

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΦΙΛΟΣΟΦΙΚΗ ΣΧΟΛΗ- ΤΜΗΜΑ ΦΙΛΟΛΟΓΙΑΣ

ΤΟΜΕΑΣ ΓΛΩΣΣΟΛΟΓΙΑΣ

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Διεπιστημονικό–Διαπανεπιστημιακό Πρόγραμμα

Μεταπτυχιακών Σπουδών «ΤΕΧΝΟΓΛΩΣΣΙΑ»

Σοφία Μαλτέζου

Εξόρυξη Γνώμης: Δημιουργία Ελληνικού Λεξικού Πόρου

Διπλωματική Εργασία

ΑΘΗΝΑ 2014

Φύλλο Εξέτασης

Περίληψη

Η παρούσα ερευνητική εργασία περιγράφει τη διαδικασία δημιουργίας ενός ελληνικού λεξικού πόρου που μπορεί να αξιοποιηθεί σε συστήματα *Εξόρυξης Γνώμης* (*Opinion Mining*). Βασικός στόχος αυτής της διαδικασίας ήταν η συγκέντρωση και κατηγοριοποίηση μιας ομάδας *υποκειμενικών όρων* (*subjectivity clues*), κυρίως επιθέτων και επιρρημάτων, οι οποίοι χρησιμοποιούνται κατά την εκφορά *αξιολογικού λόγου* σε ένα κείμενο. Το εγχείρημα αυτό εστίασε περισσότερο στη χρήση του αξιολογικού λόγου ως γλωσσικού μέσου έκφρασης *στάσης* (*Attitude*) προς κάποιο αντικείμενο αξιολόγησης, υπό το πρίσμα της θεωρίας *Appraisal*, και κυρίως των όρων *κριτική* (*Judgment*), η οποία απευθύνεται προς πρόσωπα, και *αξιολόγηση* (*Appreciation*), η οποία έχει ως αντικείμενο αναφοράς πράγματα, γεγονότα, φαινόμενα κ.λ.π.

Η μεθοδολογική προσέγγιση που ακολουθήθηκε στηρίζεται στην επεξεργασία πραγματικών δεδομένων μέσω της χρήσης σωμάτων κειμένων προκειμένου να διερευνηθεί ο βαθμός υποκειμενικότητας (*subjectivity*) καθώς και ο σημασιολογικός προσανατολισμός (*polarity*) κάθε όρου σε σχέση με το κειμενικό του περιβάλλον.

Τα κύρια στάδια σύνθεσης του λεξικού *υποκειμενικών όρων* περιλαμβάνουν την απόδοση στα ελληνικά μιας επιλεγμένης λίστας αγγλικών επιθέτων με προσδιορισμένη τιμή πολικότητας, τον έλεγχο της σταθερότητας αυτής της τιμής μέσω της διερεύνησης του *αξιολογικού δυναμικού* των επιθέτων σε σχέση με το συμφραστικό τους πλαίσιο, την επέκταση της αρχικής λίστας με προσθήκη νέων όρων και την κατηγοριοποίησή τους βάσει συγκεκριμένων σημασιολογικών και συντακτικών κριτηρίων.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Χάρη Παπαγεωργίου, για την επιστημονική καθοδήγηση και τη ψυχολογική στήριξη κατά τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας.

Επίσης, θα ήθελα να απευθύνω θερμές ευχαριστίες σε όλους τους συμφοιτητές μου, της «Τεχνολογισίας VI», για τις ουσιαστικές και γόνιμες σχέσεις συνεργασίας που αναπτύξαμε καθ'όλη τη διάρκεια φοίτησής μας.

Επίσης, ευχαριστώ όλα τα μέλη της εξεταστικής επιτροπής για τις παρατηρήσεις και τα σχόλιά τους για την εργασία μου.

Πίνακας περιεχομένων

Κεφάλαιο 1: Εισαγωγή	6
Κεφάλαιο 2: Θεωρητικό Υπόβαθρο	8
2.1. Sentiment Analysis	8
2.2. Σχετική Έρευνα – Λεξικά.....	10
2.2.1. SentiWordNet.....	12
2.2.2. Linguistic Inquiry and Word Count	13
2.2.3. MPQA	15
2.2.4. Bing Liu's OpinionLexicon	16
2.2.5. General Inquirer.....	16
2.2.6. ANEW	16
Κεφάλαιο 3: Ελληνικό Λεξικό Υποκειμενικών Όρων	17
3.1. Σκοπός	17
3.2. Θεωρία Appraisal	18
3.2.1. Attitude.....	19
3.2.1.1 Συναισθηματική Κατάσταση.....	19
3.2.1.2 Κριτική Προσώπων	20
3.2.1.3. Εκτίμηση	20
3.3.Μεθοδολογία	21
3.4. Υλοποίηση.....	24
3.4.1. Μετάφραση Αγγλικών Όρων-Φορέων Υποκειμενικότητας.....	24
3.4.2. Αξιολόγηση Υποκειμενικών Όρων	25
3.4.2.1. Ισχύς Υποκειμενικότητας.....	27
3.4.2.2. Προσδιορισμός Σημασιολογικού Προσανατολισμού	27
3.4.2.3.Affective State	28
3.4.3. Επέκταση Λεξικού	29
3.4.3.1. Προσθήκη Επιρρημάτων	29
3.4.4. Κατηγοριοποίηση Όρων	33
Κεφάλαιο 4: Συμπεράσματα	39
Κεφάλαιο 5: Συζήτηση-Μελλοντικές Προεκτάσεις	40
Βιβλιογραφία	42

1. Εισαγωγή

Τα τελευταία χρόνια έχει παρατηρηθεί η εκδήλωση ενός ραγδαία αυξανόμενου ερευνητικού ενδιαφέροντος για τη μελέτη του πεδίου της *Εξόρυξης Γνώμης* (*Opinion Mining*) ή αλλιώς *Ανάλυσης Συναισθήματος* (*Sentiment Analysis*). Και οι δύο όροι χρησιμοποιούνται εναλλακτικά στη βιβλιογραφία για να δηλώσουν ένα ευρύτατο φάσμα ζητημάτων του πεδίου της Επεξεργασίας Φυσικής Γλώσσας (*Natural Language Processing*), που σχετίζονται σε γενικές γραμμές με την αναζήτηση και τον προσδιορισμό των θετικών και αρνητικών απόψεων, στάσεων και συναισθημάτων που εκφράζονται σε ένα κείμενο (Wilson et al., 2005). Τα περισσότερα συστήματα *εξόρυξης γνώμης*, είτε πραγματοποιούν αυτήν την αναζήτηση και κατηγοριοποίηση της γνώμης σε επίπεδο κειμένου (*document level*) και προτάσεων (*sentence level*) είτε σε επίπεδο των λέξεων και των οντοτήτων που αναπαριστούν (*entity and feature level*), δίνουν ιδιαίτερη έμφαση στην ανάπτυξη λεξικών πόρων (*lexical resources*), που ονομάζονται συνήθως *sentiment lexicon* και *opinion lexicon* (Liu, 2012). Τέτοιου είδους λεξικοί πόροι, όπως π.χ. το Sentiwordnet (Baccianella, Esuli, Sebastiani, 2010) περιέχουν *opinion words* ή *sentiment words*, λέξεις, δηλαδή, και φράσεις που χρησιμοποιούνται συνήθως για την έκφραση θετικής ή αρνητικής άποψης και συναισθημάτων, οι οποίες διακρίνονται αντίστοιχα σε *θετικές* (π.χ. *όμορφος, υπέροχος, συναρπαστικός*) και *αρνητικές* (π.χ. *άσχημος, κακός, απαίσιος*).

Η ανάπτυξη λεξικών πόρων συνιστά μια από τις πιο τυπικές προσεγγίσεις κατά την ενασχόληση με το πεδίο της *εξόρυξης γνώμης*. Πολυάριθμα λεξικά, ποικίλων μεθοδολογικών προσεγγίσεων και εφαρμογών έχουν αναπτυχθεί προκειμένου να αξιοποιηθούν σε εφαρμογές ανάλυσης συναισθήματος (π.χ. (*General Inquirer*, Stone et al., 1966, *MPQA subjectivity lexicon* Wilson, Wiebe and Hoffmann, 2005, *SentiWordNet 3.0*, Baccianella, Esuli and Sebastiani, 2010, κ.α.) Η επικρατέστερη πρακτική που ακολουθείται είναι η συλλογή- με αυτοματοποιημένο συνήθως τρόπο- λέξεων και εκφράσεων με θετικό ή αρνητικό σημασιολογικό φορτίο και η συνακόλουθη επισημείωσή τους βάσει της αρχικής τους σημασίας, θετικής ή αρνητικής. Ωστόσο, αυτή η πρακτική δεν είναι πάντα ασφαλής, καθώς είναι σύνηθες το φαινόμενο να εμφανίζονται σε ένα ουδέτερο κείμενο αρκετές λέξεις θετικά ή αρνητικά φορτισμένες. Π.χ. στην πρόταση *Το Α αποτελεί ικανή και αναγκαία*

συνθήκη για το *B*, το επίθετο *ικανός*, το οποίο έχει θετική αρχική σημασία, δεν επηρεάζει τον σημασιολογικό προσανατολισμό της πρότασης, όπως π.χ. στην πρόταση *Ο Χ είναι ικανός υπάλληλος*, μέσω της οποίας εκφράζεται μια θετική αξιολογική κρίση. Και αντίστροφα, η καταγραφή θετικών και αρνητικών όρων, με μοναδικό γνώμονα το αρχικό σημασιολογικό τους φορτίο, θα απέκλειε λέξεις όπως *ωμός* ή *ρηχός*, λέξεις με ουδέτερη αρχική σημασία, οι οποίες όταν χρησιμοποιούνται μεταφορικά προσδίδουν αρνητικό προσανατολισμό στην φράση/πρόταση στην οποία περιέχονται (π.χ. *ωμό κρέας/ωμός τρόπος, ρηχά νερά/ρηχός άνθρωπος*). Στις περιπτώσεις αυτές, γίνεται σαφές ότι θα έπρεπε να εξεταστεί και το κειμενικό περιβάλλον κάθε εξεταζόμενου όρου, ώστε να έχουμε μια ολοκληρωμένη άποψη για το *αξιολογικό δυναμικό (evaluative potential)* του κάθε όρου, τις πιθανότητες, δηλαδή, που έχει να χρησιμοποιηθεί στον αξιολογικό λόγο προκειμένου να εκφράσει συναίσθημα, κρίση ή εκτίμηση (Taboada, M. & Grieve, J. 2006).

Παρά το γεγονός ότι έχει δοθεί ιδιαίτερη έμφαση σε τέτοιου είδους προσεγγίσεις, όσον αφορά την δημιουργία λεξικών πόρων αγγλικής γλώσσας, δεν έχουν αναπτυχθεί ανάλογα λεξικά για την ελληνική γλώσσα.

Εντοπίζοντας αυτό το κενό στην ερευνητική προσέγγιση αυτού του πεδίου, κρίθηκε σκόπιμη η σύνθεση ενός ελληνικού λεξικού υποκειμενικών λέξεων και φράσεων (*opinion words*), που θα μπορεί να χρησιμοποιηθεί σε εφαρμογές *εξόρυξης γνώμης*. Η μεθοδολογική προσέγγιση καθώς και τα στάδια που ακολουθήθηκαν κατά την πορεία σύνθεσής του είναι και το θέμα που θα μας απασχολήσει σε αυτή την εργασία.

Αρχικά, θα ορίσουμε το ερευνητικό πεδίο της εργασίας μας (*Sentiment Analysis*) αναφέροντας τις επικρατέστερες μεθοδολογικές προσεγγίσεις, προκειμένου να γίνει αντιληπτός τόσο ο στόχος της, όσο και η πορεία υλοποίησής του.

Στη συνέχεια, θα παρουσιαστούν οι κυριότεροι διαθέσιμοι λεξικοί πόροι αγγλικής γλώσσας, προκειμένου να κατανοηθεί αφ'ενός ο τρόπος δημιουργίας τους και αφ'ετέρου ο τρόπος λειτουργίας τους σε εφαρμογές *Ανάλυσης Συναισθήματος*.

Ακολούθως, στο τρίτο κεφάλαιο, αναλύεται εκτενώς η πορεία σύνθεσης του λεξικού μας, από την αρχική πηγή έμπνευσης και τη στοχοθεσία έως τη μεθοδολογική προσέγγιση και την υλοποίησή της. Αρχικά, θα αναλυθεί η θεωρία *Appraisal* (Martin and White, 2005), μια γλωσσολογική θεωρία που συστηματοποιεί τους δυνατούς

τρόπους μέσω των οποίων εκφράζονται απόψεις και συναισθήματα σε ένα κείμενο. Η θεωρία αυτή όρισε τη βασική κατευθυντήρια γραμμή για το τί είδους λέξεις πρέπει να συμπεριλάβουμε στο λεξικό και έθεσε τα κριτήρια τόσο για τη συλλογή των *όρων άποψης* όσο και για την κατηγοριοποίησή τους. Οι άξονες της θεωρίας που θα μας απασχολήσουν περισσότερο είναι η *κριτική προσώπων (judgement)* και η *αξιολόγηση πραγμάτων (appreciation)*, καθώς και οι τρόποι πραγμάτωσής τους μέσα στα κείμενα.

Θα εστιάσουμε κυρίως στα στάδια που ακολούθησε, δηλαδή, τη συλλογή των *υποκειμενικών όρων*, την αξιολόγησή τους ως προς το *αξιολογικό δυναμικό* τους και την κατηγοριοποίησή τους σε ομάδες κοινών χαρακτηριστικών. Επίσης, θα αναφερθούμε στα μέσα συλλογής των *υποκειμενικών όρων*, καθώς και τους τρόπους με τους οποίους εμπλουτίστηκαν με νέους. Τέλος, θα εξετάσουμε τους δυνατούς τρόπους αξιοποίησής του σε εφαρμογές κειμενικής ανάλυσης.

2. Θεωρητικό υπόβαθρο

2.1. Sentiment Analysis

Η *Ανάλυση Συναισθήματος (Sentiment Analysis)* ή αλλιώς η *Εξόρυξη Γνώμης (Opinion Mining)* είναι ο τομέας του Natural Language Processing που ασχολείται με την υπολογιστική ανάλυση των απόψεων, των συναισθημάτων, των εκτιμήσεων, των αξιολογήσεων και των στάσεων των ανθρώπων προς οντότητες, όπως άτομα, προϊόντα, υπηρεσίες, θέματα, γεγονότα και τα χαρακτηριστικά τους (Liu, 2012). Παρά το γεγονός ότι πρόκειται για σχετικά πρόσφατο τομέα της υπολογιστικής-γλωσσολογικής έρευνας, έχει αναδειχθεί σε ένα αρκετά δραστήριο ερευνητικό τομέα, κυρίως λόγω της πληθώρας των εφαρμογών που προσφέρει σχεδόν σε κάθε πεδίο (βιομηχανία προϊόντων, marketing, οικονομικές και πολιτικές επιστήμες, Μέσα Μαζικής Επικοινωνίας, ψυχολογία κ.α.) αλλά και των ποικίλων ερευνητικών ζητημάτων που έχει θέσει.

Η μελέτη του *Sentiment Analysis* έχει έως τώρα υλοποιηθεί σε τρία βασικά επίπεδα:

α) Επίπεδο Κειμένου (*document level*): Το ζητούμενο σε αυτό το επίπεδο είναι να προσδιοριστεί η ευρύτερη στάση του συγγραφέα, θετική ή αρνητική, σε ένα κείμενο που περιλαμβάνει κρίσεις και απόψεις (*opinionated text*). Για παράδειγμα, σε ένα

κείμενο το οποίο περιέχει κριτική για ένα προϊόν, η ανάλυση συναισθήματος εξετάζει τη συνολική θετική ή αρνητική στάση που αυτό απηχεί. Το μειονέκτημα αυτής της προσέγγισης είναι ότι θεωρεί ως δεδομένο ότι σε ένα κείμενο η κριτική έχει μόνο ένα αντικείμενο αναφοράς, οπότε πρακτικά δεν είναι εφαρμόσιμη σε περιπτώσεις κειμένων που περιέχουν π.χ. σύγκριση δύο διαφορετικών προϊόντων.

Οι περισσότερες τεχνικές που χρησιμοποιούνται για αυτό το επίπεδο ανάλυσης βασίζονται στην *επιβλεπόμενη εκμάθηση μηχανής (supervised machine learning)*, η οποία έχει ως βασικό στόχο την αυτοματοποίηση ενός συνόλου *χαρακτηριστικών (features)*, όπως τις μορφολογικές κατηγορίες (*Part-Of-Speech tagging*), τις συντακτικές σχέσεις, το φαινόμενο της άρνησης κ.α.

β) Επίπεδο Πρότασης (*sentence level*): Ο τρόπος ανάλυσης σε αυτό το επίπεδο εστιάζει στην πρόταση και τον ακριβή προσδιορισμό της θετικής, αρνητικής ή ουδέτερης στάσης που εκφράζει. Η προσέγγιση αυτή συχνά συνδέεται στη βιβλιογραφία με τον όρο *subjectivity classification* (Wiebe, Bruce and O'Hara, 1999), που διαχωρίζει τις προτάσεις που περιέχουν γεγονότα της πραγματικότητας (*objective sentences*) από αυτές που περιέχουν υποκειμενικές κρίσεις (*subjective sentences*). Ωστόσο, δεν είναι βέβαιο ότι η γνώμη μπορεί να ανιχνευθεί μόνο σε υποκειμενικές εκφράσεις, αφού υπάρχουν και περιπτώσεις κατά τις οποίες υποδηλώνεται και σε αντικειμενικές εκφράσεις (π.χ. *Πέρασαν ούτε δυο μήνες από τότε που αγόρασα το αυτοκίνητο και αμέσως χάλασε*). Τα δύο βασικά ζητήματα που ερευνώνται μέσω αυτής της προσέγγισης είναι ο έλεγχος της υποκειμενικότητας μιας πρότασης και στη συνέχεια ο ακριβής προσδιορισμός του *σημασιολογικού προσανατολισμού* της. Επίσης, συχνά συνυπολογίζονται πολλοί παράμετροι, όπως το φαινόμενο της άρνησης (*negation*), π.χ. *καθόλου καλός*, το θέμα της τροπικότητας (*modality*), η αμφισημία των λέξεων, ο συντακτικός ρόλος των λέξεων στην πρόταση κ.α. (Wilson et al., 2005).

γ) Επίπεδο Οντότητας και Χαρακτηριστικών (*entity and feature level*): Το επίπεδο αυτό, (για το οποίο χρησιμοποιείται και ο όρος *aspect-based*), επιχειρεί μια περισσότερο λεπτομερή ανάλυση σε σχέση με τους δύο προηγούμενους, καθώς εστιάζει στην ίδια την άποψη και όχι σε ανάλυση δομικών στοιχείων της γλώσσας (κείμενο, πρόταση, φράση). Βασίζεται κυρίως στη ιδέα ότι μια υποκειμενική κρίση αποτελείται από ένα *συναίσθημα (sentiment)* και έναν στόχο (*target*) στον οποίο

απευθύνεται, ο οποίος στα περισσότερα συστήματα κειμενικής ανάλυσης αναπαρίσταται μέσω οντοτήτων (*entities*). Στόχος αυτής της ανάλυσης είναι να αναζητήσει τα εκφραζόμενα συναισθήματα και απόψεις προς τα αντικείμενα-στόχους, αλλά και τα επιμέρους χαρακτηριστικά τους (*aspects*). Για παράδειγμα στην πρόταση «*Η ποιότητα κλήσης του iPhone είναι καλή, αλλά η διάρκεια ζωής της μπαταρίας είναι μικρή*», αξιολογούνται δύο διαφορετικές όψεις της ίδιας οντότητας (η ποιότητα κλήσης και η διάρκεια μπαταρίας). Αυτού του είδους η ανάλυση συνεξετάζει και άλλους παράγοντες που σχετίζονται με την έκφραση της άποψης, όπως το πρόσωπο που εκφράζει την άποψη (*opinion holder*) αλλά και τον χρόνο της έκφρασης (*time*).

Οι τεχνικές που συνήθως αξιοποιούνται για την ανάλυση αυτού του επιπέδου στοχεύουν στην εξόρυξη των χαρακτηριστικών των αξιολογούμενων οντοτήτων (*feature extraction*) και στον προσδιορισμό-κατηγοριοποίηση αυτών των χαρακτηριστικών (*feature sentiment classification*) ως προς το τρίπτυχο θετικό-αρνητικό-ουδέτερο, που υλοποιείται κυρίως μέσω των προσεγγίσεων που περιλαμβάνουν *επιβλεπόμενη εκμάθηση μηχανής* και δημιουργία λεξικών πόρων. (Liu, 2012)

2.2. Σχετική Έρευνα – Λεξικά

Όπως είδαμε, ένας από τους πιο κοινούς τρόπους ερευνητικής προσέγγισης στο πεδίο της *ανάλυσης συναισθήματος* (*sentiment analysis*), είναι η δημιουργία λεξικών πόρων αποτελούμενων από όρους που εκφράζουν άποψη, θετική ή αρνητική (*opinion words*). Στη βιβλιογραφία οι όροι αυτοί ονομάζονται επίσης *polar words*, *opinion-bearing words* και *sentiment words/expressions* (Wilson et al., 2005, Liu, 2010). Πρόκειται για λέξεις-φράσεις προκαθορισμένου *σημασιολογικού προσανατολισμού* (*semantic orientation*) ως προς το υποκειμενικό τους περιεχόμενο (Taboada, M. & Grieve, J. 2006), που λειτουργούν ως βασικοί πυρήνες κατά τις εφαρμογές κειμενικής ανάλυσης.

Κατά τον Liu, (2010) για τη δημιουργία τέτοιων λεξικών (*opinion lexicons*), υπάρχουν τρεις βασικές μεθοδολογικές προσεγγίσεις: α) η μη αυτοματοποιημένη

αναζήτηση και συλλογή όρων, η οποία, αν και χρονοβόρα, συνήθως συνδυάζεται με τις αυτοματοποιημένες μεθόδους για τη διόρθωση λαθών των τελευταίων.

β) Η μέθοδος που στηρίζεται σε λεξικά και στην τεχνική *bootstrapping*, η οποία έγκειται στη συλλογή κάποιων αρχικών όρων με προκαθορισμένο *προσανατολισμό*, που λειτουργούν ως *σπόροι όρων* (*seed terms*) και στο διαδοχικό εμπλουτισμό τους με συνώνυμα και αντώνυμα μέσω κάποιου ηλεκτρονικού λεξικού, όπως π.χ. το WordNet. Η τεχνική αυτή παρουσιάζει το μειονέκτημα της αδυναμίας εντοπισμού των ιδιαίτερων σημασιολογικών αποχρώσεων που μπορεί να λαμβάνει ένας όρος κατά τη χρήση του (π.χ. *cheap trick/ticket*, Hatzivassiloglou and McKeown, 1997).

γ) Η μέθοδος που βασίζεται στην αξιοποίηση ηλεκτρονικών σωμάτων κειμένων (*corpus-based approach*) και στην αναζήτηση σε αυτά συντακτικών ή συμφραστικών κανόνων που διέπουν τις σχέσεις μιας αρχικής ομάδας λέξεων. Η λογική της τεχνικής αυτής έγκειται στη χρήση κάποιας αρχικής ομάδας επιθέτων-υποκειμενικών όρων, καθώς και των γλωσσικών κανόνων που διέπουν τις μεταξύ τους σχέσεις, ώστε να προκύψουν επιπρόσθετοι ανάλογοι όροι (Hatzivassiloglou and McKeown, 1997). Ένα τέτοιο είδος γλωσσικών κανόνων είναι η συμπλεκτική σύνδεση δύο όρων, που συνήθως δηλώνει ότι τα συνδεόμενα μέρη, εν προκειμένω τα επίθετα (π.χ. *το χ κινητό είναι κομψό και πρακτικό*), παρουσιάζουν τον ίδιο *σημασιολογικό προσανατολισμό* (*orientation*). Η τεχνική αυτή, που ονομάζεται *συνεκτικότητα συναισθήματος* (*sentiment consistency*), έχει χρησιμοποιηθεί ευρέως από πολλούς ερευνητές με σημαντικές βελτιώσεις στην αρχική σύλληψη. Οι Kanayama και Nasukawa, 2006 (όπως αναφέρει ο Liu, 2010) επέκτειναν την αρχική ιδέα εισάγοντας τις έννοιες της *ενδο-προτασιακής* και της *δια-προτασιακής συνεκτικότητας* (*intra-sentential* και *inter-sentential sentiment consistency*) αναζητώντας αναλογίες *σημασιολογικού προσανατολισμού* και σε επίπεδο ολόκληρης περιόδου λόγου και όχι μεμονομένων προτάσεων. Επίσης, οι Qiu et al, 2009 (όπως αναφέρει ο Liu, 2010) εφάρμοσαν τη μέθοδο της *διάδοσης* (*propagation*), μέσω της οποίας εντοπίζονται συγκεκριμένες συντακτικές σχέσεις μεταξύ των *όρων άποψης* και των *χαρακτηριστικών* (*features*) των στόχων-αντικειμένων αναφοράς αυτών των όρων. Οι σχέσεις αυτές μπορούν διαδοχικά να χρησιμοποιηθούν για την αναγνώριση νέων *όρων άποψης* και νέων *χαρακτηριστικών* έως την εξάντλησή τους. Ακόμη, οι Ding και Liu, 2008 (όπως αναφέρει ο Liu, 2010) ακολούθησαν τη μέθοδο εντοπισμού των *όρων άποψης* και του *σημασιολογικού προσανατολισμού* τους σε σχέση με τα *χαρακτηριστικά* του εκάστοτε

αντικειμένου αναφοράς, αποδεικνύοντας ότι μια λέξη μπορεί να έχει διαφορετικό προσανατολισμό σε διαφορετικά συμφραζόμενα ακόμα και κατά τη χρήση της στο ίδιο πεδίο. Ωστόσο, παρά τις βελτιώσεις που έχει επιδεχθεί, η μεθοδολογική προσέγγιση που στηρίζεται στα σώματα κειμένων, παρουσιάζει το μειονέκτημα της αδυναμίας τους να περιλαμβάνουν όλες τις λέξεις μιας γλώσσας.

Προς κατανόηση της μεθοδολογίας που ακολουθείται κατά τη σύσταση των λεξικών πόρων που προορίζονται για αξιοποίηση σε συστήματα συναισθηματικής ανάλυσης καθώς και του τρόπου λειτουργίας τους, παρουσιάζονται επιλεκτικά τα παρακάτω:

2.3.1. SentiWordNet

(Baccianella Stefano, Andrea Esuli and Fabrizio Sebastiani., 2010)

Λεξικός πόρος ελεύθερης πρόσβασης, που χρησιμοποιεί ως βασικό δομικό στοιχείο τα σύνολα συνωνύμων (*synsets*) της λεξιλογικής βάσης δεδομένων WordNet. Κάθε σύνολο συνωνύμων (*synset*) καθορίζεται από τρεις βαθμολογίες *Obj(s)*, *Pos(s)* και *Neg(s)*, που περιγράφουν πόσο αντικειμενικοί, θετικοί ή αρνητικοί είναι οι όροι που το απαρτίζουν. Η μέθοδος που χρησιμοποιείται για την ανάπτυξη του SentiWordNet βασίζεται στην ποσοτική ανάλυση των ορισμών (*glosses*) κάθε συνόλου συνωνύμων (*synset*) και στη χρήση των παραγόμενων διανυσματικών αναπαραστάσεων κάθε όρου για την ημιαυτόματη ταξινόμηση των *synsets*. Οι τρεις βαθμολογίες, προϊόν συνδυασμού των αποτελεσμάτων που παράγονται από ένα σύνολο οκτώ τριμερών ταξινομητών, χαρακτηρίζονται από παρόμοια επίπεδα ακρίβειας, αλλά παρουσιάζουν διαφορετική συμπεριφορά ταξινόμησης.

Η χρήση των συνόλων συνωνύμων (*synsets*) και όχι των ίδιων των όρων ως βασικών μονάδων ανάπτυξης του λεξικού προσφέρει τη δυνατότητα διερεύνησης των ποικίλων σημασιολογικών αποχρώσεων του ίδιου όρου καθώς και των ιδιοτήτων που λαμβάνουν σε σχέση με την έκφραση γνώμης. (Baccianella, Esuli, Sebastiani, 2010).

Στον πίνακα που ακολουθεί, αποτυπώνεται η βασική δομή του λεξικού.

POS	ID	PosScore	NegScore	SynsetTerms	Gloss
a	1740	0.125	0	able#1	(usually followed by `to') having the necessary means or [...]
a	2098	0	0.75	unable#1	(usually followed by `to') not having the necessary means or [...]
a	2312	0	0	dorsal#2 abaxial#1	facing away from the axis of an organ or organism; [...]
a	2527	0	0	ventral#2 adaxial#1	nearest to or facing toward the axis of an organ or organism; [...]
a	2730	0	0	acrosopic#1	facing or on the side toward the apex
a	2843	0	0	basisopic#1	facing or on the side toward the base
a	2956	0	0	abducting#1 abducent#1	especiallyofmuscles; [...]
a	3131	0	0	adductive#1 adducting#1 adducent#1	especiallyofmuscles; [...]
a	3356	0	0	nascent#1	beingbornorbeginning; [...]
a	3553	0	0	emerging#2 emergent#2	comingintoexistence; [...]

Πίνακας 3. Απόσπασμα από το λεξικό SentiWordNet.(Παράδειγμα από την ιστοσελίδα <http://sentiment.christopherpotts.net/lexicons.html#resources>)

2.3.2. Linguistic Inquiry and Word Count

(James W. Pennebaker, Roger J. Booth, and Martha E. Francis, 2007)

Το λεξικό LIWC2007 αποτελεί τον κεντρικό πυρήνα λειτουργίας της ομώνυμης εφαρμογής ανάλυσης κειμένου. Η βασική μορφή του λεξικού αποτελείται από σχεδόν 4.500 λέξεις και ρίζες λέξεων, εκ των οποίων η κάθε μία μπορεί να ανήκει σε μία ή περισσότερες κατηγορίες λέξεων ή υπολεξικά που αναφέρονται κυρίως σε ψυχολογικές και διανοητικές καταστάσεις αλλά και σε γραμματικά, δομικά και παραγωγικά στοιχεία (π.χ μέρος του λόγου, μήκος λέξεων, επιφωνήματα). Κατά τη διαδικασία ανάλυσης κειμένου εξετάζεται κάθε λέξη χωριστά και σε περίπτωση αντιστοίχισής της με μία ή περισσότερες κατηγορίες του λεξικού, αυξάνεται ο αντίστοιχος μετρητής του συγκεκριμένου χαρακτηριστικού. Για παράδειγμα, η λέξη *cried* ανήκει στις εξής πέντε κατηγορίες λέξεων: *sadness*, *negative emotion*, *overall*

affect, verb, και *past tense verb*. Έτσι, αν εντοπιστεί στο κείμενο-στόχο, προσμετράται χωριστά κάθε μία βαθμολογία (*score*) αυτών των κατηγοριών.

Η δημιουργία του λεξικού περιελάμβανε πολλά στάδια κατά τα οποία επεκτάθηκε τόσο σε εύρος κατηγοριών όσο και σε πλήθος λέξεων που τις απαρτίζουν. Η *Συλλογή λέξεων (Word Collection)*, (δημιουργία των κατηγοριών και εμπλουτισμός τους με λήμματα), οι *Φάσεις Βαθμονόμησης των Αξιολογητών (Judges' Rating Phases)* (πολύπλευρη διαδοχική αξιολόγηση της βαθμολογικής κλίμακας κάθε κατηγορίας από τρεις αντικειμενικούς κριτές), η *Ψυχομετρική Αξιολόγηση (Psychometric Evaluation)* (αφαίρεση των σπάνια χρησιμοποιούμενων κατηγοριών και προσθήκη νέων βάσει ανάλυσης εκτεταμένου αριθμού κειμένων) και ο *Εκσυγχρονισμός-Προσθήκες (Updates and Expansions)* (αναδιαμόρφωση της δομής του λεξικού με έμφαση σε όλα τα είδη του γραπτού και προφορικού λόγου) αποτελούν τα σημαντικότερα βήματα ανάπτυξης του λεξικού που καθόρισαν την τελική τρέχουσα μορφή του, η οποία υποστηρίζει 82 γλώσσες. Στους παρακάτω πίνακες υπάρχουν παραδείγματα ανάλυσης κειμένου μέσω της εφαρμογής LIWC καθώς και λέξεων για κάθε μια από τις κατηγορίες που απαρτίζουν το λεξικό .

LIWC Results			
<i>Details of Writer: No Details</i>			
<i>Date/Time: 27 August 2013, 5:17 pm</i>			
LIWC Dimension	Your Data	Personal Texts	Formal Texts
Self-references (I, me, my)	0.32	11.4	4.2
Socialwords	10.02	9.5	8.0
Positiveemotions	1.79	2.7	2.6
Negativeemotions	3.06	2.6	1.6
Overallcognitivewords	5.06	7.8	5.4
Articles (a, an, the)	10.34	5.0	7.2
Bigwords (> 6 letters)	27.74	13.1	19.6

The text you submitted was 948 words in length.

Πίνακας 4. Παράδειγμα ανάλυσης κειμένου μέσω της εφαρμογής LIWC (<http://www.liwc.net/tryonline.php>)

2.3.3. Multi Perspective Question Answering Subjectivity Lexicon (MPQA)

(Theresa Wilson, Janyce Wiebe, Paul Hoffmann, 2005)

Το λεξικό MPQA Subjectivity Lexicon, μέρος του συστήματος εξόρυξης γνώμης Opinion Finder, είναι μια λίστα 8221 υποκειμενικών στοιχείων (*subjectivity clues*), κυρίως λέξεων και φράσεων, δομημένων σε ισάριθμες σειρές. Για κάθε *subjectivity clue* δίνονται πληροφορίες σχετικά με τον τύπο υποκειμενικότητάς του (*strongsubj/weaksubj*), το μήκος του στοιχείου σε λέξεις (*len*), τη μορφή εμφάνισης του στοιχείου (*word*) και το μέρος του λόγου που αντιπροσωπεύει (*pos*), όπως επίσης και για το μορφολογικό χαρακτήρα (*stemmed/unstemmed*) και την a priori πολικότητα (*prior polarity*) του κάθε στοιχείου ανεξάρτητα από το κειμενικό περιβάλλον. Στον πίνακα που ακολουθεί, φαίνεται η δομή του λεξικού. Για παράδειγμα το στοιχείο 1 έχει χαρακτηριστεί με την ετικέτα *weaksubj*, καθώς παρουσιάζει υποκειμενική σημασία σε συγκεκριμένα κειμενικά περιβάλλοντα, αποτελεί απλή λέξη και επίθετο. Επίσης, δεν έχει υποστεί επεξεργασία *stemming* και έχει αρνητική πολικότητα.

	Strength	Length	Word	Part-of-speech	Stemmed	Polarity
1.	type=weaksubj	len=1	word1=abandoned	pos1=adj	stemmed1=n	priorpolarity =negative
2.	type=weaksubj	len=1	word1=abandonment	pos1=noun	stemmed1=n	priorpolarity =negative
3.	type=weaksubj	len=1	word1=abandon	pos1=verb	stemmed1=y	priorpolarity =negative
4.	type=strongsubj	len=1	word1=abase	pos1=verb	stemmed1=y	priorpolarity =negative
5.	type=strongsubj	len=1	word1=abasement	pos1=anypos	stemmed1=y	priorpolarity =negative
6.	type=strongsubj	len=1	word1=abash	pos1=verb	stemmed1=y	priorpolarity =negative

Πίνακας 5. Απόσπασμα από το λεξικό υποκειμενικότητας MPQA

(παράδειγμα από την ιστοσελίδα: <http://sentiment.christopherpotts.net/lexicons>)

2.3.4. Bing Liu's Opinion Lexicon

(Minqing Hu and Bing Liu, 2004)

Πρόκειται για μια λίστα θετικών (2006) και αρνητικών (4783) λέξεων της αγγλικής γλώσσας που εκφράζουν γνώμη ή συναίσθημα. Παρά το γεγονός ότι χαρακτηρίζεται από ποικιλία υφολογικών επιπέδων, καθώς περιλαμβάνει σκόπιμα αρκετές μορφολογικές παραλλαγές και παραφθορές των λέξεων (π.χ. *zealous*) και εκφράσεις της καθομιλουμένης που απαντούν συχνά σε κοινωνικά μέσα δικτύωσης (*bastard, 2-faced*), ο ίδιος ο δημιουργός του τονίζει ότι η ύπαρξη ενός αντίστοιχου λεξικού δεν εξασφαλίζει απαραίτητα και ακρίβεια στην διαδικασία εξόρυξης γνώμης.

2.3.5. General Inquirer

(Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, Daniel M. Ogilvie, 1966).

Το λεξικό του ομώνυμου συστήματος κειμενικής ανάλυσης, αποτελεί στην πραγματικότητα ένα σύστημα 182 κατηγοριών (*tag categories*), εκ των οποίων η κάθε μία αποτελείται από ένα σύνολο λέξεων και των αντίστοιχων ερμηνειών τους. Δύο βασικές εκτενείς κατηγορίες είναι αυτές των θετικών (1915) και αρνητικών (2291) λέξεων. Οι περισσότερες κατηγορίες είναι δομημένες κυρίως σε δύο υπολεξικά, το **HarvardIV-dictionary**, που περιλαμβάνει κατηγορίες όπως: λέξεις που δείχνουν χαρά, πόνο, αρετή, λέξεις που φανερώνουν υπερβολή ή εγκράτεια, λέξεις που απαντούν σε ένα συγκεκριμένο χώρο κ.α. και το **Lasswell value dictionary**, που διακρίνει τα λήμματα σε τέσσερις τομείς σεβασμού (*ισχύς, ευθύτητα, υπόληψη, δεσμός*) και σε τέσσερις τομείς ευημερίας (*πλούτος, ευεξία, φώτιση και ικανότητα*). Παράλληλα, υπάρχουν και οι “*marker categories*”, δύο κατηγορίες μορφολογικών και σημασιολογικών δεικτών, όπως *άρθρο, αριθμητικό, πρόθεση* και *αρσενικό, θηλυκό, χώρος* και *χρόνος*.

2.3.6. Affective Norms for English Words (ANEW)

(Margaret M. Bradley and Peter J. Lang, 1999)

Το ANEW αποτελεί ένα σύνολο κανονιστικών συναισθηματικών αξιολογήσεων για ένα μεγάλο αριθμό λέξεων της αγγλικής γλώσσας. Ο στόχος δημιουργίας του ήταν να σχηματιστεί ένα σύστημα λεκτικού υλικού, αξιολογημένου υπό το πρίσμα τριών

βασικών διαστάσεων, της ευχαρίστησης (*pleasure*), της διέγερσης (*arousal*) και της κυριαρχίας (*dominance*), συμπληρώνοντας το υπάρχον Διεθνές Σύστημα Συναισθηματικής Εικόνας (*International Affective Picture System*) (IAPS, Lang, Bradley & Cuthbert, 1999) και το Διεθνές Σύστημα Συναισθηματικών Ψηφιοποιημένων Ήχων (*International Affective Digitized Sounds*) (IADS; Bradley & Lang, 1999), με συλλογές οπτικών και ηχητικών ερεθισμάτων αντίστοιχα, που περιλαμβάνουν αυτές τις συναισθηματικές αξιολογήσεις. Για τη διαδικασία αξιολόγησης, (στην οποία συμμετείχαν φοιτητές Εισαγωγικής Ψυχολογίας), αξιοποιήθηκε το *Self-Assessment Manikin* (SAM), ένα σύστημα συναισθηματικής αξιολόγησης, με διπολικές κλίμακες βαθμολόγησης κάθε συναισθηματικής κατάστασης (*pleasure, arousal, dominance*), το οποίο εφαρμόστηκε σε ένα σύνολο 1040 λέξεων, κυρίως ουσιαστικών, επιθέτων και ρημάτων. Σε κάθε λέξη-λήμμα του λεξικού αντιστοιχεί μια μέση τιμή για κάθε μία από τις τρεις κλίμακες αξιολόγησης (*pleasure, arousal, dominance*).

Η δημιουργία του λεξικού βασίζεται στη θεωρία των συναισθηματικών διαστάσεων, σύμφωνα με την οποία τα συναισθηματικά ερεθίσματα αξιολογούνται βάσει συναισθηματικών διπολικών διαστάσεων και όχι διακριτών συναισθηματικών κατηγοριών.

3. Ελληνικό Λεξικό Υποκειμενικών Όρων

3.1. Σκοπός

Παρά το γεγονός ότι έχουν αναπτυχθεί πολυάριθμοι και αποτελεσματικοί στο ζήτημα της *Εξόρυξης Γνώμης* λεξικοί πόροι για την αγγλική γλώσσα, δεν έχει δημιουργηθεί ανάλογος για την ελληνική. Αυτή η διαπίστωση αποτέλεσε και το έναυσμα για την ενασχόληση με αυτή την προσπάθεια δημιουργίας ενός λεξικού υποκειμενικών όρων.

Ο βασικός στόχος της δημιουργίας του Λεξικού *υποκειμενικών όρων* ήταν η σύνθεση ενός ελληνικού λεξικού πόρου, που θα μπορεί να αξιοποιηθεί σε ένα αυτοματοποιημένο σύστημα εξόρυξης γνώμης. Ως *υποκειμενικοί όροι* (*subjectivity clues*), ορίζονται οι λέξεις και οι φράσεις που μπορούν να χρησιμοποιηθούν για να εκφράσουν εσωτερική στάση, γνώμη, άποψη κ.λ.π. (Wilson et al., 2005). Οι λεξικοί πόροι (*lexical resources*) που προορίζονται για χρήση στον τομέα της *Ανάλυσης*

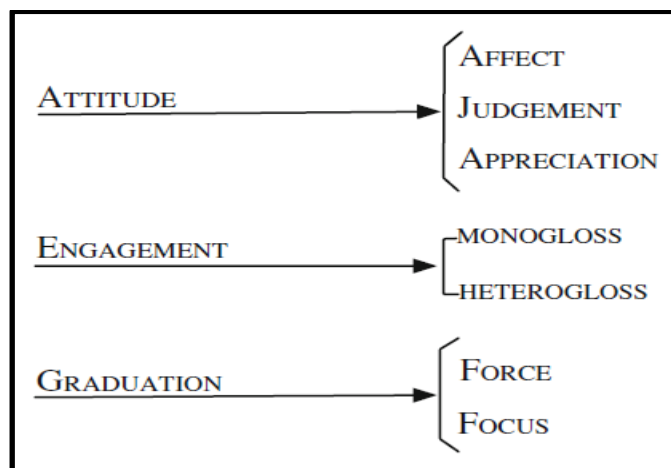
Συναισθημάτων (Sentiment Analysis) περιέχουν όρους, χαρακτηρισμένους με μια συγκεκριμένη τιμή *πολικότητας*, θετικής ή αρνητικής βάσει του σημασιολογικού φορτίου τους. Με τον όρο *πολικότητα (polarity)* εννοούμε το θετικό ή αρνητικό *σημασιολογικό προσανατολισμό* μιας λέξης (Wilson et al.,2005).

Το εγχείρημα αυτό εστίασε περισσότερο στη χρήση του αξιολογικού λόγου ως γλωσσικού μέσου έκφρασης *στάσης (Attitude)* προς κάποιο αντικείμενο αξιολόγησης, υπό το πρίσμα της θεωρίας *Appraisal*, και κυρίως των όρων *κριτική (Judgment)*, η οποία απευθύνεται προς πρόσωπα και *αξιολόγηση (Appreciation)*, η οποία έχει ως αντικείμενο αναφοράς πράγματα, καταστάσεις, γεγονότα κ.λ.π. (Martin, 2000, 2003). Συγκεκριμένα, επιλέχθηκε η αναζήτηση και συγκέντρωση επιθέτων που χρησιμοποιούνται για την εκφορά υποκειμενικού λόγου, καθώς η έρευνα στον τομέα αυτό έχει αποδείξει στενή σχέση μεταξύ της παρουσίας των επιθέτων σε μια πρόταση και το βαθμό υποκειμενικότητάς της (Hatzivassiloglou and Wiebe, 2000). Αυτή η διαπίστωση έχει θεωρηθεί ως η ένδειξη ότι τα επίθετα συνιστούν ασφαλείς δείκτες υποκειμενικότητας και αξιολογικού λόγου, κατευθύνοντας συχνά την έρευνα στην *Ανάλυση Συναισθήματος* προς μια προσέγγιση βασισμένη στην επισημείωση των μερών του λόγου (*P-O-S tagging*). Όπως συνοψίζουν και οι Wilks και Stevenson, (1998) « η επισημείωση των μερών του λόγου μπορεί να θεωρηθεί μια ακατέργαστη μορφή αποσαφήνισης της έννοιας κάθε λέξης».

3.2. Θεωρία *Appraisal*

Πρόκειται για μια θεωρία της συστημικής λειτουργικής γλωσσολογίας για την αξιολόγηση σε ένα κείμενο (Martin and White, 2005). Ουσιαστικά, είναι μια προσπάθεια συστηματοποίησης της δυνατότητας της γλώσσας να εκφράσει γνώμες και στάσεις σε ένα κείμενο (Taboada, Grieve, 2004). Αποτελείται από τρία υποσυστήματα που λειτουργούν αλληλεπιδραστικά: Το υποσύστημα *Attitude(στάση)*, που σχετίζεται με την έκφραση προσωπικών συναισθημάτων, είτε πρόκειται για συναισθηματικές αντιδράσεις, κριτική συμπεριφοράς ή αξιολόγηση πραγμάτων, το υποσύστημα *Engagement(δέσμευση)*, το οποίο εξετάζει το βαθμό δέσμευσης του συγγραφέα στις θέσεις και εκτιμήσεις που εκφράζονται είτε από τον ίδιο (*monogloss*), είτε από άλλους συγγραφείς (*heterogloss*), και το υποσύστημα

Graduation (διαβάθμιση), το οποίο περιγράφει τις γλωσσικές λειτουργίες μέσω των οποίων εντείνεται ή μειώνεται η ισχύς της εκφραζόμενης άποψης.



Πίνακας 1. Σχηματική απεικόνιση της θεωρίας Appraisal. (Read, J., Carroll, J. 2010).

3.2.1. Attitude

Το υποσύστημα *Attitude* διακρίνεται σε τρεις πτυχές της ανθρώπινης προσωπικότητας: τον ψυχισμό, την ηθική και την αισθητική. Όλοι οι επιμέρους τύποι κάθε κατηγορίας του υποσυστήματος *Attitude* μπορούν επιπροσθέτως να κατηγοριοποιηθούν ανάλογα με την *πολικότητά* τους, αν έχουν δηλαδή, θετικό ή αρνητικό σημασιολογικό προσανατολισμό, (Liu, 2010) διακρίνοντας, έτσι, τους θετικούς από τους αρνητικούς όρους.

3.2.1.1. Συναισθηματική κατάσταση

Το σύστημα *appraisal* περιλαμβάνει τέσσερις υποκατηγορίες συναισθηματικής κατάστασης: *Προδιάθεση(Dis/Inclination)*, που αφορά όρους που εκφράζουν επιθυμία ή αποστροφή προς κάποια κατάσταση (π.χ. θελκτικός, αβάσταχτος, σιχαμερός), *Ευτυχία/Αυστηχία(Un/Happiness)*, που σχετίζεται με την ψυχική κατάσταση (π.χ. ευχαριστημένος, κεφάλτος, οδυνηρός), *Αν-/Ασφάλεια(Un/Security)*, που περιλαμβάνει συναισθήματα ασφάλειας και ηρεμίας σε σχέση με το κοινωνικό

περιβάλλον (π.χ. αξιосέβαστος, ισορροπημένος, ανήσυχος, έκπληκτος, άνετα, άβολα) και *Ικανοποίηση/Απογοήτευση(Dis/Satisfaction)*, που αντιστοιχεί στις προσωπικές προσδοκίες (π.χ. ικανοποιημένος, βαρετός).

3.2.1.2 Κριτική προσώπων

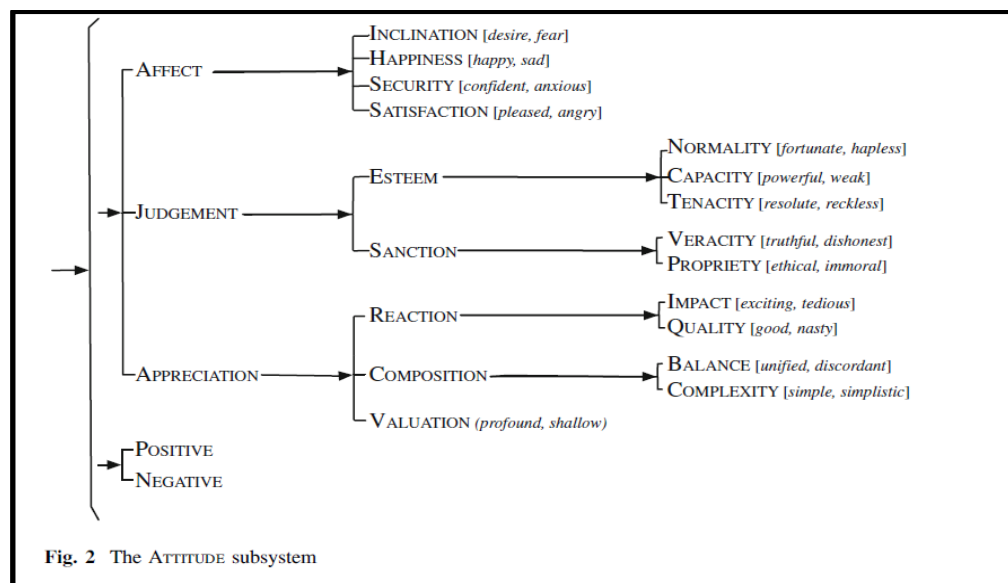
Η κριτική ατόμων διακρίνεται σε δύο βασικές κατηγορίες: α) την *Υπόληψη(Esteem)*, που σχετίζεται με την κανονικότητα ως προς τη κοινωνική συμπεριφορά ενός ατόμου (π.χ. μοντέρνος, ξεπερασμένος, τυχερός, λογικός) τις ικανότητές του (π.χ. ισχυρός, ευρηματικός, ευφυής, αποτυχημένος) και την αξιοπιστία του (π.χ. αξιόπιστος, αφοσιωμένος, συνεπής, ασυνεπής) και β) το *Κύρος(Sanction)*, που αφορά στην ακεραιότητα και την εντιμότητα ενός χαρακτήρα (π.χ. ειλικρινής, έντιμος, ευθύς, διαβρωμένος) καθώς και την κοινωνική ορθότητα (π.χ. καλός, ευγενής, δίκαιος, άξεστος, θρασύς).

3.2.1.3. Εκτίμηση

Στην κατηγορία αυτήν ανήκουν οι αξιολογικές κρίσεις που εκφράζονται για υλικά αγαθά, γεγονότα, φυσικά και κοινωνικά φαινόμενα και καταστάσεις, οι οποίες μπορούν να διακριθούν σε τρία βασικά επίπεδα: α) την *Αντίδραση(Reaction)*, που σχετίζεται με τον αντίκτυπο (π.χ. συναρπαστικός, συνταρακτικός, αδιάφορος) και την ποιότητα ενός πράγματος ή γεγονότος (π.χ. καλός, άξιος, άψογος, κακός), β) τη *Σύνθεση(Composition)*, που αναφέρεται στην ισορροπία (π.χ. αρμονικός, άτακτος, ισορροπημένος, άναρχος) και τη συνθετότητα ενός αντικειμένου ή κατάστασης (π.χ. ακριβής, επιτηδευμένος, ιδιόρρυθμος, περίπλοκος, μπερδεμένος) και γ) την *Εκτίμηση(Valuation)* που περιγράφει την αξία κάποιου πράγματος (π.χ. εξαιρετικός, ευτελής, άχρηστος).

Στην παρούσα διαδικασία ανάπτυξης ενός λεξικού όρων που εκφράζουν άποψη, το υποσύστημα *Attitude* ήταν αυτό που όρισε το βασικό πλαίσιο αναζήτησης και προσδιορισμού του *αξιολογικού δυναμικού* κάθε όρου, με το σκεπτικό ότι κάθε γνώμη μπορεί να έχει την εξής μορφή: α) να σχολιάζει ένα πράγμα (*appreciation*), β) να κρίνει ένα πρόσωπο (*judgment*) ή γ) να εκφράζει προσωπική συναισθηματική κατάσταση (*affect*) (Taboada, M. & Grieve, J. 2006). Θα εστιάσουμε κυρίως στις κατηγορίες *judgment* (κριτική) και *appreciation* (εκτίμηση), όροι υπό το πρίσμα των

οποίων, θα υλοποιηθεί τόσο η συγκέντρωση των λέξεων όσο και η αξιολόγηση και κατηγοριοποίησή τους.



Πίνακας 2. Σχηματική απεικόνιση του υποσυστήματος Attitude. (Read, J., Carroll, J. 2010).

3.3. Μεθοδολογία

Από τα πρώτα στάδια σχεδιασμού και οργάνωσης αυτού του λεξικού, έγινε αντιληπτό ότι η απλή καταγραφή όρων (επιθέτων) με *a priori* (Wilson et al., 2005) προσδιορισμένη πολικότητα (θετική ή αρνητική) δεν θα ήταν αρκετή για τη δημιουργία ενός αξιόπιστου γλωσσικού πόρου που θα μπορεί να αξιοποιηθεί σε ένα αυτοματοποιημένο σύστημα εξόρυξης γνώμης. Τα περισσότερα λεξικά που χρησιμοποιούνται σε συστήματα συναισθηματικής ανάλυσης γραπτού λόγου, αποτελούνται κυρίως από λίστες λέξεων και όχι σημασιών. Ωστόσο, πολλές λέξεις έχουν τόσο υποκειμενική όσο και αντικειμενική σημασία (Liu, 2012). Συνεπώς, τα συστήματα αυτά συχνά καταλήγουν στην εξαγωγή αριθμητικών συμπερασμάτων, συχνά εσφαλμένων και παραπλανητικών, κυρίως γιατί η *υποκειμενικότητα* και η *πολικότητα* των λέξεων συνιστούν ιδιότητες που τις περισσότερες φορές εξαρτώνται από το κειμενικό πλαίσιο στο οποίο εντάσσονται και δεν μπορούν να προσδιοριστούν εκ των προτέρων (Hatzivasiloglou & McKeown, 1997). Από την άλλη πλευρά,

τέτοιου είδους προσεγγίσεις, δεν μπορούν πάντα να αποτυπώσουν με ακρίβεια τις πολυποίκιλλες σημασιολογικές και υφολογικές διαστάσεις του ανθρώπινου λόγου, ιδιαίτερα όταν αυτός αντικατοπτρίζει τις πολυσύνθετες εκφάνσεις του ψυχικού κόσμου(συναισθήματα, αντιλήψεις, στάσεις κ.λ.π). Για παράδειγμα, το επίθετο *τυφλός* κατά την κυριολεκτική χρήση του, ως ουσιαστικοποιημένου επιθέτου, δηλώνει μια αντικειμενική πληροφορία για το πρόβλημα όρασης κάποιου ανθρώπου, στη φράση *τυφλή εμπιστοσύνη* χρησιμοποιείται μεταφορικά εκφράζοντας τη θετική στάση κάποιου ανθρώπου προς ένα πρόσωπο ή ένα αντικείμενο, ενώ στη φράση *τυφλό χτύπημα* το ίδιο επίθετο χρησιμοποιείται μεταφορικά εκφράζοντας όμως αρνητική έννοια και προκαλώντας ανάλογα συναισθήματα.

Από τις παραπάνω διαπιστώσεις έγινε σαφές ότι, για να έχουμε περισσότερο ακριβή και έγκυρα αποτελέσματα από τη χρήση ενός λεξικού κατά την κειμενική ανάλυση, θα έπρεπε να ληφθεί υπ'οψιν όχι μόνο ο αρχικός σημασιολογικός προασανατολισμός αλλά και η *συμφραστική πολικότητα (contextual polarity)* κάθε όρου, η οποία δεν συμπίπτει πάντα με την *a priori* πολικότητα (Wilson et al., 2005). Με τον όρο *συμφραστική πολικότητα*, εννοούμε το σημασιολογικό προασανατολισμό που αποκτά μια λέξη εντός του κειμενικού περιβάλλοντός της. Στην περίπτωση αυτή, δεν αρκεί μόνο η διερεύνηση της αρχικής λεξικής σημασίας κάποιου όρου αλλά και οι ιδιαίτερες σημασιολογικές αποχρώσεις που αποκτά κατά την εμφάνισή του σε κάποιο συμφραστικό πλαίσιο.

Συνεπώς, επιλέχθηκε ένας χειροκίνητος τρόπος ανάπτυξης και διαδοχικής επέκτασης του λεξικού, που περιελάμβανε συλλογή λέξεων και αξιολόγησή τους ως προς τη λειτουργία τους ως φορέων άποψης, ο οποίος υλοποιήθηκε κυρίως με δύο μέσα:

α)τη χρήση ερμηνευτικών λεξικών, εντύπων [28], και ηλεκτρονικών [3, 8, 30], τόσο για τη συλλογή των αρχικών λέξεων που θα εντάσσονταν στο λεξικό όσο και για τον προσδιορισμό του αρχικού σημασιολογικού φορτίου τους αλλά και για την διεύρυνσή του μέσω της αναζήτησης συνωνύμων και αντωνύμων των όρων.

α)την αξιοποίηση σωμάτων κειμένων για την καταγραφή του κειμενικού περιβάλλοντος κάθε όρου, για την αποτύπωση της χρήσης του σε εκφράσεις που δηλώνουν άποψη. Δόθηκε ιδιαίτερη έμφαση σε αυτόν τον τρόπο, καθώς η γλωσσική έρευνα που στηρίζεται σε σώματα κειμένων προσφέρει χρήσιμα στοιχεία για το περιεχόμενο των εξεταζόμενων όρων σε αυθεντικά κείμενα και πραγματικές

επικοινωνιακές περιστάσεις, παρέχοντας, έτσι, τη δυνατότητα εξαγωγής συμπερασμάτων με υψηλό βαθμό ακρίβειας.

Από τα ποικίλλα μέσα γλωσσικής ανάλυσης που προσφέρουν συνήθως τα σώματα κειμένων, η επεξεργασία των δεδομένων-επιθέτων βασίστηκε κυρίως σε δύο τρόπους, τη δημιουργία *συμφραστικών πινάκων (concordances)* των εξεταζόμενων όρων και την κατάρτιση *λίστας συχνότητας* των συγκεκριμένων γλωσσικών τύπων στα αντίστοιχα σώματα κειμένων. Οι *συμφραστικοί πίνακες* (λίστες αριθμημένων προτάσεων στις οποίες εμφανίζεται ο εξεταζόμενος όρος μαζί με τα συμφραζόμενά του) έδωσαν τη δυνατότητα καταγραφής των συχνότερων *συνάψεων (collocations)* κάθε επιθέτου. Με τον όρο *σύναψη*, ο Sinclair (1991) (όπως αναφέρει η Φραγκάκη, 2010) ορίζει τη «συνεμφάνιση δύο ή περισσότερων λέξεων σε μικρή απόσταση μεταξύ τους μέσα σε ένα κείμενο». Η καταγραφή αυτών των συνάψεων είχε ως στόχο τη συγκέντρωση συντακτικής και σημασιολογικής πληροφορίας για κάθε ένα επίθετο, ώστε να αναδειχθεί η χρήση του κατά την εκφορά αξιολογικού λόγου.

Παράλληλα, μια σημαντική παράμετρος που πρέπει να ληφθεί υπόψιν κατά την ανάλυση του αξιολογικού λόγου είναι ο βαθμός διαφάνειας της αξιολόγησης, κατά πόσο, δηλαδή, πρόκειται για *άμεση* ή *έμμεση* έκφραση γνώμης. Ως *άμεση (explicit)* άποψη θεωρείται αυτή που δηλώνεται εμφανώς σε μια *υποκειμενική πρόταση*, η οποία περιέχει αντικειμενικές πληροφορίες για τον κόσμο, ενώ *έμμεση (implicit)* αυτή η οποία εμπεριέχει προσωπικά συναισθήματα ή πεποιθήσεις και υπολανθάνει σε μια *αντικειμενική πρόταση* (Liu, 2010). Στην παρούσα ερευνητική εργασία εστιάζουμε κυρίως στην έκφραση εμφανούς άποψης και όχι στις περιπτώσεις που αυτή υπολανθάνει μέσω διαφόρων γλωσσικών φαινομένων (π.χ. μέσω της ειρωνίας).

Επίσης, σε αυτή τη μελέτη, κατά την αξιολόγηση της υποκειμενικότητας και της πολικότητας κάθε όρου δεν εξετάζεται το γλωσσικό φαινόμενο της άρνησης (*negation*) και η επίδρασή του στο σημασιολογικό φορτίο μιας πρότασης (π.χ. *είναι αξιόπιστη εφαρμογή/δεν είναι αξιόπιστη εφαρμογή*). Ωστόσο, κατά την καταγραφή της συγκεκριμένης πληροφορίας, συγκεντρώθηκε ικανός αριθμός γλωσσικών φαινομένων, μέσω των οποίων δηλώνεται η άρνηση, που μπορούν να αξιοποιηθούν σε μελλοντική έρευνα.

3.4. Υλοποίηση

Η πορεία σύνθεσης του λεξικού συνοψίζεται σε τέσσερις διαδοχικές φάσεις: α) την απόδοση στην ελληνική γλώσσα επιλεγμένων αγγλικών επιθέτων που εκφράζουν άποψη, β) τον έλεγχο του αξιολογικού δυναμικού κάθε επιθέτου και του βαθμού αξιοπιστίας του ως δείκτη υποκειμενικού-αξιολογικού λόγου, κυρίως βάσει του λεξιλογικού ορισμού και του κειμενικού περιβάλλοντός τους, γ) την επέκταση της αρχικής λίστας επιθέτων με την προσθήκη συγγενών εννοιολογικά (συνώνυμα) και μορφολογικά (ομόρριζα επιρρήματα) όρων και δ) την κατάταξή τους σε συγκεκριμένες ομάδες βάσει σημασιολογικών και συντακτικών κριτηρίων. Η φάση διεύρυνσης του λεξικού πρακτικά υλοποιούταν παράλληλα κατά τα δύο πρώτα στάδια, αφού συνιστούσε μια διαρκή διαδικασία εμπλουτισμού του, που διέτρεχε όλη την πορεία συλλογής λέξεων.

3.4.1. Μετάφραση αγγλικών όρων-φορέων υποκειμενικότητας

Αρχικά, επιλέχθηκε από τις διάφορες κατηγορίες του συστήματος *attitude*, όπως περιγράφεται στη θεωρία *Appraisal* (Martin and White, 2005), μια ομάδα προσδιορισμένης πολικότητας επιθέτων της αγγλικής γλώσσας, αντιπροσωπευτικών δειγμάτων των διαφόρων σημασιολογικών και υφολογικών διαστάσεων του προφορικού και γραπτού λόγου, που λειτούργησαν ως σπόροι όρων (*seed terms*) (Whitelaw et al., 2005). Στη συνέχεια, με τη χρήση των ηλεκτρονικών λεξικών *word reference* [8] και *Collins dictionary* [3] τα αγγλικά επίθετα αποδόθηκαν σε ανάλογους ελληνικούς όρους ποικίλων σημασιολογικών αποχρώσεων, με μέση αναλογία 1:3, διευρύνοντας έτσι το σώμα του λεξικού ποσοτικά και ποιοτικά. Κατά τη διαδικασία μετάφρασης, δόθηκε ιδιαίτερη βαρύτητα στη συγκέντρωση ικανού αριθμού συνωνύμων για κάθε ελληνικό όρο, ώστε να προκύψει μια ευρεία ομάδα επιθέτων, διαβαθμισμένης σημασιολογικής έντασης αλλά και ποικίλων υφολογικών επιπέδων. Για παράδειγμα, για το επίθετο *big* χρησιμοποιήθηκαν οι αντίστοιχοι ελληνικοί όροι *μεγάλος*, *δυνατός*, *ισχυρός*, *σημαντικός*, που απεικονίζουν τις διαφορετικές χρήσεις της λέξης, το επίθετο *chesty* αποδόθηκε στα αντίστοιχα ελληνικά επίθετα *αλαζονικός*, *υπερφίαλος*, *υπεροπτικός*, *ματαιόδοξος*, *φαντασμένος*, *ξίπασμένος*, ώστε να υπάρχει ποικιλία ύφους (επίσημο, ανεπίσημο, οικείο, υβριστικό), ενώ το επίθετο *excruciating* μεταφράστηκε στα αντίστοιχα ελληνικά επίθετα *έντονος*, *δριμύς*, *σφοδρός*,

ανυπόφορος και *αβάσταχτος* που αποτυπώνουν την κλιμάκωση στην ένταση της έννοιας.

Παράλληλα, η διαδικασία απόδοσης των επιθέτων από τη μία γλώσσα στην άλλη, λειτούργησε και ως ένα αρχικό στάδιο ελέγχου του ρόλου τους ως δεικτών υποκειμενικότητας, αφού η επιλογή των ελληνικών όρων πραγματοποιήθηκε υπό το πρίσμα της χρήσης τους ως αξιολογικών στοιχείων, εστιάζοντας πολλές φορές στη μεταφορική σημασία τους. Έτσι, για τους αγγλικούς όρους που εμφανίζουν ταυτόχρονα δηλωτική και συνυποδηλωτική σημασία επιλέχθηκαν οι αντίστοιχοί τους ελληνικοί που χρησιμοποιούνται είτε κυριολεκτικά ή μεταφορικά αποκλείοντας την αμιγώς δηλωτική σημασία. Π.χ. για το επίθετο *acid*, δεν συμπεριλήφθηκε η ερμηνεία *όξινος* (κυριολεκτική σημασία) αλλά οι ερμηνείες *οξύς*, *δριμύς*, *καυστικός*, *ξινός* που αποδίδουν τόσο τη λογική όσο και τη συγκινησιακή χρήση της γλώσσας, η οποία μας ενδιαφέρει περισσότερο για τη διερεύνηση στάσεων και απόψεων.

3.4.2. Αξιολόγηση Υποκειμενικών Όρων

Στο στάδιο αυτό, επιχειρήθηκε μια συστηματική διαδικασία ελέγχου της υποκειμενικότητας (*subjectivity*) και της πολικότητας (*polarity*), συμφραστικής και *a priori* των ελληνικών όρων, που προέκυψαν από το προηγούμενο στάδιο της μετάφρασης. Μία λέξη ή φράση μπορεί να χαρακτηριστεί ως *υποκειμενική έκφραση* (*subjective expression*), όταν χρησιμοποιείται για να εκφράσει θετική ή αρνητική γνώμη, συναίσθημα, στάση, σκέψη κ.λ.π. (Wilson et al., 2005). Συνεπώς, το πρώτο βήμα ήταν να εξεταστεί αν τα συγκεκριμένα επίθετα χρησιμοποιούνταν αυτοτελώς ή περιέχονταν σε τέτοιου είδους *υποκειμενικές εκφράσεις* και στη συνέχεια να διερευνηθεί η σταθερότητα της υποκειμενικότητάς τους, αν δηλαδή εμφανίζονταν πάντα ως φορείς υποκειμενικής έκφρασης (θετικής ή αρνητικής) ή αν είχαν και ουδέτερη χρήση (π.χ. *βαρύ φορτίο/ύφος*). Επίσης, εξετάστηκε το είδος του προσανατολισμού των επιθέτων ως προς το δίπολο θετικό-αρνητικό και αντίστοιχα η σταθερότητα αυτού του προσανατολισμού, αν έχουν, δηλαδή, πάντα θετικό ή αρνητικό σημασιολογικό φορτίο σε σχέση με την περίπτωση χρήσης τους (π.χ. *φοβερό κινητό/έγκλημα*).

Η πορεία διερεύνησης των υποκειμενικών όρων υλοποιήθηκε σε δύο φάσεις: α) τη διερεύνηση του σημασιολογικού ορισμού τους σε λεξικά και β) την καταγραφή του συμφραστικού περιβάλλοντός τους.

α) Αρχικά, επιχειρήθηκε μια λεπτομερής εξέταση της σημασιολογικής ερμηνείας κάθε επιθέτου, όπως αποτυπώνεται σε Ερμηνευτικά-Σημασιολογικά Λεξικά. Με τη χρήση των λεξικών, ηλεκτρονικής (Τριανταφυλλίδη, 1998) και έντυπης (Μπαμπινιώτη, 2006) μορφής εξετάστηκε ο ορισμός κάθε επιθέτου, συμπεριλαμβανομένων και των σημασιολογικών και πραγματολογικών πληροφοριών που τον συνοδεύουν, ώστε να προκύψει ένα πρώτο στάδιο αξιολόγησης των όρων ως προς το κριτήριο της πολικότητας. Εκτός από την αξιοποίηση των παραδειγμάτων χρήσης κάθε επιθέτου, που παρείχαν σημαντική πληροφορία για τις συνάψεις του, καταγράφηκαν επίσης τα συνώνυμα και τα αντώνυμά τους, τα οποία ακολούθως ελέγχονταν ως προς τα ίδια κριτήρια (*υποκειμενικότητα/πολικότητα*) και προσθέτονταν στην αρχική λίστα. Αυτός ο μη αυτοματοποιημένος τρόπος συγκέντρωσης όρων για το λεξικό, αφ' ενός αποδείχτηκε ιδιαίτερα χρονοβόρος, αφ' ετέρου όμως ελαχιστοποιούσε τα περιθώρια εσφαλμένων αποτελεσμάτων σε σχέση με μια αυτοματοποιημένη μέθοδο, καθώς η αναζήτηση των νέων όρων υλοποιούνταν υπό την τήρηση συγκεκριμένων κριτηρίων. Εξάλλου, λεξικά όπως το WordNet, που θα μπορούσαν να χρησιμοποιηθούν για την αυτοματοποιημένη διεύρυνση της αρχικής ομάδας λέξεων με συνώνυμα και αντώνυμα, δεν υποστηρίζουν την ελληνική γλώσσα.

β) Το κειμενικό περιβάλλον κάθε όρου αναζητήθηκε τόσο σε σώματα κειμένων όσο και σε μηχανές αναζήτησης. Συγκεκριμένα, χρησιμοποιήθηκαν ο Εθνικός Θησαυρός Ελληνικής Γλώσσας (ΙΕΛ), το Σώμα Νέων Ελληνικών Κειμένων από την Πύλη για την Ελληνική Γλώσσα (ΚΕΓ) και το Σώμα Κειμένων GkWaC από το σύστημα αναζήτησης σωμάτων κειμένων SketchEngine, καθώς και η μηχανή αναζήτησης Google μέσω των οποίων πραγματοποιήθηκε μια συστηματική καταγραφή των συνάψεων κάθε επιθέτου κατά τη χρήση τους στο γραπτό λόγο. Η διερεύνηση των όρων, μέσω των μηχανών αναζήτησης, στο διαδίκτυο (web) και ιστολόγια (blogs), επιλέχθηκε κυρίως γιατί σε αυτά αποτυπώνεται η τρέχουσα χρήση της γλώσσας και υποστηρίζουν έτσι μια περισσότερο συγχρονική προσέγγιση αποτύπωσης των γλωσσικών φαινομένων που μας ενδιαφέρουν.

3.4.2.1. Ισχύς Υποκειμενικότητας

Η καταγραφή αυτή των συνάψεων των επιθέτων με άλλους όρους, κυρίως ουσιαστικά και επιρρήματα, επικεντρώθηκε στη χρήση τους κατά την εκφορά αξιολογικού λόγου χωρίς να αποκλειστεί και η χρήση τους σε ουδέτερου περιεχομένου εκφράσεις, προκειμένου να ελεγχθεί η σταθερότητα, η ισχύς της υποκειμενικότητάς τους. Για παράδειγμα, από την αναζήτηση του κειμενικού περιβάλλοντος του επιθέτου *ευγενής* προέκυψε ότι συνήθως προσδιορίζει τις λέξεις *καλοσύνη, κυρία, άνθρωπος, καταγωγή, πρόθεση, προσφορά, χορηγία, τρόπος, αίσθημα, άμιλλα, παραχώρηση, υλικό, αέριο*, ενώ αρκετά συχνά χρησιμοποιείται, ιδιαίτερα σε κείμενα ιστορικού χαρακτήρα, ως ουσιαστικοποιημένο επίθετο για να δηλώσει την αντίστοιχη κοινωνική τάξη μιας ιστορικής περιόδου. Στην περίπτωση αυτή, όπως και στις φράσεις *ευγενές αέριο/υλικό* το επίθετο έχει ουδέτερο σημασιολογικό φορτίο και δεν αντικατοπτρίζει υποκειμενικό λόγο, ενώ στις φράσεις *ευγενής άνθρωπος/τρόπος* έχει θετικό σημασιολογικό φορτίο και χρησιμοποιείται ως αξιολογικός όρος. Από την άλλη πλευρά, αναδείχτηκε μια κατηγορία επιθέτων, με θετικό ή αρνητικό υποκειμενικό χαρακτήρα, τα οποία διατηρούν τον αρχικό σημασιολογικό φορτίο σε όλες τις συνάψεις τους, εμφανίζοντας έτσι ταύτιση *a priori* και συμφραστικής πολικότητας, όπως π.χ. *αζιέπαινος, βαρετός, θαυμάσιος* κ.λ.π. Δεν αναφερόμαστε, βέβαια, στις περιπτώσεις κατά τις οποίες υπάρχει στην πρόταση άρνηση ή υπολανθάνουσα άποψη π.χ. μέσω της ειρωνίας, γλωσσικά φαινόμενα τα οποία δεν εξετάζονται στην παρούσα μελέτη.

3.4.2.2. Προσδιορισμός Σημασιολογικού Προσανατολισμού

Το δεύτερο επίπεδο διερεύνησης του αξιολογικού δυναμικού του κάθε επιθέτου, που στην ουσία, συντελούνταν παράλληλα προς το προηγούμενο περιελάμβανε τον ακριβή προσδιορισμό της πολικότητας για κάθε επίθετο, αν εκφράζει δηλαδή, πάντα θετική ή αρνητική σημασία, ή αν αυτή διαφοροποιείται σε διαφορετικά συμφραστικά πλαίσια. Στη φάση αυτή, τηρήθηκε η εξής γραμμή: για τις περιπτώσεις όπου ένα επίθετο εμφανιζόταν τουλάχιστον σε μία *υποκειμενική έκφραση* έχοντας θετικό ή αρνητικό προσανατολισμό, (οπότε είχε ήδη εξασφαλιστεί το πρώτο κριτήριο ότι αποτελεί *polarword*), χαρακτηριζόταν με την θετική ή αρνητική πολικότητα της συγκεκριμένης έκφρασης με την ετικέτα *positive* ή *negative* π.χ.

απολαυστικός: *positive*, αποκρουστικός: *negative*. Εάν εμφανιζόταν σε τουλάχιστον μία θετική και μία αρνητική υποκειμενική έκφραση, χαρακτηριζόταν ως θετική/αρνητική με την ετικέτα *positive/negative*, π.χ. δεινός ρήτορας/δεινή κατάσταση. Σε περιπτώσεις χρήσης επιθέτων με ουδέτερη ή θετική σημασία χρησιμοποιήθηκε επίσης η ετικέτα *positive*, π.χ. φωτεινό χρώμα/μυαλό, ενώ αντίστοιχα σε περιπτώσεις χρήσης με ουδέτερη ή αρνητική σημασία χρησιμοποιήθηκε η ετικέτα *negative* π.χ. αντικανονική στάθμευση/συμπεριφορά (Wilson et al., 2005).

3.4.2.3. Affective state

Το τελευταίο επίπεδο αξιολόγησης των επιθέτων σχετίζεται με τον προσδιορισμό του αξιολογικού δυναμικού τους (*evaluative potential*) (Taboada, M. & Grieve, J. 2006) με βασικές συνιστώσες τους όρους *judgment* και *appreciation* του υποσυστήματος *Attitude* της θεωρίας *Appraisal*. Με δεδομένο ότι κάθε επίθετο που συνιστά υποκειμενική έκφραση, αποτελεί ένα εν δυνάμει μέσο αξιολογικού λόγου, κρίθηκε σκόπιμο να προσδιοριστεί ο τύπος αξιολόγησης, εάν πρόκειται δηλαδή για κριτική ενός προσώπου (*judgment*) ή αξιολόγηση ενός πράγματος, προϊόντος, γεγονότος ή μιας ιδέας (*appreciation*) ή αν πρόκειται για μέσο έκφρασης μιας εσωτερικής ψυχικής κατάστασης. Καθώς η συγκεκριμένη μελέτη εστιάζει περισσότερο στην αξιολόγηση εξωτερικών οντοτήτων (προσώπων –πραγμάτων) (Whitelaw et al., 2005), δεν έχει εξεταστεί η κατηγορία *affect*, η οποία αναφέρεται σε συγκεκριμένες συναισθηματικές αντιδράσεις. Αντίθετα, χρησιμοποιήθηκε ο όρος *mood*, ο οποίος σχετίζεται με την ευρύτερη ψυχική κατάσταση, τη διάχυτη διάθεση του συγγραφέα, που έχει συνήθως χαμηλότερη ένταση και μεγαλύτερη διάρκεια από ένα συγκεκριμένο συναίσθημα που προκλήθηκε από ένα ερέθισμα (Scherer, 1984). Έτσι, χρησιμοποιήθηκαν δύο ετικέτες χαρακτηρισμού των επιθέτων: α) *attitude*, για τις περιπτώσεις επιθέτων, που μπορούν να χρησιμοποιηθούν για να εκφράσουν κριτική προς ένα άτομο ή πράγμα και β) *mood/attitude*, για τα επίθετα τα οποία μπορούν, παράλληλα με την έκφραση άποψης, να υποδηλώσουν τη στάση του ομιλητή προς το αντικείμενο σχολιασμού. Π.χ. για το επίθετο *συνεπής* (*συνεπής υπάλληλος/πολιτική/σύστημα*), επιλέχθηκε η ετικέτα *attitude*, καθώς χρησιμοποιείται για την άσκηση θετικής κριτικής προς το αντίστοιχο ουσιαστικό που συνοδεύει. Ανάλογα, για το επίθετο *απαίσιος* (*απαίσιος καιρός/θέαμα/συμπεριφορά*), επιλέχθηκε ο χαρακτηρισμός *mood/attitude*, καθώς, όταν

χρησιμοποιείται, δηλώνει αφ' ενός μια αρνητική αξιολογική κρίση, αλλά αφ' ετέρου τη δυσαρέσκεια του συγγραφέα προς το χαρακτηριζόμενο πρόσωπο ή πράγμα.

3.4.3. Επέκταση λεξικού

Όπως προαναφέρθηκε, η επέκταση της αρχικής ομάδας *υποκειμενικών όρων* δεν αποτέλεσε μια απολύτως διακριτή φάση κατά τη πορεία ανάπτυξης του λεξικού πόρου, αλλά περισσότερο μια σπειροειδούς μορφής διαδικασία διεύρυνσής του. Έτσι, η συγκέντρωση επιθέτων υλοποιήθηκε τόσο κατά το στάδιο μετάφρασης των αγγλικών όρων, με την προσθήκη συνωνύμων, όσο και κατά το στάδιο διερεύνησης της σημασιολογικής ταυτότητας και του συμφραστικού περιβάλλοντος για κάθε ένα από αυτά, με την προσθήκη συγγενών σημασιολογικά όρων. Π.χ. κατά την απόδοση του όρου *awkward*, προέκυψαν τα ελληνικά επίθετα *αδέξιος*, *άχαρος*, *άγαρμος* ενώ στη συνέχεια κατά την αναζήτηση ορισμών και συγκειμένου, προστέθηκαν τα συνώνυμα *άκομπος*, *άγαρμος* και *σκαιός* καθώς και τα αντώνυμα *κομψός* και *χαριτωμένος*.

Επίσης, κατά το στάδιο εντοπισμού του συμφραστικού πλαισίου των επιθέτων, καταγράφηκε μια ευρεία ομάδα ουσιαστικών και επιτατικών όρων (*intensifiers*), με τα οποία συνάπτονται συνήθως κατά τη χρήση τους στο λόγο. Όσον αφορά τα ουσιαστικά, καταγράφηκαν μόνο τα κοινά, συγκεκριμένα (π.χ. *άνθρωπος*, *άτομο*, *βιβλίο*, *έργο* κ.λ.π.) και αφηρημένα (π.χ. *μέθοδος*, *ζωή*, *συμπεριφορά*, κ.λ.π.) και όχι τα κύρια ονόματα.

3.4.3.1. Προσθήκη Επιρρημάτων

Κατά τη δεύτερη φάση επέκτασης του λεξικού, ακολούθησε η συγκέντρωση των ομόρριζων με τα επίθετα επιρρημάτων, για την οποία εφαρμόστηκε ανάλογη μέθοδος με αυτήν της συλλογής επιθέτων.

Η συμπερίληψη των επιρρημάτων σε αυτόν το γλωσσικό πόρο θεωρήθηκε σκόπιμη, καθώς τα επιρρήματα, ως προσδιορισμοί των ρημάτων, χρησιμοποιούνται ευρέως στο λόγο για την αξιολόγηση πράξεων, γεγονότων και εν γένει ενεργειών (π.χ. *φέρεται ανάλγητα*, *καταναλώνω άσκοπα*, *λειτουργεί άπογα* κ.λ.π.). Έτσι, θα μπορούσαν να αξιοποιηθούν στο γλωσσικό πόρο, όπως και τα επίθετα, ως δείκτες αξιολογικών

κρίσεων κατά την κειμενική ανάλυση. Επίσης, από την προηγούμενη διαδικασία καταγραφής συγκειμένου των επιθέτων, διαπιστώθηκε εκτεταμένη χρήση των επιρρημάτων, κυρίως των ποσοτικών (*πολύ, λίγο, αρκετά* κ.λ.π.), επιτατικών (*απολύτως, έντονα, δραματικά*) και τροπικών (*εντελώς, παντελώς, ολοκληρωτικά*), ως *intensifiers*. Αυτή η λειτουργία τους, μπορεί να προσφέρει επιπρόσθετη πληροφορία για το βαθμό έντασης της ιδιότητας που περιγράφει το επίθετο, συνεισφέροντας, με αυτόν τον τρόπο, στη συνολική αποτίμηση της αξιολογικής ισχύος ενός εκφωνήματος. Για παράδειγμα, στις φράσεις *εντελώς ανεύθυνη στάση, άκρως υποτιμητικός χαρακτηρισμός, ιδιαίτερος δύσκολη προσπάθεια*, το θετικό ή αρνητικό σημασιολογικό φορτίο των επιθέτων επιτείνεται από τα επιρρήματα που τα συνοδεύουν, καθιστώντας σαφέστερη τον σημασιολογικό προσανατολισμό των εκφράσεων.

Προκειμένου να συσταθεί ο κατάλογος των επιρρημάτων ακολουθήθηκε η εξής διαδικασία: Για κάθε επίθετο της ελληνικής πλέον γλώσσας προστέθηκε -στις περιπτώσεις που ήταν δυνατό- το ομόριζο επίρρημα με κατάληξη -α ή/και -ως. Η διερεύνηση των επιρρημάτων στα ηλεκτρονικά σώματα κειμένων, στα λεξικά και στο διαδίκτυο έδειξε ότι η καταγραφή και των δύο τύπων (-α/-ως) επιβαλλόταν, κυρίως για δύο λόγους:

α) οι ομιλητές της γλώσσας χρησιμοποιούν λιγότερο ή περισσότερο και τους δύο τύπους, οπότε θα έπρεπε να συμπεριληφθούν, προκειμένου να έχουμε ασφαλή αποτελέσματα σε μια εφαρμογή Ανάλυσης Συναισθήματος με κύριο εργαλείο το λεξικό. Π.χ. το επίρρημα *βαρέως* (φρ. *φέρω βαρέως*) χρησιμοποιείται εξίσου συχνά σε σχέση με το αντίστοιχο *βαριά* (φρ. *το παίρνω βαριά*) προκειμένου να περιγράψει μια δυσάρεστη συναισθηματική κατάσταση.

β) Η εναλλαγή μεταξύ του τύπου της δημοτικής και του αντίστοιχου λόγιου τύπου, αφ' ενός διαφοροποιεί σκόπιμα ή μη το υφολογικό επίπεδο του εκφωνήματος και αφ' ετέρου σε αρκετές περιπτώσεις δηλώνει σημασιολογική διαφορά. Π.χ. οι τύποι *έντονα* και *έντιμα* φανερώνουν περισσότερο οικείο ύφος σε σχέση με τους τύπους *εντόνως* και *εντίμως* αντίστοιχα, που αντικατοπτρίζουν περισσότερο επίσημο και τυπικό ύφος. Αυτή η υφολογική διαφορά μπορεί να προσφέρει και επιπρόσθετη πληροφορία τόσο για την επικοινωνιακή περίσταση και το κειμενικό είδος όπου εντάσσεται το συγκεκριμένο εκφώνημα όσο και για την «ταυτότητα» του ίδιου του συγγραφέα (*opinion holder*). Αξίζει να σημειωθεί ότι δύσχρηστοι τύποι επιρρημάτων

όπως *ακατονομάστως* και *αξιοθαυμάστως* (φρ. *λοιδορώ ακατονομάστως, παρουσιάζω αξιοθαυμάστως*) εντοπίστηκαν κατά συντριπτική πλειοψηφία σε κείμενα θρησκευτικού χαρακτήρα ή εκκλησιαστικού περιεχομένου.

Από την άλλη πλευρά, στις περιπτώσεις επιρρημάτων όπως *αδιάκριτα*(=χωρίς διακριτικότητα)/*αδιακρίτως*(=ανεξαρτήτως) κρίθηκε σκόπιμη η συμπερίληψη και των δύο τύπων, καθώς διαφοροποιούνται σημασιολογικά και πραγματολογικά και εμφανίζονται σε διαφορετικό συμφραστικό πλαίσιο (Μπαμπινιώτης, 2006).

Για την αξιολόγηση των επιρρημάτων ακολουθήθηκε ανάλογη μέθοδος αξιολόγησης υποκειμενικότητας και πολικότητας. Διερευνήθηκε μέσω λεξικών [28,30] ο σημασιολογικός ορισμός τους και καταγράφηκε το συμφραστικό τους πλαίσιο στο διαδίκτυο και σε σώματα κειμένων [9,27,29]. Ομοίως, τηρήθηκε η ίδια μέθοδος χαρακτηρισμού με αυτήν των επιθέτων, όσον αφορά τον σημασιολογικό προσανατολισμό και την ισχύ της υποκειμενικότητας του κάθε επιρρήματος.

Συγκεκριμένα, όσον αφορά την πολικότητα, χρησιμοποιήθηκαν οι ετικέτες *positive*, για επιρρήματα με θετικό προσανατολισμό (π.χ. *απολαυστικά, εγκάρδια, ευφύως*), *negative* για επιρρήματα με αρνητική σημασία (π.χ. *εχθρικά, βαρετά, επικυνδύνως*) ή *positive/negative* για επιρρήματα με θετική ή/και αρνητική σημασία (π.χ. *πουλάει/φέρεται τρελά, εξαπατώ/δηλώνω απροκάλυπτα*). Επίσης, όσον αφορά τη συναισθηματική κατάσταση(*affective state*) χρησιμοποιήθηκαν οι χαρακτηρισμοί *attitude*, για τις περιπτώσεις επιρρημάτων, που μπορούν να χρησιμοποιηθούν για να εκφράσουν κριτική προς ένα άτομο ή πράγμα (π.χ. *αδιάκριτα, αδέξια, ενάρετα*) και *mood/attitude*, για τα επιρρήματα τα οποία μπορούν να υποδηλώσουν ταυτόχρονα και τη συναισθηματική κατάσταση του ομιλητή προς το αντικείμενο αναφοράς (π.χ. *εξοργιστικά, επαχθώς, θαυμάσια*). Ακόμη, έγινε διάκριση μεταξύ των επιρρημάτων με *a priori* και εξαρτώμενη από το συγκεκριμένο υποκειμενικότητα, διαχωρίζοντας έτσι τους όρους με ισχυρό υποκειμενικό χαρακτήρα (π.χ. *αξιοθαύμαστα, στυγνά, ποταπά*) από αυτούς με αδύναμη υποκειμενική χρήση (π.χ. *τραυματίζω/πλήττω θανάσιμα, αναπαριστώ/διηγούμαι γραφικά*).

Ένα βασικό πρόβλημα που εντοπίστηκε κατά τη διερεύνηση των επιρρημάτων στα δύο από τα τρία ηλεκτρονικά σώματα κειμένων που χρησιμοποιήθηκαν (Ηλεκτρονικό Σώμα Κειμένων της Πύλης για την Ελληνική Γλώσσα και SketchEngine) ήταν η αδυναμία εξαγωγής ακριβών αποτελεσμάτων για τη συχνότητα εμφανίσεων του

εξεταζόμενου όρου, λόγω της σύμπτωσης του τύπου του επιρρήματος (σε –α) με την ονομαστική και αιτιατική πληθυντικού του ουδετέρου του αντίστοιχου επιθέτου. Στον παρακάτω πίνακα, φαίνεται ότι ο τύπος *καταπληκτικά* λειτουργεί ως επίρρημα μόνο σε τρεις από τις πέντε προτάσεις, ενώ στις υπόλοιπες λειτουργεί ως επίθετο.

Αντίθετα, η δυνατότητα διαχωρισμού των δύο τύπων ήταν εφικτή μόνο στον Εθνικό Θησαυρό της Ελληνικής Γλώσσας, λόγω των διαθέσιμων κριτηρίων αναζήτησης. Συνεπώς, η συγκέντρωση συμφραστικής πληροφορίας αποδείχτηκε μια αρκετά χρονοβόρα διαδικασία με περιορισμένα σχετικά αποτελέσματα.

καταπληκτικά (19) [<i>καταπληκτικός - A:Nnp:Anp:Vnp, καταπληκτικά - ADV</i>]
M0384 P008 L025 ...το ταλέντο ως χώρα να βγάζουμε καταπληκτικά ταλέντα και την εξαιρετική ικα...
M1053 P003 L003 ...ανική, αλλά γιατί συνδεόταν με καταπληκτικά γεγονότα". Πόσοι, όμως, από μ...
M2539 P003 L005 ... στιγμή που ο αντίπαλος έπαιξε καταπληκτικά δεν είχαμε καμία ελπίδα. Θα έλ...
M3061 P006 L018 ... σκορ της ημέρας (4-0!) και τα καταπληκτικά σε εκτέλεση τέρματα που σημείω...
M3174 P004 L006 ...του αυστραλιανού Όπεν. "Πέρασα καταπληκτικά αυτές τις μέρες. Συγκινή

Πίνακας 6. Συμφραστικός πίνακας από το Ηλεκτρονικό Σώμα Κειμένων της Πύλης για την Ελληνική Γλώσσα.

Μετά και την ολοκλήρωση των προσθηκών, συνολικά συγκεντρώθηκαν 2144 *υποκειμενικοί όροι* (810 επίθετα και 1334 επιρρήματα με καταλήξεις σε –α και –ως), ένα σύνολο λέξεων προσδιορισμένης συμφραστικής πολικότητας, τα οποία έπρεπε να ομαδοποιηθούν βάσει των κοινών τους χαρακτηριστικών ως προς τον *σημασιολογικό προσανατολισμό* και το *δυναμικό υποκειμενικότητας*, προκειμένου να αξιοποιηθούν σε ένα σύστημα εξόρυξης γνώμης.

3.4.4. Κατηγοριοποίηση όρων

Το τελευταίο στάδιο σύνθεσης του λεξικού πόρου περιελάμβανε την ταξινόμηση τόσο των συλλεχθέντων υποκειμενικών όρων όσο και της συμφραστικής πληροφορίας που καταγράφηκε για κάθε ένα από αυτούς, με βασικά κριτήρια ταξινόμησης αυτά που χρησιμοποιήθηκαν και κατά τον έλεγχο αξιολόγησης του αξιολογικού δυναμικού τους.

Συγκεκριμένα τα κριτήρια ήταν τα εξής:

- 1) Η ισχύς υποκειμενικότητας (*subjectivity*), η σταθερότητα της υποκειμενικότητας ενός όρου κατά τη χρήση του στο λόγο
- 2) Η πολικότητα του κάθε όρου (*polarity*), δηλαδή ο θετικός ή αρνητικός σημασιολογικός προσανατολισμός του
- 3) Το είδος στάσης (*attitude*) που εκφράζει, αν χρησιμοποιείται, δηλαδή, για κριτική προσώπου ή πράγματος ή για έκφραση συναισθήματος

Με βάση αυτά τα κριτήρια, κάθε επίθετο και επίρρημα προσδιορίστηκαν με ένα σετ χαρακτηριστικών που προσδιορίζουν το αξιολογικό δυναμικό του, την ισχύ υποκειμενικότητας (*subjectivity type*), την πολικότητα (*polarity*) και τον τύπο ψυχικής κατάστασης (*affective state*). Έτσι, για παράδειγμα, το επίθετο *υπέροχος* έχει λάβει τα εξής χαρακτηριστικά:

Subjectivity type: strong

Polarity: positive

Affective state: attitude

Συνολικά, ο λεξικός πόρος κατανεμήθηκε σε δύο βασικές ενότητες: α) το κυρίως μέρος, που περιλαμβάνει τους αξιολογικούς όρους (*subjectivity terms*), κατανεμημένους σε έξι κατηγορίες επιθέτων και οχτώ κατηγορίες επιρρημάτων, β) την ενότητα συγκειμένου, που περιλαμβάνει τις συχνότερες συνάψεις των

υποκειμενικών όρων και διακρίνεται σε δύο μεγάλες κατηγορίες ουσιαστικών (για τα επίθετα) και ρημάτων (για τα επιρρήματα), οι οποίες διακρίνονται σε επιμέρους υποκατηγορίες ανάλογα με την κατηγορία επιθέτων και επιρρημάτων που προσδιορίζουν αντίστοιχα.

A) Υποκειμενικοί όροι

i) Επίθετα

Πρόκειται για την ομάδα των επιθέτων-υποκειμενικών όρων, ταξινομημένων κατά αλφαβητική σειρά η οποία αποτελείται από τις εξής κατηγορίες επιθέτων:

- **Επίθετα 1:** Επίθετα με προκαθορισμένη πολικότητα και ισχυρό δείκτη υποκειμενικότητας. Πρόκειται για επίθετα, που κατά κανόνα, μεταφέρουν το σημασιολογικό τους φορτίο στις συνάψεις τους με τα ουσιαστικά, συνιστώντας, έτσι, τους ασφαλέστερους δείκτες υποκειμενικότητας σε ένα κείμενο. Διακρίνονται επιμέρους σε θετικά και αρνητικά, ως προς τον σημασιολογικό προσανατολισμό τους και σε αυτά που εκφράζουν μόνο άποψη ή και συναίσθημα. Αξίζει να σημειωθεί ότι οι περισσότεροι όροι της λίστας λαμβάνουν αρνητική τιμή ως προς την πολικότητα, π.χ. *αδιάντροπος, αδίστακτος, απεχθής, αποτρόπαιος, φρικτός* κ.α. αλλά υπάρχουν και ορισμένοι θετικού σημασιολογικού φορτίου, όπως *εξαίρετος, θαυμάσιος και γαμάτος*.
- **Επίθετα 2:** Επίθετα με προκαθορισμένη πολικότητα, αλλά εξαρτώμενο από το συγκεκριμένο δείκτη υποκειμενικότητας. Πρόκειται για τα επίθετα των οποίων ο δείκτης υποκειμενικότητας καθορίζεται είτε από το προηγούμενο επίρρημα ή/και από το χαρακτηριζόμενο ουσιαστικό είτε γενικά από το κειμενικό περιβάλλον του όρου. Έτσι, ανάμεσα στις περιπτώσεις συνάψεων όπως *νόμο αβάσιμη κατηγορία και αβάσιμος ισχυρισμός, ιεραρχικά ανώτερος υπάλληλος και ανώτερος άνθρωπος, πνευματικά δυσκίνητος άνθρωπος και δυσκίνητος μηχανισμός* παρατηρείται εμφανής διαφορά ως προς τον δείκτη υποκειμενικότητας. Διακρίνονται επιμέρους σε θετικούς και αρνητικούς όρους, ως προς τον τύπο πολικότητας και σε αυτούς που είναι μόνο φορείς γνώμης ή και συναίσθηματος.

- **Επίθετα 2n:** Επίθετα με προκαθορισμένη πολικότητα, αλλά ισχύ υποκειμενικότητας εξαρτώμενη συγκεκριμένα από το χαρακτηριζόμενο ουσιαστικό που ακολουθεί. Στις περιπτώσεις χρήσης αυτών των επιθέτων η υποκειμενικότητα καθορίζεται κυρίως από τη σύναψή τους με ουσιαστικά και όχι άλλου είδους μέρη του λόγου. Διακρίνονται επιμέρους σε θετικά και αρνητικά επίθετα, ως προς τον σημασιολογικό προσανατολισμό τους σε αυτά που εκφράζουν μόνο άποψη ή και συναίσθημα. Είναι η λίστα με τις περισσότερες εγγραφές όρων, καθώς τα περισσότερα επίθετα αυτής της κατηγορίας έχουν κυριολεκτική και μεταφορική χρήση και δεν συνιστούν πάντα ισχυρούς δείκτες υποκειμενικότητας. Π.χ. *αιματηρή συμπλοκή/αιματηρή θυσία, ευγενικός άνθρωπος/ ευγενική καταγωγή.*
- **Επίθετα 3:** Επίθετα χωρίς προκαθορισμένη πολικότητα και μεταβαλλόμενη ισχύ υποκειμενικότητας, που εξαρτώνται από το συγκείμενο του επιθέτου. Πρόκειται για επίθετα που εμφανίζουν θετικό ή αρνητικό σημασιολογικό φορτίο ανάλογα με το γλωσσικό περιβάλλον χρήσης τους και αντίστοιχα κλιμακούμενο δείκτη υποκειμενικότητας. Π.χ. *σοβαρό πρόβλημα/σοβαρός επιστήμονας, τρομερό ενδιαφέρον/τρομερό έγκλημα και φοβερή καταστροφή/φοβερή εντύπωση.*
- **Επίθετα 3i:** Επίθετα χωρίς προκαθορισμένη πολικότητα και κλιμακούμενο δείκτη υποκειμενικότητας, χαρακτηριστικά που εξαρτώνται από το συγκείμενο του επιθέτου και λειτουργούν κυρίως ως επιτατικά επίθετα (*intensifiers*). Πρόκειται για μια ιδιαίτερη κατηγορία επιθέτων, καθώς χωρίς να διαθέτουν συγκεκριμένη θετική ή αρνητική εννοιολογική απόχρωση, επιτείνουν έντονα το σημασιολογικό φορτίο του ουσιαστικού που χαρακτηρίζουν. Π.χ. *ιλιγγιάδης ανάπτυξη/ποσό και σφοδρός πολέμιος/έρωτας.*
- **Επίθετα 4:** Επίθετα με ουδέτερο αρχικό σημασιολογικό φορτίο και εξαρτώμενο δείκτη υποκειμενικότητας. Πρόκειται για επίθετα που ενώ δεν έχουν σαφή θετική ή αρνητική τιμή πολικότητας, εμφανίστηκαν κατά τις συνάψεις τους να χρησιμοποιούνται κατά την αξιολόγηση προσώπου ή

πράγματος και την έκφραση συναισθημάτων, π.χ. *παραδοσιακές αρχές*(θετική έννοια), *διφορούμενος άνθρωπος* (αρνητική έννοια), *υποκειμενική κρίση* (αρνητική έννοια).

	Positive	Negative	Positive/Negative
Επίθετα 1	21,57%	78,43%	
Επίθετα 2	40,00%	60,00%	
Επίθετα 2n		54,99%	
Επίθετα 3			11,87%
Επίθετα 3i			2,47%
Επίθετα 4			1,24%

Πίνακας 7. Ποσοστιαία κατανομή επιθέτων

ii) Επιρρήματα:

Η οργανωμένη κατά αλφαβητική σειρά λίστα επιρρημάτων με καταλήξεις σε –α και –ως, ομόριζων των επιθέτων, η οποία αποτελείται από τις εξής ομάδες:

- **Επιρρήματα 1:** Επιρρήματα με προκαθορισμένη πολικότητα και ισχυρό δείκτη υποκειμενικότητας. Πρόκειται για επιρρήματα, που όπως και τα αντίστοιχα επίθετα, κατά κανόνα μεταφέρουν το σημασιολογικό τους φορτίο στις συνάψεις τους με τα ρήματα, συνιστώντας, έτσι, σημαντικούς φορείς υποκειμενικότητας. Οι περισσότεροι όροι της λίστας έχουν αρνητική πολικότητα, ενώ συχνά συνοδεύονται από ποσοτικά και επιτακτικά επιρρήματα π.χ. *πολύ, εντελώς, αρκετά* που ενισχύουν ακόμα περισσότερο τη σημασία τους. Π.χ. *φέρομαι ανήθικα, παραβιάζω αναίσχυντα, αισθάνομαι απαίσια*.
- **Επιρρήματα 1i:** Επιρρήματα με προκαθορισμένη πολικότητα και ισχυρό δείκτη υποκειμενικότητας, που λειτουργούν συνήθως και ως intensifiers. Πρόκειται για την κατηγορία επιρρημάτων συχνότατης χρήσης κατά την έκφραση γνώμης, καθώς συνοδεύουν όχι μόνο ρήματα και επίθετα αλλά και

ολόκληρες φράσεις επιτείνοντας το σημασιολογικό τους φορτίο. Π.χ. *πονάω* *αφόρητα/αφόρητα βαρετός, περνάω* *θεικά/θεικά όμορφος*.

- **Επιρρήματα 2:** Επιρρήματα με προκαθορισμένη πολικότητα και εξαρτώμενο από το συγκείμενο δείκτη υποκειμενικότητας. Τα επιρρήματα αυτά δεν λειτουργούν πάντα ως ισχυροί φορείς άποψης αυτόνομα, αλλά σε σχέση με το ρήμα που χαρακτηρίζουν, κυρίως λόγω της κυριολεκτικής και μεταφορική χρήσης των περισσοτέρων. Π.χ. *στέκομαι/συμπεριφέρομαι άκαμπτα, νιώθω/φέρομαι ζεστά*.
- **Επιρρήματα 2i:** Επιρρήματα με προκαθορισμένη πολικότητα και εξαρτώμενο από το συγκείμενο δείκτη υποκειμενικότητας, που λειτουργούν και ως *intensifiers*. Πρόκειται για επίσης ευρείας χρήσης επιρρήματα κατά την έκφραση άποψης, καθώς συνοδεύουν επίθετα και ρήματα, αλλά δεν συνιστούν πάντα ασφαλείς δείκτες υποκειμενικότητας, η οποία καθορίζεται συνήθως από το συγκείμενό τους. Π.χ. *πληγώνω/κοιμάμαι βαριά και βαριά κοινότοπος, νιώθω παγερά/παγερά αδιάφορος*.
- **Επιρρήματα 3 :** Επιρρήματα χωρίς σταθερή τιμή πολικότητας και δείκτη υποκειμενικότητας εξαρτώμενο από το συγκείμενο. Πρόκειται για ιδιαίτερη κατηγορία επιρρημάτων, καθώς, παρουσιάζουν διαφορές σε σχέση με τα ομόριζά τους επίθετα όσον αφορά τον σημασιολογικό προσανατολισμό. Έτσι, ενώ τα περισσότερα από αυτά έχουν έντονο σημασιολογικό φορτίο, δεν είναι σταθερής πολικότητας, καθώς αυτή καθορίζεται κυρίως από το ρήμα που προσδιορίζουν, πράγμα που καθιστά αναγκαία την εξέταση και του συγκειμένου τους κατά την αναζήτηση άποψης σε ένα κείμενο (π.χ. *βλάπτω/σκέφτομαι σοβαρά, εμπιστεύομαι/υπακούω τυφλά*).
- **Επιρρήματα 3i:** Επιρρήματα χωρίς σταθερή τιμή πολικότητας και δείκτη υποκειμενικότητας εξαρτώμενο από το συγκείμενο, που λειτουργούν και ως *intensifiers*. Επίσης ιδιαίτερη κατηγορία επιρρημάτων, με ισχυρό διπολικό σημασιολογικό φορτίο που εξαρτάται κυρίως από τα χαρακτηριζόμενα επίθετα και ρήματα. Π.χ. *λυπάμαι/χαίρομαι απερίγραπτα*,

ερωτεύομαι/σκέφτομαι τρελά, εκνευρίζομαι/περνάω φοβερά και φοβερά ρομαντικός/εχθρικός.

- **Επιρρήματα 4:** Επιρρήματα χωρίς συγκεκριμένη τιμή πολικότητας και δείκτη υποκειμενικότητας. Πρόκειται για επιρρήματα που ενώ δεν παρουσιάζουν αρχικά κάποια συγκεκριμένη σημασιολογική απόχρωση, συχνά χρησιμοποιούνται στην αξιολόγηση πράξεων. Π.χ *εξετάζω αναλυτικά, μιλάω διαφορούμενα, σκέφτομαι παραδοσιακά.*
- **Επιρρήματα 4i:** Επιρρήματα με αρχικά ουδέτερη πολικότητα που λειτουργούν κυρίως ως *intensifiers*. Είναι επιρρήματα που κυρίως ενισχύουν το σημασιολογικό φορτίο των χαρακτηριζόμενων όρων. Π.χ. *συμφωνά απόλυτα/απόλυτα κατανοητός, καταστρέφω ολοκληρωτικά/ολοκληρωτικά ανίκανος, ικανοποιώ πλήρως/πλήρως κατανοητός.*

	Positive	Negative	Positive/Negative
Επιρρήματα 1	15,60%	84,40%	
Επιρρήματα 1i	20,51%	79,49%	
Επιρρήματα 2	46,76%	53,24%	
Επιρρήματα 2i	47,25%	52,75%	
Επιρρήματα 3			5,95%
Επιρρήματα 3i			7,80%
Επιρρήματα 4			2,47%
Επιρρήματα 4i			2,01%

Πίνακας 8. Ποσοστιαία κατανομή επιρρημάτων

B) Συγκείμενο

Στην ενότητα αυτή παρατίθενται ανά όρο οι καταγεγραμμένες συχνότερες συνάψεις κάθε επιθέτου. Έχει γίνει βασική κατηγοριοποίηση σε ουσιαστικά, επίθετα και ρήματα και επιμέρους ταξινόμηση ανά πολικότητα (θετική/αρνητική) με ποσοτικά στοιχεία και ποσοστιαία κατανομή επί του συνόλου της κάθε κατηγορίας. Οι κατηγορίες είναι οι εξής:

- **Ουσιαστικά:** Η λίστα ουσιαστικών που εμφανίστηκαν ως οι συχνότεροι προσδιοριζόμενοι όροι από κάθε κατηγορία επιθέτων. Αποτελείται κυρίως από κοινά ονόματα, συγκεκριμένα (π.χ. *άνθρωπος, άτομο, κινητό*) και αφηρημένα (*συμπεριφορά, γεγονός, πολιτική*). Η λίστα αυτή μπορεί να αξιοποιηθεί και σε μια εφαρμογή εξόρυξης χαρακτηριστικού (*feature extraction*) που θα εστιάζει στη συχνότητα εμφάνισης των ουσιαστικών ή των ονοματικών φράσεων που τα περιέχουν. Η λίστα έχει καταρτιστεί με το σκεπτικό ότι εφ'όσον κάθε γνώμη που εκφράζεται σε ένα κείμενο έχει έναν στόχο-αντικείμενο αναφοράς, είναι σημαντικό να αναγνωρίζεται σε ένα κείμενο όχι μόνο κάθε έκφραση άποψης αλλά και το αντικείμενο που προσδιορίζει (Liu, 2012).
- **Ρήματα:** Η ταξινομημένη ανά κατηγορία επιρρημάτων λίστα ρημάτων που συνήθως συνοδεύονται από αυτά. Πρόκειται για τα ρήματα που εμφανίζουν τη μεγαλύτερη συχνότητα συνεμφάνισης με τα επιρρήματα που έχουν συμπεριληφθεί ως υποκειμενικοί όροι στο λεξικό. Τα ρήματα λόγου (*απαντώ, ερωτώ, δηλώνω* κ.λ.π.), συναισθημάτων (*νιώθω, αισθάνομαι, χαίρομαι* κ.λ.π.) και σκέψης-κρίσης (*θεωρώ, νομίζω, πιστεύω* κ.λ.π.) είναι αυτά με τις συχνότερες συνάψεις με τα επιρρήματα που έχουν καταγραφεί.

4. Συμπεράσματα

Με την παρούσα εργασία πραγματοποιήθηκε η συγκέντρωση μιας λίστας επιθέτων και επιρρημάτων-φορέων υποκειμενικού λόγου και η κατηγοριοποίησή τους με βασικά κριτήρια τον βαθμό υποκειμενικότητας (*subjectivity*) και την τιμή

πολικότητάς (*polarity*) τους, με σκοπό να αξιοποιηθεί ως γλωσσικός πόρος σε εφαρμογές *εξόρυξης γνώμης*.

Η μεθοδολογική προσέγγιση που ακολουθήθηκε, η αναζήτηση της *συμφραστικής πολικότητας* των εξεταζόμενων όρων σε πραγματικά δεδομένα σωμάτων κειμένων, έδωσε τη δυνατότητα να μελετηθεί μεγάλος αριθμός επιθέτων και επιρρημάτων και να διερευνηθεί ο ρόλος τους ως φορέων άποψης, στάσης και συναισθημάτων σε ένα κείμενο.

Παράλληλα, όμως, ανέδειξε και τη δυσκολία κατηγοριοποίησης των όρων αυτών σε αυστηρά οριοθετημένες ομάδες λέξεων κυρίως λόγω της αμφισημίας της γλώσσας και της σημασιολογικής ποικιλίας που αποκτά κάθε λέξη σε σχέση με το συμφραστικό της περιβάλλον. Συνεπώς, ειδικά τα επίθετα και τα επιρρήματα με εξαρτώμενο δείκτη υποκειμενικότητας και τιμή πολικότητας χρειάζονται περισσότερη σημασιολογική πληροφορία, ώστε να μπορέσουν να αξιοποιηθούν αποτελεσματικά με την παράλληλη χρήση συντακτικών κανόνων (*patterns*) είτε σε συστήματα *μη επιβλεπόμενης εκμάθησης μηχανής* (*unsupervised machine learning*) είτε σε εφαρμογές *επιβλεπόμενης εκμάθησης μηχανής* (*supervised machine learning*).

Η θεωρία *appraisal* έθεσε το πλαίσιο αναζήτησης και κατηγοριοποίησης των επιθέτων και επιρρημάτων σε σχέση με τον τύπο άποψης που εκφράζεται σε ένα κείμενο (συναισθηματική κατάσταση, κριτική προσώπων ή αξιολόγηση πραγμάτων) δίνοντας τη δυνατότητα να εξεταστεί κάθε όρος σε διαφορετικές περιπτώσεις εκφοράς αξιολογικού λόγου και να χαρακτηριστεί κάθε όρος με ένα σετ χαρακτηριστικών ως προς την υποκειμενικότητα (*subjectivity type*) την πολικότητα (*polarity*) και τον τύπο άποψης (*affective state*)

5. Συζήτηση-Μελλοντικές Προεκτάσεις

Αναμφισβήτητα, η δημιουργία ενός λεξικού πόρου μπορεί να λειτουργήσει ως ένα πρώτο στάδιο προσέγγισης της αναγνώρισης και της κατηγοριοποίησης των γλωσσικών φαινομένων μέσω των οποίων εκφέρεται η άποψη σε γραπτά κείμενα. Ωστόσο, η χρήση ενός λεξικού, ακόμα και με εξειδικευμένη πληροφορία για τη χρήση κάθε λέξης σε συγκεκριμένα φραστικά πλαίσια, δεν μπορεί να εξασφαλίσει απόλυτη ακρίβεια στην εξόρυξη γνώμης και συναισθημάτων, καθώς δεν μπορεί να

περιέχει όλους τους δυνατούς τρόπους με τους οποίους μπορούν αυτά να εκφραστούν. Και αντίστροφα, δεν είναι βέβαιο ότι ένας όρος που συμπεριλαμβάνεται ως υποκειμενικός (*polar word*), θα εκφράζει άποψη σε κάθε κείμενο που εμφανίζεται (π.χ. *θέλω να αγοράσω ένα καλό βιβλίο*).

Στις περιπτώσεις αυτές, χρειάζεται να δομηθούν συγκεκριμένοι συντακτικοί-σημασιολογικοί κανόνες, η αξιοπιστία των οποίων πρέπει να ελεγχθεί σε σώματα κειμένων (*testing corpus*) ώστε σε συνδυασμό με τη χρήση του λεξικού πόρου να μπορούν να αξιοποιηθούν σε κάποια κατάλληλα σχεδιασμένη εφαρμογή *Ανάλυσης Συναισθήματος*.

Επίσης, καλό θα ήταν να γίνει και ένα δεύτερο στάδιο χαρακτηρισμού των όρων από διαφορετικό ερευνητή, ώστε να διασταυρωθεί η ορθότητα των χαρακτηριστικών κάθε επιθέτου και επιρρήματος υπό το πρίσμα διαφορετικού γλωσσικού αισθητηρίου.

Όσον αφορά τη θεωρία *appraisal*, και τη μοντελοποίηση που όρισε σε σχέση με την έκφραση γνώμης, θα μπορούσε σε μελλοντικό βήμα ενασχόλησης με το θέμα, να αξιοποιηθούν επιπλέον και οι παράγοντες του φορέα της άποψης (*appraiser*) αλλά και του αντικειμένου αναφοράς της (*appraised*), ώστε να προκύψει ένα πληρέστερο σετ χαρακτηριστικών για κάθε όρο.

6.Βιβλιογραφία:

Ξενόγλωσση:

- [1] Baccianella S., Essuli, A. and Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of LREC 10.
- [2] Bradley, M.M., & Lang, P.J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida
- [3] *Collins dictionary*, διαθέσιμο στην ιστοσελίδα <http://www.collinsdictionary.com/translator> Τελευταία πρόσβαση στις 15 Μαρτίου 2013
- [4] *General Inquirer*, διαθέσιμο στην ιστοσελίδα <http://www.wjh.harvard.edu/~inquirer/> Τελευταία επίσκεψη στις 20 Ιουνίου 2013
- [5] Hatzivassiloglou, V. and McKeown K. R., (1997). *Predicting the semantic orientation of adjectives*. In Proceedings of ACL 1997.
- [6] Hatzivassiloglou, V. and Wiebe J., (2000). *Effects of Adjective Orientation and Gradability on Sentence Subjectivity*. In Proceedings of the international Conference on Computational Linguistics (COLING, 2000)
- [7] Hu, M. and Liu, B. (2004). *Mining and Summarizing Customer Reviews*. ACM SIGKDD-2
- [8] Kellog, M, *Word Reference* (1999), Διαθέσιμο στην ιστοσελίδα <http://www.wordreference.com/http://translate> Τελευταία πρόσβαση στις 20 Μαρτίου 2013
- [9] Kilgarriff, and A. Pavel, R., *Sketch Engine*, διαθέσιμο στην ιστοσελίδα <http://www.sketchengine.co/> Τελευταία πρόσβαση στις 20 Ιουνίου 2013
- [10] Liu, B., (2010). *Sentiment Analysis and Subjectivity*, An chapter in Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkha and F. J. Damerau).
- [11] Liu, B., (2012). *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers.
- [12] Marcus, M.P., Marcinkiewicz, M.A. & Santorini, B. (1993) *Building a Large Annotated Corpus of English: The Penn Treebank*, Association for Computational Linguistics
- [13] Martin, J.R. and White, P. (2005). *The Language of Evaluation, Appraisal in English*, Palgrave Macmillan, London, New York.
- [14] Nasukawa, T. and Yi, J. (2003). *Sentiment Analysis : Capturing Favorability Using Natural Language Processing*, K-CAP 2003, Florida, USA.
- [15] Nielsen, F. A. (2011). *A new ANEW: Evaluation of a word list for sentiment analysis in microblogs* DTU Informatics, Technical University of Denmark, Lyngby, Denmark.
- [16] Pang B., Lee, L. (2008). *Opinion Mining and Sentiment Analysis*, Foundations and Trends in Information Retrieval.

- [17] Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). *Linguistic Inquiry and Word Count: LIWC2001*. Mahwah, NJ: Erlbaum Publishers
- [18] Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). *Linguistic Inquiry and Word Count: LIWC2007*. Austin, TX
- [19] Read, J., Carroll, J. (2010). *Annotating Expressions of Appraisal in English*, Διαθέσιμο στην ιστοσελίδα Springerlink.com
- [20] Scherer, Klaus R. 1984. Emotion as a Multicomponent Process: A model and some cross-cultural data. In P. Shaver, ed., *Review of Personality and Social Psych* 5.
- [21] Taboada, M., & Grieve, J. (2006). *Analyzing appraisal automatically*, American Association for Artificial Intelligence
- [22] Valitutti, A., Strapparava, C. & Stock, O. (2004). *Developing affective lexical resources*. *PsychNology Journal* 2: 61-83
- [23] Whilson T., Wiebe J., & Hoffman P. (2005). *Recognising Contextual Polarity in Phrase-Level Sentiment Analysis*. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing
- [24] Whitelaw, C., Garg, N., & Argamon, S. (2005). *Using Appraisal Groups for Sentiment Analysis*. In Proceedings of the conference on information and knowledge management (CIKM).
- [25] Wiebe, J., Bruce, R. and O'Hara, T. (1999). *Development and use of a gold-standard data set for subjectivity classifications*. In *Proceedings of the Association for Computational Linguistics (ACL-1999)*.
- [26] Wilks Y. and Stevenson M., (1998). *The grammar of sense: Using part-of-speech tags as a first step in semantic disambiguation*, *Journal of Natural Language Engineering*, vol. 4, pp. 135–144.

Ελληνόγλωσση:

- [27] Εθνικός Θησαυρός της Ελληνικής Γλώσσας (ΕΘΕΓ™), Ινστιτούτο Επεξεργασίας Λόγου. Διαθέσιμο στην ιστοσελίδα <http://hnc.ilsp.gr/find.asp> Τελευταία πρόσβαση : 20 Σεπτεμβρίου 2013
- [28] Μπαμπινιώτης, Γ., (2006). *Λεξικό της Νέας Ελληνικής Γλώσσας*, Αθήνα : Κέντρο Λεξικολογίας
- [29] Σώμα Νέων Ελληνικών Κειμένων από την “Πύλη για την ελληνική γλώσσα” (<http://www.greek-language.gr>) του Κέντρου Ελληνικής Γλώσσας (ΚΕΓ) . Τελευταία πρόσβαση στις 30 Οκτωβρίου 2013
- [30] Τριανταφυλλίδη, Μ., (1998). *Λεξικό της Κοινής Νεοελληνικής*, Θεσσαλονίκη: Ινστιτούτο Νεοελληνικών Σπουδών Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, διαθέσιμο στην ιστοσελίδα : Πύλη για την Ελληνική Γλώσσα <http://www.greek-language.gr/> Τελευταία πρόσβαση : 15 Σεπτεμβρίου 2013
- [31] Φραγκάκη, Γεωργία Μ., (2010). *Ο αξιολογικός ρόλος του επιθέτου και η χρήση του ως δείκτη ιδεολογίας: μελέτη βασισμένη σε σώματα κειμένων δημοσιογραφικού λόγου*. Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών (ΕΚΠΑ). Σχολή Φιλοσοφική. Τμήμα Φιλολογίας. Τομέας Γλωσσολογίας.