντίστοιχο λεξικό

Ο κανόνας 1 ενεργοποιείται σε προτάσεις όπως :

“So, no, I would not advise taking a supplement unless your doctor specifically recommends it.”

Ο κανόνας 2 αντίστοιχα ενεργοποιείται για προτάσεις όπως:

“But since I'm stuck with you, how about giving me a name?”

Το αποτέλεσμα της γραμματικής για τις παραπάνω προτάσεις είναι το εξής:

Direct: *“insight = recommendation”*. Και στις δύο περιπτώσεις είναι το ίδιο, όπως προκύπτει από τη σημασιολογική κατηγορία στην οποία εντάσσονται τόσο τα ρήματα της κατηγορίας VB2 όσο και η έκφραση Expr2.

Neg_Direct: Στην πρώτη πρόταση εντοπίζεται το *“not”* και επιστρέφει Neg_Direct = true, ενώ στη δεύτερη πρόταση δεν ενεργοποιείται καθόλου επειδή λείπει αρνητικό στοιχείο.

Object_of_Direct: Στην πρώτη περίπτωση είναι το *“taking”* και στην δεύτερη το *“giving”*.

4. Εφαρμογή – Αξιολόγηση

4.1. Εφαρμογή

Για την υλοποίηση του εργαλείου χρησιμοποιήθηκε η πλατφόρμα επεξεργασίας φυσικής γλώσσας GATE. Σε αυτήν υπάρχει ένα σύνολο εργαλείων για εξαγωγή πληροφορίας από κείμενα και για διάφορων ειδών επεξεργασία. Πρόκειται για ανοιχτό λογισμικό το οποίο αρχικά αναπτύχθηκε από το Πανεπιστήμιο του Sheffield και πλέον χρησιμοποιείται και εμπλουτίζεται από ένα μεγάλο πλήθος προγραμματιστών, χρηστών, διδασκόντων και μαθητών.

Η εφαρμογή της γραμματικής έγινε σε δύο στάδια. Στην πρώτη φάση, έγινε μια πρώτη επεξεργασία του σώματος κειμένων. Πρόκειται για μια φάση προ-επεξεργασίας εφαρμόζοντας εργαλεία ενσωματωμένα στο GATE. Το κείμενο που προέκυψε από αυτή την επεξεργασία αποτέλεσε την είσοδο για την εφαρμογή της γραμματικής.

Για την προ-επεξεργασία χρησιμοποιήθηκαν συγκεκριμένα αρθρώματα (modules) του συστήματος εξαγωγής πληροφορίας ANNIE, το οποίο αποτελεί ένα από τα σημαντικά εργαλεία του GATE. Πιο συγκεκριμένα, δημιουργήθηκε μια αλυσίδα αρχικής επεξεργασίας (pipeline) στην οποία εντάχθηκαν τρία από τα υποσυστήματα του ANNIE και δύο άλλα εργαλεία του GATE, τα οποία περιγράφονται συνοπτικά παρακάτω.

Document Reset PR: Αυτό αναλαμβάνει να σβήσει από τα κείμενα οποιαδήποτε επισημείωση μπορεί να είχε γίνει με άλλα εργαλεία.

ANNIE English Tokenizer: Χωρίζει το κείμενο στα μικρότερα συστατικά του, όπως αριθμούς, σημεία στίξης και ελάχιστες λεξικές μονάδες.

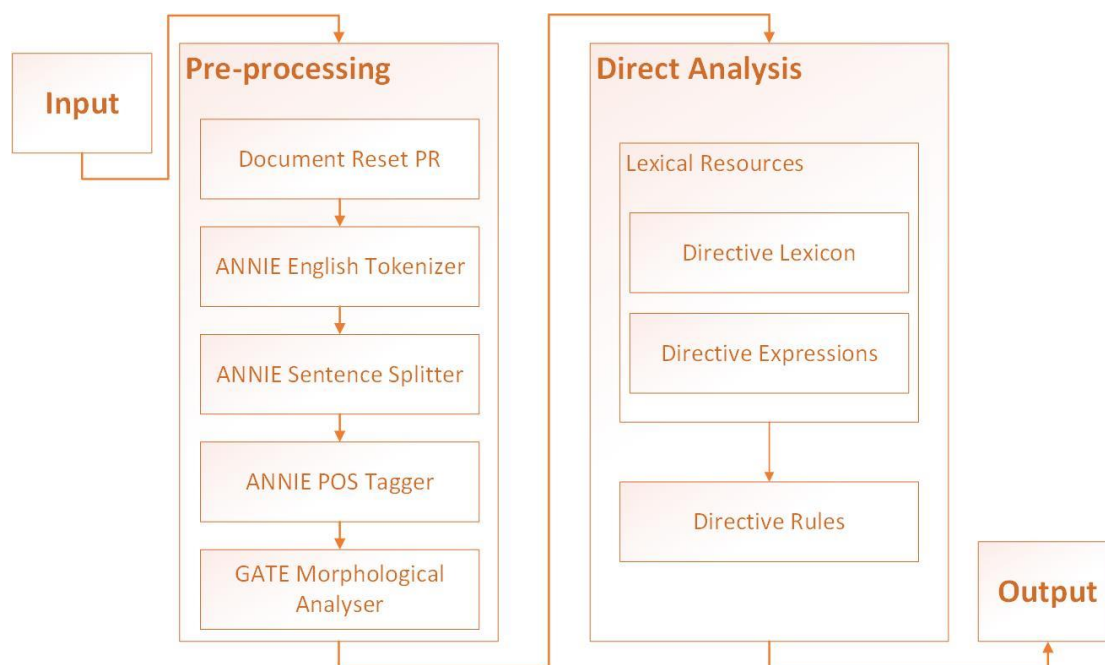
ANNIE English Sentence Splitter: Χωρίζει το κείμενο σε προτάσεις.

ANNIE POS Tagger: Δίνει σε κάθε λεκτική μονάδα και σύμβολο μια ετικέτα για την συντακτική του κατηγορία, δηλαδή το μέρος του λόγου στο οποίο ανήκει.

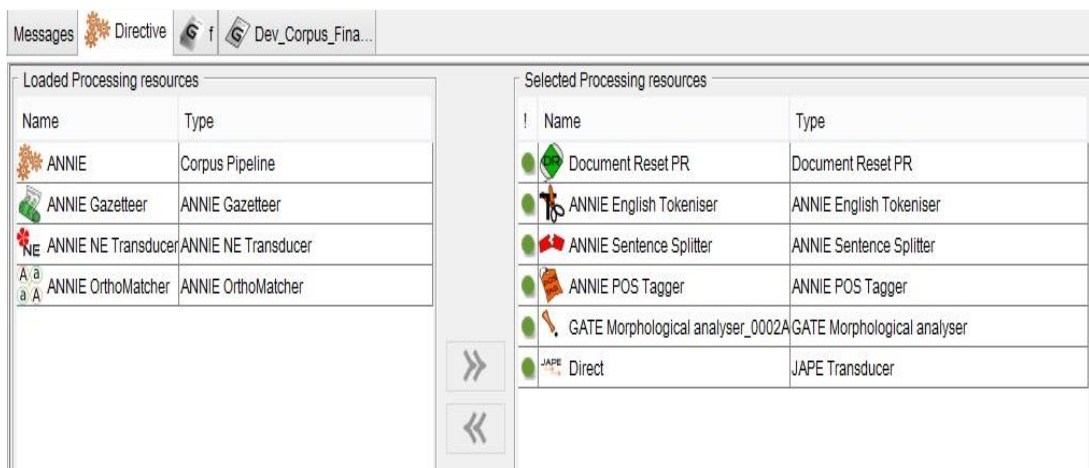
GATE Morphological Analyzer: Διακρίνει σε κάθε λεκτική μονάδα το λήμμα και τα προσφύματα από τα οποία αποτελείται.

Η παραπάνω διαδικασία εκτελείται διαδοχικά, με τη σειρά που αναφέρθηκε. Κάθε εργαλείο παίρνει σαν είσοδο το αποτέλεσμα της επεξεργασίας του προηγούμενου. Στο τέλος αυτής της αλυσίδας προστίθεται η γραμματική που περιγράφηκε. Είσοδο σε αυτήν αποτελεί το αποτέλεσμα της προ-επεξεργασίας. Η γραμματική αποτελεί ουσιαστικά τη φάση της επεξεργασίας του κειμένου, εφαρμόζοντας τα λεξικά και τους κανόνες.

Στο σχήμα που ακολουθεί φαίνονται τα στάδια προ-επεξεργασίας (Preprocessing) και επεξεργασίας (Direct Analysis) όπως περιγράφηκαν πιο πάνω.



Η αλυσίδα επεξεργασίας του σώματος κειμένου όπως εμφανίζεται στο GATE, παρουσιάζεται παρακάτω.



Το τελικό κείμενο που προκύπτει ως έξοδος από όλη τη διαδικασία επεξεργασίας, φέρει ξεχωριστές ετικέτες για κάθε στοιχείο που επισημαίνουν τα εργαλεία προ-επεξεργασίας και επιπλέον τις ετικέτες που προστίθενται από την εφαρμογή της γραμματικής, δηλαδή των λεξικών και των κανόνων. Κάθε ένα από αυτά σημειώνεται με διαφορετικό χρώμα, ενώ στο κάτω μέρος, κάτω από το επισημειωμένο κείμενο, φαίνονται τα αποτελέσματα των κανόνων, σύμφωνα με αυτά που έχουν οριστεί να εμφανίζονται. Στην παρακάτω οθόνη είναι εμφανή όλα τα παραπάνω.

want us to pay the entire expense of healthcare through our taxes for goodness sakes noooooooooooooo do n t
 they need to be postponed until hey can be held fairly and with adequate security and please do n t go off on some tangent
 about how we wo n t accept the results of the election anyway because we ll just assume it was rigged i m not in the mood for
 red herring
 fell for that fallacy how easy you are controlled the rest is just nitpicking and handwring you meant lies and distortion please
 do n t make me correct you again
 i would suggest that perhaps you should not paint all people of faith with the same broad brush
 The judge said, "I adjure you to answer truthfully."
 I solemnly adjure you to maintain it.
 " With all the solemn authority of the Bureau, I admonish her sternly not to go around blabbing our conversation to the world. "
 And such is the legacy I would leave which I admonish him to respect.
 Furthermore, I admonish all who have engaged in this item of gossip over Mr. Arnett's past to put up or shut up.
 Now if it's there -- so if they're doing it, I would advise them do if after 4 PM.
 if anyone watching this program, you know, is doing that, I'd advise them to get a full blood panel with a general practitioner,
 I advise ye to rest your mind for a while. Do not grieve over your loss
 I advise any homemaker to set up a dedicated IRA in addition to a spouse's retirement
 Nevertheless, I wouldn't advise everyone to cut back on their voice service.
 So, no, I would not advise taking a supplement unless your doctor specifically recommends it.
 Although teenagers have climbed Everest, I would advise waiting till you're more developed, physically and mentally.
 Normally I will do better for a loan. But on this I advise you sell it outright. "
 And I advise people not to do it.
 That Is why I changed my recommendation from saying more is better. Now I advise walking two miles in 30 minutes three
 times a week
 I advise that we negotiate with both the British and this sorcerer.

Direct
 DirectiveLex
 DirectiveLexExpr
 Neg_Direct
 Object_of_Direct
 Sentence
 SpaceToken
 Split
 Token
▶ Original markings

Type	Set	Start	End	Id	Features
Object_of_Direct		71004	71007	48420	{object=set, rule=1}
Neg_Direct		71086	71089	48423	{negated=true, rule=1}
Direct		71090	71096	48421	{insight=recommendation, rule=1}
Object_of_Direct		71109	71112	48422	{object=cut, rule=1}
Neg_Direct		71159	71162	48426	{negated=true, rule=3}
Direct		71163	71169	48424	{insight=recommendation, rule=3}
Object_of_Direct		71170	71176	48425	{object=take, rule=3}
Direct		71287	71293	48427	{insight=recommendation, rule=3}
Object_of_Direct		71294	71301	48428	{object=wait, rule=3}

4.2. Αξιολόγηση

Για την αξιολόγηση της αποτελεσματικότητας του εργαλείου, χρησιμοποιήθηκαν δύο σώματα κειμένων. Συγκεκριμένα, τα test corpora αποτέλεσαν τα αντίστοιχα test data του Task 4 του SemEval 2014 που αφορούσε σε Aspect Based Sentiment Analysis. Τα corpora αυτά είναι δύο:

- Ένα για κριτικές φορητών υπολογιστών (**Laptop Reviews**)
- Και ένα για κριτικές εστιατορίων (**Restaurant Reviews**)

Ειδικότερα, πρόκειται για συλλογές προτάσεων και όχι ολόκληρων κειμένων. Κάθε ένα από τα δύο αυτά σώματα αποτελείται από 800 μεμονωμένες προτάσεις, άρα η αξιολόγηση έγινε σε σύνολο 1600 προτάσεων. Αν ληφθεί υπόψη ότι το σώμα με το οποίο έγινε η ανάπτυξη του εργαλείου (development corpus) περιέχει 4241 κείμενα, υφίσταται μια καλή ποσοτική αναλογία με το σώμα δοκιμής.

Ένα επιπλέον σημαντικό χαρακτηριστικό των δεδομένων στα οποία δοκιμάστηκε η γραμματική των ΚΛΠ, είναι ότι προέρχονται από δύο διαφορετικές θεματικές περιοχές, δηλαδή κριτικές για φορητούς υπολογιστές και για εστιατόρια. Ακόμα σπουδαιότερο όμως είναι το γεγονός ότι αυτά τα θεματικά πεδία είναι διαφορετικά από εκείνα που χρησιμοποιήθηκαν αρχικά για τη δημιουργία του εργαλείου. Ο λόγος που επελέγη να γίνει αυτό είναι επειδή εξ αρχής στόχος ήταν η γραμματική να είναι ανεξάρτητη θεματικού πεδίου, να μπορεί δηλαδή να έχει αποτελέσματα σε οποιαδήποτε συλλογή προτάσεων ή κειμένων, χωρίς περιορισμούς σημασιολογικούς.

Προκειμένου να γίνει η αξιολόγηση, δημιουργήθηκε ένα Annotation Schema (**DirectivesAnnotationSchema**). Αυτό είναι ένα αρχείο σε γλώσσα xml, το οποίο αναγνωρίζεται από το GATE και με το οποίο μπορεί κανείς να καθορίζει τύπους επισημειώσεων. Χρησιμοποιώντας αυτό, τα σώματα προτάσεων που χρησιμοποιήθηκαν για την αξιολόγηση επισημειώθηκαν χειρωνακτικά ως προς τα στοιχεία που θα έπρεπε να επιστρέφει σωστά το εργαλείο ανάλυσης ΚΛΠ. Το αποτέλεσμα αυτής της διαδικασίας αποτέλεσε το gold corpus, το οποίο συγκρίθηκε με το αντίστοιχο που επισημειώθηκε αυτόματα από τη μηχανή εφαρμόζοντας τη γραμματική και μετρήθηκε με αυτόν τον τρόπο η αποτελεσματικότητά της.

Για την ποσοτική μέτρηση των αποτελεσμάτων χρησιμοποιήθηκαν τρία μέτρα που εφαρμόζονται ευρέως στην ανάκτηση πληροφορίας (Information Retrieval). Αυτά είναι η **ακρίβεια** (precision), η **ανάκληση** (recall) και ο **αρμονικός μέσος** (f-measure). Προτού παρατεθούν τα αριθμητικά αποτελέσματα, κρίνεται σκόπιμο να οριστούν οι τρεις αυτές μετρικές.

Η **ακρίβεια** (precision) είναι το μέτρο το οποίο σχετίζεται με την ορθότητα των αποτελεσμάτων του εργαλείου. Με άλλα λόγια, μετράει σε τι ποσοστό όσα εντόπισε η εφαρμογή είναι σωστά και έπρεπε να εντοπιστούν. Η **ανάκληση** (recall) μετράει την αποτελεσματικότητα του εργαλείου ως προς τα αποτελέσματα. Δηλαδή, μετρά στην ουσία το ποσοστό των σωστών επισημειώσεων που το σύστημα έκανε σε σχέση με όλα τα σωστά αποτελέσματα που υπάρχουν μέσα στο κείμενο ή τα κείμενα στα οποία γίνεται ο έλεγχος. Τέλος, ο **αρμονικός μέσος** (f-measure) λαμβάνει υπόψη τόσο την ακρίβεια όσο και την ανάκληση. Γενικότερα μεταξύ των δύο αυτών μέτρων υπάρχει αλληλεπίδραση, οπότε όσο αυξάνεται η ακρίβεια τόσο μειώνεται η ανάκληση και αντίστροφα. Ο αρμονικός μέσος μετράει πόσο καλή ισορροπία υπάρχει μεταξύ των δύο αυτών. Πρέπει να σημειωθεί βέβαια ότι ο αρμονικός μέσος θεωρεί ότι τόσο η ακρίβεια

όσο και η ανάκληση έχουν την ίδια βαρύτητα για την αποτίμηση της αποτελεσματικότητας. Αυτό ωστόσο δεν ισχύει πάντα, αφού κάθε εργαλείο που κατασκευάζεται έχει μια συγκεκριμένη στόχευση και κατά συνέπεια μπορεί η μία μετρική να θεωρείται σημαντικότερη από την άλλη. Έτσι συμβαίνει και στην παρούσα εργασία στην οποία δίνεται μεγαλύτερη σημασία στην ακρίβεια, δηλαδή στην ορθότητα των αποτελεσμάτων.

Παρακάτω παρατίθενται τα ποσοτικά αποτελέσματα της αξιολόγησης του εργαλείου σε δύο πίνακες, έναν για τα Laptops και έναν για τα Restaurants.

Laptops						
	Found	Correct	Precision	Existed	Recall	F-measure
Direct	20	17	85%	40	42%	57%
Insight	20	17	85%	40	42%	57%
Object	20	17	85%	39	44%	58%
Negated	-	-	-	-	-	-

Η αξιολόγηση έγινε με βάση τα στοιχεία που έχουν οριστεί από τους κανόνες να φέρνει ως αποτελέσματα. Αυτά είναι τρία: η λέξη ή φράση που χαρακτηρίζεται ως ΚΛΠ (**Direct**), το Αντικείμενό της (**Object**) και την άρνηση (**Negated**) όταν αυτή εμφανίζεται. Η σημασιολογική κατηγορία (Insight) στην οποία ανήκει κάθε κατευθυντικό εκφώνημα αντιμετωπίζεται από τους κανόνες ως χαρακτηριστικό του Direct και δεν παίρνει ξεχωριστή ετικέτα. Ωστόσο, στη μέτρηση των αποτελεσμάτων έχει σημασία να ελεγχθεί κατά πόσο είναι σωστή η κατηγορία στην οποία εντάσσεται κάθε ΚΛΠ που εντοπίζεται. Έτσι, εφαρμόζονται οι τρεις μετρικές που περιγράφηκαν, Precision, Recall και F-measure, για κάθε στοιχείο ξεχωριστά. Στο συγκεκριμένο σώμα δεν βρέθηκαν καθόλου στοιχεία άρνησης, γι' αυτό και δεν υπάρχει μέτρηση για αυτά. Επίσης, θεωρήθηκε σημαντικό να σημειωθούν για κάθε κατηγορία τα αποτελέσματα που βρέθηκαν (**Found**), πόσα από αυτά ήταν σωστά (**Correct**) και πόσα υπήρχαν σε όλο το σώμα προτάσεων (**Existed**), με βάση τα οποία υπολογίστηκαν οι τιμές για την ακρίβεια και την ανάκληση.

Ο αντίστοιχος πίνακας για τα Restaurants είναι ο εξής:

Restaurants						
	Found	Correct	Precision	Existed	Recall	F-measure
Direct	26	25	96%	88	28%	43%
Insight	26	25	96%	88	28%	43%
Object	26	25	96%	83	30%	46%
Negated	5	5	100%	17	29%	45%

Το πρώτο που γίνεται εύκολα αντιληπτό είναι το μεγάλο ποσοστό ακρίβειας που παρατηρείται τόσο στο θεματικό πεδίο των φορητών υπολογιστών όσο και σε εκείνο των εστιατορίων. Μάλιστα στο δεύτερο είναι αισθητά υψηλότερο και στο ζήτημα της άρνησης αγγίζει το 100%. Αυτό το αποτέλεσμα ταιριάζει με την προσέγγιση που ακολουθήθηκε στην εργασία, δηλαδή την έμφαση στην ακρίβεια. Επιπλέον, παρατηρώντας τη διαφορά μεταξύ των δύο θεματικών, σε επίπεδο ακρίβειας, ίσως στο πεδίο των εστιατορίων οι τρόποι με τους οποίους εκφράζεται η κατευθυντικότητα είναι πιο εύκολο να προβλεφθούν σε σχέση με τους φορητούς υπολογιστές.

Είναι επίσης ξεκάθαρο αυτό που ειπώθηκε παραπάνω για την αλληλεπίδραση μεταξύ ακρίβειας και ανάκλησης. Φαίνεται ότι τα ποσοστά ανάκλησης είναι αρκετά χαμηλά και στα δύο σώματα κειμένων. Ως προς το θεματικό πεδίο, εδώ δείχνουν να αντιστρέφονται τα πράγματα και να αποδεικνύεται ότι στις κριτικές εστιατορίων χρησιμοποιούνται γλωσσικά μέσα έκφρασης που η μηχανή αδυνατεί να εντοπίσει πολύ περισσότερο από ότι στους φορητούς υπολογιστές.

Γενικότερα, όπως προκύπτει από το ποσοστό του αρμονικού μέσου, στους φορητούς υπολογιστές υπάρχει μεγαλύτερη ισορροπία μεταξύ των όσων εντοπίζονται σωστά και των όσων θα έπρεπε να αναγνωρίζονται. Αντιθέτως, στα εστιατόρια φαίνεται ότι οι τρόποι με τους οποίους προβλέπει το εργαλείο ότι εκφράζονται ΚΛΠ είναι σωστοί, υπάρχουν όμως αρκετά άλλα γλωσσικά μέσα έκφρασης τα οποία δεν έχουν προβλεφθεί.

Συνοψίζοντας, το εργαλείο επιτυγχάνει μεγάλη ακρίβεια εις βάρος της ανάκλησης. Αυτό σημαίνει ότι μένουν περιθώρια για βελτίωση προκειμένου να επιτευχθεί μεγαλύτερη ισορροπία ανάμεσα στα δύο αυτά μέτρα σύγκρισης. Τέλος, θα ήταν εξίσου χρήσιμο να γίνει αξιολόγηση σε περισσότερα κείμενα, ώστε να βγουν πιο ασφαλή συμπεράσματα σχετικά με τα λάθη και τις δυνατότητες βελτίωσης. Κι αυτό γιατί, παρά τον ικανό αριθμό προτάσεων που συνολικά χρησιμοποιήθηκαν, οι

εμφανίσεις ΚΛΠ είναι σχετικά λίγες. Ίσως σε ένα μεγαλύτερο σώμα να ήταν περισσότερες, θα μπορούσε όμως να έχει να κάνει και με το θεματικό πεδίο στο οποίο ανήκουν τα κείμενα.

4.3. Σχολιασμός αποτελεσμάτων – Ανάλυση λαθών

Εδώ σχολιάζονται τα αριθμητικά αποτελέσματα που προέκυψαν από την αξιολόγηση του υλικού, σε σχέση με τα θέματα που απασχολούν τη συγκεκριμένη εργασία. Γίνεται αποτίμηση της αποτελεσματικότητας του εργαλείου και απόπειρα να δοθεί κάποια ερμηνεία σχετικά με τα λάθη, στον βαθμό που αυτό είναι εφικτό. Επιπλέον, καταγράφονται τα συμπεράσματα που συνάγονται από τα ευρήματα σχετικά με τις ΚΛΠ και τα σημεία και τους τρόπους με τους οποίους θα μπορούσαν να βελτιωθούν οι επιδόσεις του εργαλείου. Σχολιάζονται πρώτα τα αποτελέσματα για τους φορητούς υπολογιστές και στη συνέχεια για τα εστιατόρια.

Στα laptops, η συγκεκριμένη εφαρμογή φάνηκε να λειτουργεί εξαιρετικά καλά σε σχέση με την ακρίβεια των αποτελεσμάτων της. Πιο συγκεκριμένα, μόνο σε τρεις περιπτώσεις αυτά τα οποία εντόπισε η μηχανή ήταν λανθασμένα και δεν θα έπρεπε να έχουν εντοπιστεί. Οι δύο από αυτές σχετίζονται με τη λειτουργία των τροπικών ρημάτων. Γενικά, ο λόγος που ονομάζονται έτσι είναι επειδή μπορούν να εκφράζουν πολλές τροπικότητες, μία από τις οποίες είναι και η κατευθυντικότητα. Έγινε προσπάθεια με τους κανόνες να περιοριστούν σε αυτή τους τη χρήση, όμως σε επίπεδο επιφανειακής σύνταξης αυτό δεν είναι πάντα εφικτό. Έτσι, στην πρόταση “*If you have to ask you don't own a Mac and your just not in the know.*”, το *have to* επισημαίνεται ως κατευθυντικό με Αντικείμενο το *ask*. Επειδή όμως βρίσκεται μέσα σε υποθετικό λόγο, δεν εκφράζει κατευθυντικότητα και δεν θα έπρεπε να έχει εντοπιστεί. Οι υποθετικές προτάσεις εν γένει δημιουργούν αρκετά προβλήματα στην ανάλυση άποψης και απασχολούν αρκετά την έρευνα (βλ. και παραπάνω 2.1.1.). Ως προς το εργαλείο θα μπορούσε να υπάρξει πρόβλεψη ώστε να μην εντοπίζονται τα στοιχεία που έχουν οριστεί ως κατευθυντικά, όταν αυτά βρίσκονται μέσα σε υποθετικές προτάσεις, προηγείται δηλαδή το *If*. Για να είναι απόλυτα επιτυχείς τέτοιοι κανόνες πρέπει να λαμβάνουν υπόψη τους εξαρτήσεις που μπορεί να είναι αρκετά μακρινές (long distance dependencies), δηλαδή τα στοιχεία τους να απέχουν πολύ μεταξύ τους. Εν προκειμένω, μπορεί να παρεμβάλλονται πολλά άλλα γλωσσικά στοιχεία ανάμεσα στον υποθετικό

σύνδεσμο και το τροπικό ή άλλο ρήμα, γεγονός που καθιστά τον εντοπισμό τους με κανόνες επιφανειακής σύνταξης πολύ δύσκολο.

Συγκριτικά με την ακρίβεια λίγα πράγματα μπορούν να ειπωθούν παραπάνω. Περισσότερα όμως χρειάζεται να παρατηρηθούν σχετικά με τα προβλήματα στην ανάκληση σωστών απαντήσεων, για τους λόγους δηλαδή για τους οποίους δεν εντοπίστηκαν οι ΚΛΠ που υπάρχουν στο σώμα προτάσεων. Επιχειρείται να γίνει μια κατηγοριοποίηση των δομών εκείνων που δηλώνουν κατευθυντικότητα, δεν αναγνωρίστηκαν όμως από τη γραμματική.

Ένα πολύ σημαντικό θέμα αποτελεί η προστακτική. Οι περισσότερες δομές που δεν κατάφερε το εργαλείο να εντοπίσει, περιείχαν προστακτική. Π.χ. *“Put a SSD and use a 21" LED screen, this set up is silky smooth!”*, *“Bottom line, if you can afford it, get a Mac!”*. Η έγκλιση αυτή στην αγγλική γλώσσα δεν έχει καμία μορφολογική ή άλλη διαφοροποίηση. Κατά συνέπεια είναι εξαιρετικά δύσκολο να την ξεχωρίσει μια γραμματική με κανόνες επιφανειακής δομής. Ένας τρόπος αντιμετώπισης θα ήταν πιθανόν, να οριστεί ένας κανόνας με τον οποίο τα ρήματα στην απλή τους μορφή – όχι γερούνδια και μετοχές – που βρίσκονται στην αρχή πρότασης να λαμβάνονται ως κατευθυντικά. Αυτό θα βοηθούσε να αναγνωριστούν κάποιες περιπτώσεις, όπως όμως φαίνεται και από τα παραδείγματα που παρατίθενται, όχι όλα, αφού πολλές φορές τέτοιες δομές βρίσκονται σε άλλη θέση της πρότασης.

Θα μπορούσε επίσης να προβλεφθεί ένας υποθετικός λόγος όπως στο δεύτερο παράδειγμα παραπάνω, του οποίου η απόδοση είναι σε προστακτική και αποτελεί ΚΛΠ. Επιπλέον, υπάρχει το σημείο στίξης, το κόμμα που χωρίζει την υπόθεση από την απόδοση και έτσι θα μπορούσε να καταγραφεί σαν σχήμα σε κανόνα. Ωστόσο, υπάρχουν πολλές προτάσεις στις οποίες ο συγγραφέας παραλείπει το σημείο στίξης, ιδιαίτερα στα είδη λόγου που εξετάζονται στα οποία δεν δίνεται μεγάλη προσοχή στη μορφή. Π.χ. *“if you watch dvds once in a while like i do get this”*. Άλλωστε πολλά από τα σώματα κειμένων που επεξεργάζονται μηχανές υφίστανται προ-επεξεργασία, κατά την οποία αφαιρούνται τα σημεία στίξης. Αυτό θα καθιστούσε έναν τέτοιο κανόνα τροχοπέδη για τη λειτουργία του εργαλείου, καθώς θα έφερνε πολλά λάθος αποτελέσματα.

Μια άλλη κατηγορία προτάσεων που δεν εντοπίστηκαν, ήταν αυτές που περιείχαν εκφραστικά μέσα τα οποία δεν εμφανίζονται στα σώματα κειμένων στα οποία στηρίχθηκε η ανάπτυξη του εργαλείου. Ίσως αυτό να έχει να κάνει με το διαφορετικό θεματικό πεδίο, σε κάθε περίπτωση όμως αποτελεί ένδειξη ότι ίσως θα

πρέπει να εξεταστούν περισσότερα κείμενα ώστε να εμπλουτιστούν τα γλωσσικά μέσα έκφρασης ΚΛΠ. Τέτοια παραδείγματα περιπτώσεων που δεν εντόπισε η εφαρμογή είναι: “*This laptop is a great buy.*”, “*not something you want to miss out on!!!*”, “*You will not regret buying this computer!*”.

Ακόμα, δεν αναγνωρίστηκαν περιπτώσεις στις οποίες, αν και υπάρχει κατευθυντικό στοιχείο, δεν υπάρχει το Αντικείμενό του, π.χ. “*Very Very Highly Recommended.*”. Εφόσον τα σώματα στα οποία έγινε η αξιολόγηση περιέχουν προτάσεις και όχι ολόκληρες κριτικές ή κείμενα, είναι πιθανό το Αντικείμενο να αναφέρεται σε προηγούμενο ή επόμενο σημείο. Όμως, το εργαλείο είναι σχεδιασμένο να λειτουργεί σε επίπεδο πρότασης, εντοπίζοντας υποχρεωτικά και ένα Αντικείμενο. Ίσως θα μπορούσε να γίνει μια αλλαγή στους κανόνες και να αναγνωρίζουν τουλάχιστον κάποια επίθετα ή μετοχές, έστω κι αν δεν έχουν Αντικείμενο. Βέβαια, αυτό θα πρέπει να ελεγχθεί σε κείμενα γιατί μπορεί να αποβεί εις βάρος της ακρίβειας. Στην ίδια περίπτωση emπίπτουν και θέματα που σχετίζονται με τη συν-αναφορά (co-reference), όταν δηλαδή το Αντικείμενο είναι μια αντωνυμία, συνηθέστερα αναφορική. Τέτοιο παράδειγμα είναι το: “*I did swap out the hard drive for a Samsung 830 SSD which I highly recommend.*”, όπου η αναφορά της αντωνυμίας είναι μια ολόκληρη πρόταση. Τα ζητήματα συν-αναφοράς ωστόσο δεν έχουν εύκολες λύσεις, πόσο μάλλον σε επιφανειακό επίπεδο.

Σε ορισμένες προτάσεις, ανάμεσα στο Υποκείμενο και την κατευθυντική έκφραση παρεμβάλλονται άλλα στοιχεία τα οποία μπορούν να επιτείνουν τη σημασία του, δεν έχουν όμως προβλεφθεί από το εργαλείο. Π.χ. “*I will not hesitate to recommend it to family and friends*”, “*The criticism has waned, and now I'd be the first to recommend an Air for truly portable computing.*”. Πρέπει να εξεταστεί κατά πόσο ένας κανόνας που θα προέβλεπε έναν αριθμό tokens μεταξύ Υποκειμένου και κατευθυντικού στοιχείου θα ωφελούσε περισσότερο από όσο θα έβλαπτε. Μήπως δηλαδή έφερνε τόσα λάθος αποτελέσματα που η συνεισφορά του στην ανάκληση να είναι ασήμαντη.

Σημαντικό είναι το θέμα της σύζευξης, όταν δύο ρήματα συνδέονται παρατακτικά και το ένα από αυτά εκφράζει ΚΛΠ, όπως: “*I love it and would recommend it to everyone who is tired of the constant attention PC's require.*”. Εδώ, και πάλι το ρήμα που δηλώνει κατευθυντικότητα είναι σχετικά απομακρυσμένο από το Υποκείμενό του. Υπάρχει όμως η σύζευξη που θα μπορούσε να βοηθήσει ώστε να αναγνωρισθεί.

Η γλώσσα που χρησιμοποιείται στα κοινωνικά δίκτυα και ευρύτερα στο διαδίκτυο, δεν είναι αρκετά προσεγμένη. Προσιδιάζει περισσότερο σε προφορικό λόγο παρά σε γραπτό, με αποτέλεσμα να παρατηρούνται πολλά λάθη. Μια συχνή περίπτωση είναι αυτή της παράλειψης του αντωνυμικού υποκειμένου στην αγγλική γλώσσα που η δήλωσή του είναι υποχρεωτική λόγω της έλλειψης μορφολογίας του ρηματικού συστήματος. Είτε λόγω λάθους, είτε εκ παραδρομής πολλές φορές δεν υπάρχει ρητά εκπεφρασμένο υποκείμενο με αποτέλεσμα να μην αναγνωρίζονται οι προτάσεις αυτές από τη μηχανή. Π.χ. *“10/10 five stars would recommend.”*. Η πρόταση αυτή δεν θα μπορούσε να αναγνωρισθεί από το εργαλείο, καθώς απουσιάζει τόσο το Υποκείμενο *I* όσο και το Αντικείμενο.

Και στην ανάκληση σημαντικό θέμα αποτελούν τα τροπικά ρήματα. Καθώς μπορούν να εκφράζουν πλήθος σημασιών, με την κατευθυντικότητα να είναι μία μόνο από αυτές, είναι βέβαιο ότι θα υπάρχουν περιπτώσεις που η μηχανή δεν μπορεί να αναγνωρίσει. Μερικές από αυτές είναι δύσκολο να χαρακτηριστούν από τους ίδιους τους ανθρώπους και πιθανό να υπάρχει διαφωνία μεταξύ τους, οπότε ένα εργαλείο αυτόματης αναγνώρισης δεν θα ήταν δυνατό να έχει καλύτερη τύχη. Στην πρόταση π.χ. *“You can read all about the details of this marvelous computer on wikipedia.”*, το τροπικό ρήμα *can* χρησιμοποιείται για να κάνει μια σύσταση στον αποδέκτη, παρόλο που η κύρια χρήση του είναι να εκφράζει ικανότητα ή δυνατότητα. Αυτός είναι και ο λόγος για τον οποίο δεν έχει συμπεριληφθεί στους κανόνες παρά μόνο στην μορφή της ερώτησης που εκφράζει αίτημα με ευγενικό τρόπο (*“Honey, can you help me with the pillows?”*). Μία πιθανή λύση για το συγκεκριμένο θέμα θα μπορούσε να λαμβάνει υπόψη τη σημασία του ρήματος που λειτουργεί ως Αντικείμενο. Ρήματα που φέρουν ορισμένα σημασιολογικά χαρακτηριστικά, εντάσσονται σε κάποια πλαίσια (frames), μπορούν να εκφράζουν ΚΛΠ, ενώ άλλα όχι.

Τέλος, ένα θέμα που δημιουργεί προβλήματα στην ορθή λειτουργία των κανόνων, είναι οι ετικέτες που παίρνουν οι λεκτικές μονάδες από τον POS Tagger. Σε αρκετές περιπτώσεις, αυτές είναι λανθασμένες με αποτέλεσμα να μην μπορούν να εφαρμοστούν σωστά οι κανόνες της γραμματικής. Έτσι, για παράδειγμα στην πρόταση: *“Though please note that sometimes it crashes, and the sound quality isnt superb.”* Και ενώ υπάρχει κανόνας που προβλέπει τη δομή “please + ρήμα”, το εργαλείο δεν αναγνωρίζει τίποτα, καθώς το *note* επισημειώνεται ως NN, δηλαδή ουσιαστικό αντί για ρήμα.

Στις κριτικές εστιατορίων, όπως φάνηκε από τα αριθμητικά αποτελέσματα, το ποσοστό της ανάκλησης δεν ήταν ιδιαίτερα υψηλό. Αυτό οφείλεται σε μεγάλο βαθμό στη χρήση της προστακτικής. Όπως προκύπτει από την ανάλυση λαθών, η συντριπτική πλειοψηφία των προτάσεων στις οποίες δεν επισημειώθηκε κάποια ΚΛΠ, περιείχε ρήμα σε προστακτική, συγκεκριμένα οι 42 από τις 63, δηλαδή σχεδόν το 67%. Γενικότερα, στο συγκεκριμένο σώμα προτάσεων η προστακτική δείχνει να είναι από τα συχνότερα γλωσσικά μέσα έκφρασης ΚΛΠ. Αυτό θα μπορούσε να σχετίζεται με το θεματικό πεδίο στο οποίο ανήκει, μένει όμως να ερευνηθεί περαιτέρω σε μεγαλύτερα κείμενα από αυτή τη θεματική. Μερικά παραδείγματα παρατίθενται εδώ: *“If you're in the neighborhood, definitely stop by for a great meal.”*, *“Do not go to this restaurant.”*, *“Give it a try, menu is typical French but varied.”*.

Οι υπόλοιπες από τις προτάσεις που δεν εντοπίστηκαν περιέχουν κάποια από τα στοιχεία που αναφέρθηκαν και παραπάνω σχετικά με τους φορητούς υπολογιστές. Οι υποθετικοί λόγοι είναι ένα από αυτά: *“If your looking for nasty high priced food with a dash of ghetto scenery cheap BX A\$\$ this is the place to be!!”*, *“If you like seafood and/or Greek food you will love this place though it is not limited to just these things.”*, *“if you're looking for authentic hong kong-style food, look no further.”*. Επίσης, τρόποι έκφρασης που δε βρέθηκαν στα σώματα κειμένων που εξετάστηκαν. *“Definitely something to come down and check out.”*, *“The food is just OKAY, and it's almost not worth going unless you're getting the pialla, which is the only dish that's really good.”*, *“Its good to go there for drinks if you don't want to get drunk because you'll be lucky if you can get one drink an hour the service is so bad.”*. Προτάσεις στις οποίες λείπει το Αντικείμενο εμφανίζονται και εδώ: *“Weekends can get crowded, but still highly recommended.”*. Μπορεί ακόμα να υπάρχουν γλωσσικά λάθη στη χρήση της γλώσσας, χαρακτηριστικό του είδους του λόγου που χρησιμοποιείται στο διαδίκτυο. Για παράδειγμα, ενώ στην αγγλική γλώσσα το αντικείμενο πάντα έπεται του ρήματος ώστε να ξεχωρίζει από το υποκείμενο, καθώς δεν υπάρχουν μορφολογικά διακριτές πτώσεις, στην πρόταση *“The food is good, the teriyaki I recommend.”* το Αντικείμενο της ΚΛΠ που εκφράζεται με το *recommend*, προηγείται του ρήματος. Αυτή είναι μια δομή που δεν μπορεί να αντιμετωπιστεί με κανόνες επιφανειακής σύνταξης. Και σε αυτό το σώμα προτάσεων εμφανίστηκαν περιπτώσεις που λείπει το Υποκείμενο: *“Would not recommend anyone to go.”*, *“I love the Little Pie Company as much as anyone else who has written reviews, but must discourage anyone from visiting the Grand Central location due to their RUDE service from two sales people.”*. Τέλος,

εμφανίστηκαν και περιπτώσεις στις οποίες για κάποιον λόγο ο POS Tagger δεν αναγνωρίζει σωστά κάποιες λέξεις, οπότε με τη σειρά της η γραμματική αδυνατεί να εφαρμόσει κάποιον κανόνα: “*We requested they re-slice the sushi, and it was returned to us in small cheese-like cubes.*” Εδώ το *re-slice* δεν επισημειώνεται ως ρήμα αλλά ως ουσιαστικό με αποτέλεσμα να μην λειτουργεί ο αντίστοιχος κανόνας της γραμματικής για το *request*.

Ως γενική σύνοψη θα μπορούσε να ειπωθεί ότι η γραμματική που δημιουργήθηκε, λειτουργεί πολύ καλά σε επίπεδο ακρίβειας, όχι όμως και σε επίπεδο ανάκλησης αποτελεσμάτων. Παραπάνω καταγράφηκαν αναλυτικά ποια ήταν τα θέματα που προέκυψαν ως προς την ανάκληση, αλλά και πιθανές λύσεις που θα μπορούσαν να εφαρμοστούν σε κάθε περίπτωση προκειμένου να αυξηθεί το συγκεκριμένο ποσοστό. Στόχος για τη βελτίωση του εργαλείου είναι να επιτευχθεί μια καλύτερη ισορροπία ανάμεσα στις δύο αυτές μετρικές. Υπάρχουν δηλαδή ορισμένες μετατροπές που θα βελτίωναν αρκετά την ανάκληση, αφαιρώντας όμως κάτι από την ακρίβεια. Αυτός άλλωστε είναι και ο στόχος κάθε συστήματος, η ισορροπία, αφού είναι κατανοητό ότι ένα σύστημα με πεπερασμένες δυνατότητες αντίληψης δεν μπορεί να έχει 100% επιτυχία, τη στιγμή μάλιστα που ο ίδιος ο άνθρωπος είναι δύσκολο να το επιτύχει.

5. Συμπεράσματα – Περαιτέρω έρευνα

5.1. Συμπεράσματα

Στην παρούσα εργασία παρουσιάστηκε μια μέθοδος ανάλυσης των κατευθυντικών λεκτικών πράξεων στην αγγλική γλώσσα μέσα από κείμενα προερχόμενα από ηλεκτρονικές πηγές. Για τον σκοπό αυτόν δημιουργήθηκε ένα εργαλείο με στόχο τον εντοπισμό των δομών αυτών καθώς και των συστατικών τους στοιχείων. Η βασική διαφοροποίηση της προσπάθειας που περιγράφηκε εδώ σε σχέση με την έρευνα που έχει μέχρι στιγμής πραγματοποιηθεί στο συγκεκριμένο πεδίο, είναι ότι χρησιμοποιείται γλωσσικό υλικό που έχει παραχθεί από τους ίδιους τους χρήστες του διαδικτύου (user generated content) και μέσα σε αυτό επιχειρείται να εντοπιστούν συστάσεις (recommendations) και αιτήματα (requests), ανεξάρτητα με το θεματικό πεδίο (domain independent). Η προσέγγιση που υιοθετήθηκε εντάσσεται μέσα στα πλαίσια της έρευνας για την ανάλυση συναισθήματος και αξιοποιεί τη θεωρία των λεκτικών πράξεων. Τέλος, η μεθοδολογία που ακολουθήθηκε, στηρίζεται και καθοδηγείται από τα ίδια τα γλωσσικά δεδομένα (data driven) και στοχεύει κατά κύριο λόγο στην ακρίβεια των αποτελεσμάτων (precision oriented).

Σε πρώτη φάση εξετάστηκαν αυθεντικά σώματα κειμένων από ηλεκτρονικές πηγές (twitter, fora). Αφού εντοπίστηκαν τα βασικά γλωσσικά τους χαρακτηριστικά, εντοπίστηκαν οι προτάσεις που περιέχουν ΚΛΠ και οι τρόποι με τους οποίους αυτές εκφράζονται γλωσσικά. Τα κείμενα επισημειώθηκαν και ως προς άλλα στοιχεία, δηλαδή αν δηλώνουν υποκειμενικότητα ή αντικειμενικότητα, αν η πολικότητά τους είναι θετική ή αρνητική και αν ο συγγραφέας εκφράζει θετική κρίση (praise) απέναντι σε ένα πρόσωπο ή προϊόν, ή αρνητική (judgment). Τα αποτελέσματα αυτά δεν αξιοποιούνται πρακτικά από την εφαρμογή για τις ΚΛΠ, αφήνουν όμως περιθώρια πιο στοχευμένης μελλοντικής εργασίας.

Στη συνέχεια, τα γλωσσικά μέσα έκφρασης ΚΛΠ καταγράφηκαν σε δύο λεξικά, ένα αποκλειστικά με ρήματα και ένα με διάφορες εκφράσεις. Με βάση τα λεξικά αυτά και τη συντακτική συμπεριφορά των περιεχομένων τους, δημιουργήθηκαν κανόνες οι οποίοι αποτέλεσαν τη γραμματική των ΚΛΠ (Directives Grammar). Οι κανόνες αυτοί λειτουργούν σε επιφανειακό επίπεδο σύνταξης και επιχειρούν να εντοπίσουν τόσο τα μέσα με τα οποία εκφράζεται λεκτικά η κατευθυντικότητα, όσο και το αντικείμενο κάθε τέτοιας λεκτικής πράξης, είτε αυτό είναι μια συγκεκριμένη ονομαστική φράση, είτε η

ενέργεια που δηλώνεται από έναν ρηματικό τύπο. Επίσης, εντοπίζεται και η άρνηση, καθώς αυτή αντιστρέφει το νόημα της πρότασης και από σύσταση μπορεί να μετατραπεί σε αποτροπή.

Η εφαρμογή της γραμματικής έγινε με τη βοήθεια της πλατφόρμας επεξεργασίας φυσικής γλώσσας GATE. Αρχικά, δημιουργήθηκε μια αλυσίδα προεπεξεργασίας στην οποία επιβλήθηκαν τα κείμενα, ώστε να γίνει χωρισμός προτάσεων και μορφολογική ανάλυση των γλωσσικών δεδομένων, και στο αποτέλεσμα αυτής της αλυσίδας εφαρμόστηκε η γραμματική των ΚΛΠ, δηλαδή τα λεξικά και η γραμματική.

Η αξιολόγηση της εφαρμογής έγινε σε δύο σώματα 800 προτάσεων το καθένα από διαφορετικά θεματικά πεδία (φορητοί υπολογιστές, εστιατόρια) σε σχέση με αυτά που χρησιμοποιήθηκαν για την ανάπτυξη του εργαλείου. Τα αποτελέσματα είναι ιδιαίτερα ικανοποιητικά σε επίπεδο ακρίβειας, που ήταν και ο βασικός στόχος. Σχεδόν όλες οι δομές που εντοπίστηκαν ήταν σωστές. Ωστόσο, σε επίπεδο ανάκλησης, υπάρχουν περιθώρια βελτίωσης προκειμένου το εργαλείο να είναι σε θέση να φέρνει περισσότερα αποτελέσματα.

5.2. Περαιτέρω έρευνα – Εφαρμογές

Αρχικά, θα βοηθούσε σημαντικά στη βελτίωση του συγκεκριμένου εργαλείου ο εμπλουτισμός των σωμάτων κειμένων που χρησιμοποιήθηκαν για την ανάπτυξή του. Η χρήση περισσότερων και πιο εκτεταμένων συλλογών γλωσσικού υλικού, θα μπορούσε να αναδείξει και άλλους τρόπους έκφρασης ΚΛΠ που δεν εμφανίστηκαν σε όσα εξετάστηκαν. Με τον τρόπο αυτόν θα εμπλουτίζονταν τόσο τα λεξικά όσο και οι κανόνες. Ακόμα, θα ήταν ίσως χρήσιμο, εφόσον η εφαρμογή δημιουργήθηκε έτσι ώστε να μην εξαρτάται από συγκεκριμένο θεματικό πεδίο, να εξεταστούν κείμενα από ακόμα περισσότερες θεματικές και να διευρυνθεί έτσι η αποτελεσματικότητά της. Το ίδιο θα μπορούσε να γίνει και με τα σώματα στα οποία αξιολογήθηκε η γραμματική. Χρησιμοποιώντας περισσότερα και από περισσότερα domain, θα υπήρχαν πιο πολλές εμφανίσεις ΚΛΠ και θα βελτιωνόταν η αξιοπιστία των αποτελεσμάτων της.

Η αποτελεσματικότητα του εργαλείου, και ιδίως στο επίπεδο της ανάκλησης, θα ήταν δυνατό να βελτιωθεί με τεχνικές μηχανικής μάθησης, όπως η μάθηση με επίβλεψη (supervised learning), στην οποία το σύστημα καλείται να «μάθει» μια έννοια μέσα από ένα σύνολο δεδομένων, και η μάθηση χωρίς επίβλεψη (unsupervised learning), σύμφωνα με την οποία το σύστημα καλείται μόνο του να ανακαλύψει έννοιες

και συσχετίσεις σε ένα σύνολο δεδομένων, δημιουργώντας πρότυπα. Επιπλέον, η αποτελεσματικότητα σε επίπεδο ανάκλησης θα μπορούσε να βελτιωθεί εφαρμόζοντας τρόπους αντιμετώπισης των ζητημάτων της συν-αναφοράς (co-reference) και των εξαρτήσεων μακρινής απόστασης (long-distance dependencies).

Ένα άλλο κομμάτι που προσφέρεται ιδιαίτερα για περαιτέρω έρευνα είναι αυτό των σημασιολογικών κατηγοριών. Εδώ διακρίθηκαν δύο βασικές ομάδες με βάση αποκλειστικά την κατεύθυνση της ωφέλειας που προκύπτει μέσα από τη λεκτική πράξη. Ωστόσο, αναφέρθηκε ήδη ότι η διάκριση αυτή είναι αδρομερής και κάθε κατηγορία θα μπορούσε να χωριστεί σε μικρότερες, εξειδικεύοντας περισσότερο το είδος της κατευθυντικότητας που κάθε φορά εκφράζεται. Αυτή η λεπτομερής ανάλυση έχει νόημα να γίνει σε έρευνες με πιο ειδική στόχευση, λόγου χάρη αν επιδιώκεται η ανάπτυξη ενός εργαλείου που θα λειτουργεί σε συγκεκριμένη θεματική περιοχή.

Πρέπει επίσης να σημειωθεί ότι θεωρείται σκόπιμη η ξεχωριστή εξέταση ορισμένων υποκατηγοριών των ΚΛΠ. Αυτές είναι κυρίως οι ευχές. Η ευχή ως τροπικότητα εκφράζεται με διάφορους τρόπους και κυρίως με συγκεκριμένες δομές, για τα αγγλικά π.χ. *I hope, I wish, If only, hopefully*, καθώς και με προστακτική. Τα corpora που εξετάστηκαν βρίθουν από παρόμοια παραδείγματα: “@spamtan @trishuson @noidauson I hope you find her :(#pray”, “if only it played stand alone avi files”, “i wish i would have a real tragic love affair and get so bummed out that i d just quit my job and become a bum for a few years because i was thinking about doing that anyway”, “#prayforboston stay safe guys, please. Enough people are hurting for today. We don't want anymore :(”. Οι περιπτώσεις αυτές δεν είναι πάντα προφανείς, ιδίως ως προς την πολικότητα που εκφράζουν. Πολλές φορές εκφράζει ο συγγραφέας τους μια ελπίδα του ή εύχεται σε κάποιον να λυθεί κάποιο πρόβλημά του, εξωτερικεύοντας τη συμπάθεια ή τη συμπόνια του που μπορούν να θεωρηθούν θετικά συναισθήματα. Όταν ωστόσο εκδηλώνονται για κάποιο αρνητικό γεγονός ή εύχεται ο ομιλητής να του συμβεί κάτι που είναι κακό, όπως σε κάποια από τα παραδείγματα, τότε δεν είναι τόσο εύκολη η απόφαση αν το polarity είναι θετικό ή αρνητικό. Το ίδιο ισχύει και για την επίκληση του θείου, που είναι πολύ συχνή σε κοινωνικά δίκτυα, τουλάχιστον της αγγλικής γλώσσας που εξετάστηκαν. Π.χ. “#PrayForBoston Holy Father in Heaven, be with the injured, the first responders and their families. I pray those responsible may be found.”, “#prayforboston praying for everyone's families :(:(”. Γενικότερα το θέμα των ευχών έχει απασχολήσει την έρευνα και έχουν γίνει αρκετές μελέτες σχετικές με το ζήτημα, με πρώτη εκείνη των Goldberg et al. (2009),

όπου επιχειρείται ο εντοπισμός τους σε σώματα κειμένων προερχόμενα από πολιτικά σχόλια και κριτικές προϊόντων. Σημαντική είναι η συμβολή των Ramanand et al. (2010), οι οποίοι ενδιαφέρονται για συγκεκριμένο τύπο ευχών, ειδικότερα για προτάσεις που προτείνουν τη βελτίωση προϊόντων και προτάσεις που εκφράζουν την ευχή του συγγραφέα να αγοράσει κάποιο προϊόν. Με το ίδιο θέμα ασχολήθηκαν και οι Wu & He (2011), χρησιμοποιώντας διαφορετική μεθοδολογία.

Ένα άλλο πεδίο στο οποίο θα μπορούσε να επεκταθεί η έρευνα για τη μελέτη των ΚΛΠ, είναι εκείνο του Αντικειμένου τους. Ειδικότερα, θα είχε ενδιαφέρον να εξεταστούν τα σημασιολογικά στοιχεία του Αντικειμένου τους, κυρίως στις περιπτώσεις που αυτό είναι ένας ρηματικός τύπος. Με τον τρόπο αυτόν θα μπορούσαν να προκύψουν συμπεράσματα όσον αφορά στο τι προτείνει ή ζητά ο ομιλητής από τον Αποδέκτη. Για παράδειγμα, συστήνει να αγοράσει ένα προϊόν ή να επισκεφθεί ένα ξενοδοχείο, ζητά να γίνουν βελτιώσεις σε κάποιο προϊόν, συμβουλεύει τη διακοπή μιας συνήθειας και πολλά άλλα.

Μεγάλο ενδιαφέρον παρουσιάζει η ανάπτυξη μιας εφαρμογής που θα εντοπίζει συστάσεις και αιτήματα σε κείμενα στην ελληνική γλώσσα. Η μεθοδολογία μπορεί να είναι παρόμοια, θα πρέπει όμως να δημιουργηθούν από την αρχή γλωσσικοί πόροι και κανόνες μοντελοποίησής τους.

Συνοψίζοντας, θα μπορούσε να ειπωθεί ότι η εξέταση των ΚΛΠ προσφέρεται για γλωσσολογικές αναλύσεις διαφόρων ειδών, οι οποίες συμβάλλουν στην καλύτερη κατανόηση του φαινομένου. Επιπλέον όμως, μπορεί να βρει ιδιαίτερα πλούσιο πεδίο πρακτικών εφαρμογών, τόσο από εταιρείες που επιθυμούν να βελτιώσουν τις στρατηγικές προώθησης και πωλήσεων των προϊόντων και των υπηρεσιών τους, όσο και από κυβερνήσεις που επιδιώκουν να σφυγμομετρήσουν την κοινή γνώμη αλλά και να προωθήσουν την ασφάλεια των χωρών τους, για παράδειγμα από πιθανές τρομοκρατικές ενέργειες. Γενικότερα, στα πλαίσια της ασφάλειας στο διαδίκτυο θα μπορούσε να χρησιμοποιηθεί από υπηρεσίες δίωξης του ηλεκτρονικού εγκλήματος προκειμένου να εντοπίσουν και να αποτρέψουν εγκληματικές ενέργειες, αλλά και από τους γονείς ώστε να προστατεύσουν τα παιδιά τους, κυρίως στις επικοινωνίες τους με τρίτους. Η εφαρμογή που παρουσιάστηκε εδώ αποτελεί ένα χρηστικό όσο και γενικό εργαλείο για την ανάλυση των ΚΛΠ. Ανάλογα με τους ειδικότερους στόχους που κάθε φορά τίθενται, είναι δυνατή η περαιτέρω προσαρμογή και επέκτασή του.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Appling, D.S., Briscoe, E.J., Hayes, H., Mappus, R.L. (2013). *Towards Automated Personality Identification using Speech Acts*. In *Proceedings of the Conference on Artificial Intelligence (AAAI)*: AAAI Press.
- Austin, J. L. (1962). *How to do things with words*. Cambridge: Harvard University Press.
- Bach, K. & Harnish R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge MA: The MIT Press.
- Balabanović, M., Shoham, Y. (1997). Content-based, Collaborative Recommendation. In *Communications of the ACM* 40 (3). 66-72.
- Γεωργακοπούλου, Α. και Γούτσος, Δ. (1999). *Κείμενο και επικοινωνία*. Αθήνα: Ελληνικά Γράμματα.
- Carlos, C. S., Yalamanchi, M. (2012). Intention Analysis for Sales, Marketing and Customer Service. In *COLING 2012. The 24th International Conference on Computational Linguistics*.
- Cohen, W. W., Carvalho, V. R., Mitchell, T. M. (2010). Learning to classify email into speech acts. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg, PA, USA: Association for Computational Linguistics. 309-316.
- Ding, X., Liu, B., Yu, P. S. (2008). A Holistic Lexicon-Based Approach to Opinion Mining. In *Proceedings of the International Conference on Web Search and Web Data Mining (WSDM '08)*. 231-240.
- Ganapathibhotla, M., Liu, B. (2008). Mining Opinions in Comparative Sentences. In *Proceedings of International Conference on Computational Linguistics*. 241-248.
- Goldberg, A. B., Fillmore, N., Andrzejewski, D., Xu, Z., Gibson, B., Zhu, X. (2009). May all your wishes come true: a study of wishes and how to recognize them. In *NAACL 09 Proceedings of Human Language Technologies*. Stroudsburg, PA, USA: Association for Computational Linguistics. 263-271.
- Hu, M., Liu, B. (2004). Mining and Summarizing Customer Reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 168-177.

- Jindal, N., Liu, B. (2006). Mining Comparative Sentences and Relations. In *Proceedings of the 21st Conference on Artificial Intelligence (AAAI-06)*: AAAI Press. 1331-1336.
- Kushal, D., Lawrence, S., Pennock, D. M. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In *Proceedings of the 12th international conference on World Wide Web*. 519-528.
- Lampert, A., Dale, R., Paris, C. (2008a). The Nature of Requests and Commitments in Email Messages. In *Proceedings of the AAAI Workshop on Enhanced Messaging*. Chicago, USA: AAAI Press. 42-47.
- Lampert, A., Dale, R., Paris, C. (2008b). Requests and Commitments in Email are More Complex Than You Think: Eight Reasons to be Cautious. In *Proceedings of Australian Language Technology Workshop (ALTA 2008)*. Hobart, Australia. 55-63.
- Lampert, A., Dale, R., Paris, C. (2010). Detecting Emails Containing Requests for Action. In *HLT '10 Human Language Technologies*. Stroudsburg, PA, USA: Association for Computational Linguistics. 984-992.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. San Rafael, Calif.: Morgan & Claypool.
- Mooney, R. J., Bennett, P. N., Roy, L. (1998). In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*. AAAI Press. 70-74.
- Morgan, J. L. (1978). Two types of convention in indirect speech acts. In Cole, P. (ed.) *Syntax and Semantics. Vol. 9: Pragmatics*. New York: Academic Press. 261-280.
- Narayanan, R., Liu, B., Choudhary, A. (2009). Sentiment Analysis of Conditional Sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics. 180-189.
- Natt och Dag, J., Regnell, B., Gervasi, V., Brinkkemper, S. 2005. A linguistic-engineering approach to large-scale requirements management. In *Managing Language Requirements in Large-Scale Software Development*. Vol. 22. 32-39.
- Pang, B., Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In *Proceedings of the ACL*. 271-278.

- Pontiki, M., Papageorgiou, H. (2014). “There’s no way I would ever buy any mp3 player with a measly 4gb of storage”: Mining Intention Insights about Future Actions.
- Qadir, A. and Riloff, E. (2011). Classifying sentences as speech acts in message board posts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11)*. Stroudsburg, PA, USA: Association for Computational Linguistics. 748-758.
- Ramanand, J., Bhavsar, K. Pedanekar, N. (2010). *Wishful thinking*: finding suggestions and 'buy' wishes from product reviews. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text (CAAGET '10)*. Stroudsburg, PA, USA: Association for Computational Linguistics. 54-61.
- Searle, J. R. (1965). What is a Speech Act? In M. Black (ed.), *Philosophy in America*. London: Allen and Unwin, 221-239.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
- Searle, J. R. (1975a). A Taxonomy of Illocutionary Acts. In K. Gunderson (ed.), *Language, Mind and Knowledge*. University of Minnesota Press, 344-369.
- Searle, J. R. (1975b). Indirect speech acts. In Cole, P. & J. L. Morgan (eds.) *Syntax and Semantics 3: Speech Acts*. New York: Academic Press. 59-82.
- Searle, J. R. (1976). A classification of illocutionary acts. In *Language in Society* 5: 1-23.
- Searle, J. R. (1979). *Expression and Meaning*. Cambridge: Cambridge University Press.
- Terveen, L. G., Hill, W. C., Amento, B., McDonald, D. W., Creter, J. (1997). Phoaks: A System for Sharing Recommendations. In *Communications of the ACM* 40 (3). 59-62.
- Φιλίππακη-Warburton, E. (1992). *Εισαγωγή στη θεωρητική γλωσσολογία*. Αθήνα: Νεφέλη.
- Verschueren, J. (1980). *On Speech Act Verbs*. Amsterdam: John Benjamins.
- Wu, X., He, Z. (2011). Identifying Wish Sentence in Product Reviews. In *Journal of Computational Information Systems* 7:5. 1607-1613.

ΠΑΡΑΡΤΗΜΑ Ι – ΛΕΞΙΚΟ ΡΗΜΑΤΩΝ

Εγγραφή	Σημασιολογική κατηγορία	Συντακτική κατηγορία
adjure	request	VB1
admonish	recommendation	VB1
advise	recommendation	VB2
advocate	recommendation	VB3
allow	recommendation	VB1
appeal	request	VB4
ask	request	VB5
ask for	request	VB1
authorize/authorise	recommendation	VB6
badger	request	VB7
ban	recommendation negated	VB8
bar	recommendation negated	VB8
beg	request	VB9
beseech	request	VB10
bid	recommendation	VB10
brainwash	request	VB7
browbeat	recommendation	VB11
cajole	request	VB12
call on	request	VB10
call upon	request	VB10
can	request	VB22
caution	recommendation	VB13
claim	request	VB14
coax	request	VB7
coerce	recommendation	VB7
command	recommendation	VB10
commission	recommendation	VB1
compel	recommendation	VB1

convince	request	VB1
could	recommendation/request*	VB22
counsel	recommendation	VB1
dare	recommendation	VB1
decree	recommendation	VB15
defy	recommendation	VB1
demand	recommendation	VB3
deter	recommendation negated	VB8
dictate	recommendation	VB15
direct	recommendation	VB10
discourage	recommendation negated	VB8
dissuade	recommendation negated	VB8
drive	recommendation	VB1
encourage	recommendation	VB10
enjoin	recommendation	VB16
entice	request	VB7
entreat	request	VB1
exhort	recommendation	VB1
expect	recommendation	VB10
forbid	recommendation negated	VB16
galvanize	recommendation	VB11
goad	recommendation	VB11
have to	recommendation	VB23
hinder	recommendation negated	VB17
implore	request	VB24
importune	request	VB5
incite	recommendation	VB6
induce	recommendation	VB1
inhibit	recommendation negated	VB8
insist	recommendation	VB15
insist on	recommendation	VB18
instruct	recommendation	VB10

inveigle	request	VB7
invite	request	VB1
lead	recommendation	VB1
lobby	recommendation	VB1
mandate	recommendation	VB10
motivate	recommendation	VB10
must	recommendation	VB23
need	request	VB1
need to	recommendation	VB23
ought to	recommendation	VB23
permit	recommendation	VB1
persuade	recommendation	VB19
pester	request	VB1
plead	request	VB9
preclude	recommendation negated	VB17
prescribe	recommendation	VB6
press	recommendation	VB5
pressure	recommendation	VB7
prevent	recommendation negated	VB17
prod	recommendation	VB7
prohibit	recommendation negated	VB17
prompt	recommendation	VB1
propose	recommendation	VB3
proscribe	recommendation negated	VB17
push	recommendation	VB7
recommend	recommendation	VB3
request	request	VB3
require	recommendation	VB3
should	recommendation	VB23
spur	recommendation	VB20
suggest	recommendation	VB21
supplicate	request	VB1

tempt	recommendation	VB7
thwart	recommendation negated	VB8
urge	recommendation	VB10
want	request	VB1
warn	recommendation	VB13
will	request	VB22
wheedle	recommendation	VB11
would	request	VB22

ΠΑΡΑΡΤΗΜΑ ΙΙ – ΛΕΞΙΚΟ ΕΚΦΡΑΣΕΩΝ

Εγγραφή	Σημασιολογική κατηγορία	Συντακτική κατηγορία
how about	recommendation	Expr2
I'll thank you to	request	Expr1
put the squeeze on you to	request	Expr1
would do better to	recommendation	Expr1
I would like PRP to	request	Expr1
would you mind	request	Expr3
better	recommendation	Expr1
please	request	Expr1
why don't you	recommendation	Expr1
would PRP like to	recommendation	Expr1
what/all I advocate/propose/recommend/suggest	recommendation	Expr4
what/all I beg	request	Expr4
what/all I encourage	recommendation	Expr5
what/all I advise/require	recommendation	Expr6
what/all I expect	recommendation	Expr7
what/all I need/want	request	Expr7
what/all I ask	request	Expr8
advice/proposal/recommendation/suggestion is	recommendation	Expr9
be/make sure	recommendation	Expr9
(my) (only) request is	request	Expr10