



ΕΘΝΙΚΟ ΚΑΙ
ΚΑΠΟΔΙΣΤΡΙΑΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΦΙΛΟΣΟΦΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΦΙΛΟΛΟΓΙΑΣ



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ
ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ
ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ-ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΤΕΧΝΟΓΛΩΣΣΙΑ VI

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ:

ΑΥΤΟΜΑΤΗ ΑΝΑΛΥΣΗ ΠΡΟΘΕΣΗΣ

του Καλογερόπουλου Αθανάσιου
Φιλολόγου

ΕΠΙΒΛΕΠΩΝ:
ΠΑΠΑΓΕΩΡΓΙΟΥ ΧΑΡΗΣ,
Ερευνητής Β΄
Ινστιτούτο Επεξεργασίας του Λόγου
Τμήμα Φυσικής Γλώσσας και Εξαγωγής Γνώσης

ΑΘΗΝΑ, ΟΚΤΩΒΡΙΟΣ 2013

**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ**

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΙΝΣΤΙΤΟΥΤΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΛΟΓΟΥ

**ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ-ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΤΕΧΝΟΓΛΩΣΣΙΑ VI**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ :

**ΑΥΤΟΜΑΤΗ ΑΝΑΛΥΣΗ ΠΡΟΘΕΣΗΣ ΜΕ ΒΑΣΗ ΤΗ ΘΕΩΡΙΑ
ΤΟΥ SPEECH ACT THEORY
του Καλογερόπουλου Αθανάσιου
Φιλολόγου**

**ΕΠΙΒΛΕΠΩΝ:
ΠΑΠΑΓΕΩΡΓΙΟΥ ΧΑΡΗΣ,
Ερευνητής Β΄
Ινστιτούτο Επεξεργασίας του Λόγου
Τμήμα Φυσικής Γλώσσας και Εξαγωγής Γνώσης**

ΑΘΗΝΑ, ΟΚΤΩΒΡΙΟΣ 2013

ΕΥΧΑΡΙΣΤΙΕΣ

Στο σημείο αυτό θα ήθελα να εκφράσω τις βαθύτατες και ειλικρινές ευχαριστίες μου σε όλους εκείνους που συνέβαλλαν στην περάτωση και στην εκπόνηση της διπλωματικής μου εργασίας. Αρχικά θα ήθελα να εκφράσω την ευγνωμοσύνη μου στον κύριο Χάρη Παπαγεωργίου για την επαγγελματικότητα αλλά και το ήθος του καθ' όλη την διάρκεια της διπλωματικής. Τον ευχαριστώ για την ευκαιρία που μου έδωσε να συνεργαστώ μαζί του, για τις γνώσεις που μου μετέδωσε πάνω στην επεξεργασία φυσικής γλώσσας, την επαγγελματικότητα που τον χαρακτηρίζει και την εξέλιξη μου σαν άνθρωπο πρώτα και σαν νέο επιστήμονα ακολούθως. Η εμπειρία και οι γνώσεις που μας μετέδωσε θα αποτελέσουν πολύτιμα εφόδια, έμπνευση για το μέλλον και ο ίδιος θα είναι το πρότυπο για την μετέπειτα πορεία μου. Τέλος θα ήθελα να ευχαριστήσω την κυρία Ποντίκη Μαρία για την συνεχή και αδιάκοπη προσφορά, την πολύτιμη καθοδήγηση και την στήριξη που μου προσέφερε σε όλη την διάρκεια της εκπόνησης της διπλωματικής μου εργασίας, τον επαγγελματισμό, την οργανωτικότητα και το ήθος που την διακρίνει ως άνθρωπο.

ΠΕΡΙΛΗΨΗ

Στη συγκεκριμένη εργασία γίνεται μία προσπάθεια να δημιουργηθεί μία μηχανή αυτόματης εξόρυξη της πρόθεσης του γράφοντος μέσα από tweets, fora, blogs, κριτικές προϊόντων και γενικά σώματα κειμένων διαφόρων ειδών τα οποία εύκολα μπορούν να αντληθούν από το διαδίκτυο. Σκοπός μας είναι αυτή η μηχανή να έχει και πρακτική εφαρμογή ώστε να καταστεί δυνατόν να χρησιμοποιηθεί για περαιτέρω έρευνα του συγκεκριμένου τομέα της επεξεργασίας φυσικής γλώσσας. Αφετηρία για την εκπόνηση της συγκεκριμένης διπλωματικής θα αποτελέσει η θεωρία των speech acts(προσλεκτικών πράξεων). Για αυτό το λόγο δημιουργήθηκαν τα απαραίτητα λεξικά, οι κανόνες καθώς και τα σώματα κειμένων τα οποία ήταν αναγκαία για την διεξαγωγή των πειραμάτων. Τα αποτελέσματα που επιστράφηκαν ήταν ενθαρρυντικά καταδεικνύοντας όμως την συνέχιση και την εντατικοποίηση της εργασίας πάνω σε αυτό το θέμα με στόχο την βελτιστοποίησή τους.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ	ΣΕΛ.7-9
2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	ΣΕΛ. 10-15
2.1. ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ SPEECH ACT THEORY	ΣΕΛ. 10-15
2.1.1.ΠΑΡΟΥΣΙΑΣΗ ΤΗΣ ΘΕΩΡΙΑΣ ΤΟΥ AUSTIN(1962)	ΣΕΛ. 10-13
2.1.2.SEARLE:ΘΕΩΡΙΑ ΤΩΝ ΠΡΟΣΛΕΚΤΙΚΩΝ ΠΡΑΞΕΩΝ	ΣΕΛ. 13-15
2.2. ΝΕΟΤΕΡΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ	ΣΕΛ. 15-19
2.3. ΠΡΟΣΕΓΓΙΣΗ ΤΟΥ MANSKI	ΣΕΛ. 19
3. ΜΕΘΟΔΟΛΟΓΙΑ-ΥΛΟΠΟΙΗΣΗ	ΣΕΛ. 20-51
3.1. ΠΡΟΣΕΓΓΙΣΗ-ΘΕΜΑΤΟΣ-ΣΤΟΧΟΣ ΕΡΓΑΣΙΑΣ	ΣΕΛ. 20-21
3.2. ΠΕΡΙΓΡΑΦΗ INSIGHTS(ΠΤΥΧΩΝ ΤΗΣ ΠΡΟΘΕΣΗΣ)	ΣΕΛ. 21-24
3.3. ΛΕΞΙΚΟ (LEXICAL RESOURCES)	ΣΕΛ. 24-30
3.4. INTENTION GRAMMAR	ΣΕΛ. 30-31
3.5. INTENTION RULES	ΣΕΛ. 32-49
3.6. ΣΥΝΤΟΜΗ ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ GATE	ΣΕΛ. 50-51
4. ΠΕΡΙΓΡΑΦΗ ΣΩΜΑΤΩΝ ΚΕΙΜΕΝΟΥ	ΣΕΛ. 52-55
4.1. DEVELOPMENT CORPORA	ΣΕΛ. 52-53
4.2. EVALUATION CORPORA	ΣΕΛ. 53-55
5. ΣΥΜΠΕΡΑΣΜΑΤΑ-ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ	ΣΕΛ. 56-60
5.1. ΣΥΜΠΕΡΑΣΜΑΤΑ	ΣΕΛ. 56-57
5.2. ΣΥΓΚΡΙΣΗ ΜΕ ΑΙΑΙΟΟ	ΣΕΛ. 58
5.3. ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ	ΣΕΛ. 59-60
6. ΒΙΒΛΙΟΓΡΑΦΙΑ	ΣΕΛ. 61-63

ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ

1. ΕΙΣΑΓΩΓΗ

Η συγκεκριμένη εργασία αποτελεί μια απόπειρα προσέγγισης του τομέα που ονομάζεται «ανάλυση πρόθεσης» (**intention analysis**). Μέσα σε αυτό το πλαίσιο εντάσσεται και η εργασία την οποία εκπονώ.

Το βασικό θέμα πάνω στο οποίο εστιάζεται η έρευνα είναι αυτό του **stated intention**(εκπεφρασμένης πρόθεσης), το οποίο απασχολεί τον τομέα της ερευνητικής κοινότητας η οποία ασχολείται με σχετικά ζητήματα καθώς έχει πολλές πρακτικές εφαρμογές στην γλωσσική ανάλυση και την μελέτη του λόγου και των τρόπων έκφρασης. Το θέμα το οποίο εξετάζω αντλεί την αξία του και την βαρύτητά του από την σύνδεσή του με την **πρόβλεψη μελλοντικών πράξεων(future action prediction)**. Η πρόβλεψη μελλοντικών πράξεων μέσω της ανάλυσης της πρόθεσης του γράφοντος έχει πλειάδα πρακτικών εφαρμογών σε διάφορους τομείς της καθημερινής ζωής. Αρχικά σε θέματα ασφαλείας(security) για την πρόβλεψη αυτοκτονιών, εγκληματικών-επιθετικών ενεργειών καθώς και στον εντοπισμό ατόμων με κακόβουλη και παραβατική πρόθεση. Επίσης στον τομέα της οικονομίας και του εμπορίου ο εντοπισμός της γνώμης και της πρόθεσης των ανθρώπων για συγκεκριμένες εμπορικές ενέργειες ή οικονομικές συναλλαγές μπορεί να έχει σημαντική συμβολή στη βελτίωση των οικονομικών και εμπορικών σχέσεων. Για παράδειγμα εταιρείες οι οποίες προωθούν τα προϊόντα τους μέσα από την ανίχνευση της πρόθεσης του εμπορικού τους κοινού μπορούν να αντιληφθούν την πορεία των πωλήσεων των προϊόντων τους αλλά και να βρουν τρόπους για να βελτιώσουν τις ήδη προσφερόμενες υπηρεσίες τους ώστε να διατηρήσουν ή και να αυξήσουν το μερίδιό τους στην παγκόσμια αγορά. Η πολιτική είναι ένας ακόμα τομέας όπου η ανάλυση πρόθεσης συναντάει γόνιμο έδαφος για εφαρμογή. Από τα γκάλοπ, την πρόθεση ψήφου και την σφυγμομέτρηση της πρόθεσης του κοινωνικού συνόλου μπορούν να εξαχθούν ασφαλή συμπεράσματα τόσο για την έκβαση εκλογών όσο και την διαμόρφωση του πολιτικού σκηνικού στο σύνολό του λαμβάνοντας υπόψη όλες τις κοινωνικές παραμέτρους.

Οι τρόποι εκμείευσης της πρόθεσης του τώρα με το παρελθόν διαφέρουν σε σημαντικό βαθμό. Τα ερωτηματολόγια στο παρελθόν ήταν ο βασικός τρόπος προκειμένου να καταστεί δυνατό να εντοπιστεί η

πρόθεση των ατόμων. Για παράδειγμα ο Manski(1989) χρησιμοποιεί τα ερωτηματολόγια διότι η θεωρία του στηρίζεται στην πεποίθηση ότι οι απαντήσεις των ατόμων σε ερωτήσεις με προθετικό περιεχόμενο αποτελούν προβλέψεις για την μελλοντική τους συμπεριφορά και με αυτό τον τρόπο μπορεί να μελετηθεί η σχέση μεταξύ της εκπεφρασμένης πρόθεσης (stated intention) και της μελλοντικής συμπεριφοράς των ατόμων(αναλυτικότερα για Manski στο επόμενο κεφάλαιο). Στις μέρες όμως όπου τα μέσα κοινωνικής δικτύωσης, τα blogs όλων των ειδών και γενικά όλοι οι διαδικτυακοί πόροι είναι εύκολα προσβάσιμοι η πρόθεση των ατόμων ανιχνεύεται από ελεύθερα κείμενα(π.χ. tweeter, blogs κ.ά.) και όχι με το να ρωτάμε μέσω ερωτηματολογίων. Παράδειγμα του παραπάνω είναι η εφαρμογή που κατασκευάστηκε από το Aiaioo labs όπου εισάγεις προτάσεις και εμφανίζεται αν υπάρχει πρόθεση με βάση την κατηγοριοποίηση που έχουν υιοθετήσει(περισσότερα στο επόμενο κεφάλαιο).

Η κεντρική ιδέα πάνω στην οποία στηρίχθηκε όλη η εργασία είναι η προθετικότητα(intentionality) και κυρίως η θεωρία των **προσλεκτικών πράξεων (speech act theory)** που διατυπώθηκε αρχικά από τον Austin(1962) και στην συνέχεια από τον μαθητή του τον Searle και στην κατηγοριοποίηση την οποία πρότειναν. Συγκεκριμένα η μελέτη επικεντρώνεται στον τομέα των commissive speech act δηλαδή στις λεκτικές πράξεις που εκφράζουν την υποχρέωση ή την πρόθεση του ομιλητή να κάνει κάτι στο μέλλον. Έπειτα σύμφωνα με την πιθανότητα να κάνει κάτι ο ομιλητής μέσα στην κατηγορία αυτή διακρίνω τρεις υποκατηγορίες: 1) **επιθυμία(desire)**, δηλαδή την επιθυμία του ομιλητή να κάνει κάτι ή την ευχή του για κάτι μελλοντικό(low probability), 2) **σκοπό(purpose)**, δηλαδή τον σκοπό του ομιλητή να κάνει κάτι (medium probability) και 3) **δέσμευση(commitment)**, δηλαδή την δέσμευση του ομιλητή για μία μελλοντική πράξη(high probability). Σκοπός είναι να ανευρεθούν εκείνα τα εκφραστικά μέσα, οι δομές και τα σχήματα που θα επιτρέψουν με την μοντελοποίησή τους να επιτευχθεί η αυτόματη εξαγωγή τους από κείμενα με βάση κανόνες που θα περιγράφουν την συγκεκριμένη μοντελοποίηση. Η επίτευξη αυτού του σκοπού αποτελεί κομβικό σημείο για την συνέχεια και για την περαιτέρω επέκταση του θέματος σε πρακτικές εφαρμογές στα επίπεδα που αναφέρθηκαν προηγουμένως (πολιτικό, οικονομικό, θέματα ασφάλειας) καθώς αποτελεί τον πυρήνα πάνω στον οποίο θα στηριχθεί οποιαδήποτε

εξέλιξη. Ανιχνεύοντας την πρόθεση αρχικά αυτού που γράφει ή μιλάει υπάρχει η δυνατότητα να αναγνωριστεί σε επόμενο μελλοντικό στάδιο και το τι προτίθεται να κάνει, το αντικείμενο της πρόθεσης(object of intention) όπως για παράδειγμα να αγοράσει , να πουλήσει κάτι , να σταματήσει ή να παραιτηθεί από κάτι ή από μία συνήθεια και πολλά πράγματα ακόμη. Όλα αυτά με την προϋπόθεση ότι η εξαγωγή του πυρήνα της πρόθεσης θα έχει επιτευχθεί σε μεγάλο ποσοστό με ακρίβεια με βάση τους κανόνες.

Για όλους τους παραπάνω λόγους γίνεται έκδηλη η σημασία, η σπουδαιότητα καθώς και ο βαθμός δυσκολίας αυτού του εγχειρήματος αλλά και το ερευνητικό ενδιαφέρον και οι προκλήσεις που παρέχονται για την δημιουργία πρακτικών γλωσσικών εφαρμογών που θα συνεισφέρουν στην πρόοδο της επιστημονικής έρευνας πάνω σε αυτό το κομμάτι του επιστητού.

Σκοπός της διπλωματικής είναι:

- **να ανευρεθούν εκείνα τα εκφραστικά μέσα, οι δομές και τα σχήματα που θα επιτρέψουν την επίτευξη της αυτόματης εξαγωγής της πρόθεσης από κείμενα με βάση κανόνες που θα περιγράφουν την συγκεκριμένη μοντελοποίηση.**
- **Η εστίαση και η επικέντρωση στο πρώτο πρόσωπο, ενικό και πληθυντικό, το οποίο θα δηλώνεται και ρητώς στα κείμενα και στις προτάσεις ώστε να είναι έκδηλη η πρόθεση του γράφοντος ή του ομιλητή. Βαρύτητα θα δίνεται στο αν η πρόταση ή το κείμενο εμπεριέχεται μέσα σε εισαγωγικά καθώς και όπως είναι φυσικό στις εκφράσεις που αναφέρονται σε μελλοντικό χρόνο που είναι και το ζητούμενο της εργασίας.**
- **Στόχος είναι η δημιουργία κανόνων και λεξικών βάσει των οποίων θα εντοπίζεται η πρόθεση(intention) αρχικά στα σώματα κειμένων που δημιουργήθηκαν για αυτό τον σκοπό και στη συνέχεια σε οποιοδήποτε σώμα κειμένων.**

ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ

ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

Παλαιότερα η ενασχόληση με την εξόρυξη της πρόθεσης(intention) δεν είχε συγκεντρώσει το βλέμμα της επιστημονικής κοινότητας που ασχολούνταν με την επεξεργασία φυσικής γλώσσας λόγω του ότι η περισσότερη ερευνητική εργασία επικεντρωνόταν στην εξόρυξη της αντικειμενικής πληροφορίας από διάφορες κατηγορίες κειμένων και λιγότερο της εξόρυξης της υποκειμενικής πληροφορίας όπως είναι για παράδειγμα η πρόθεση του γράφοντος να κάνει κάτι ή να συμπεριφερθεί με έναν συγκεκριμένο τρόπο. Όσοι ασχολούνταν με τα οικονομικά πίστευαν ότι πιο πολύ ενδιαφέρουν οι πράξεις των ανθρώπων παρά η πιθανότητα ή η πρόθεση που έχουν οι άνθρωποι να συμπεριφερθούν ή να πράξουν κάτι με ένα συγκεκριμένο τρόπο στο μέλλον. Σήμερα λόγω της δραματικής αλλαγής του Διαδικτύου, της επιρροής στη ζωή του ανθρώπου με την τεράστια ανάπτυξη των μέσων κοινωνικής δικτύωσης το θέμα της εξόρυξης την πρόθεσης(intention) αποτελεί δελεαστική πρόκληση για κάθε ερευνητή που επιθυμεί να ασχοληθεί συστηματικά με τον συγκεκριμένο τομέα της επεξεργασίας φυσικής γλώσσας.

2.1 ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ SPEECH ACT THEORY

2.1.1 Παρουσίαση της θεωρίας του Austin(1962)

Ο Austin(1962) ήταν ο πρώτος που ασχολήθηκε με τη «θεωρία των έλλογων πράξεων» (Austin(1962)). Εισήγαγε τη θεωρία του σχετικά με τις έλλογες πράξεις με έναν πιο αφηρημένο και φιλοσοφικό τρόπο. Στο βιβλίο του με τίτλο «Πώς να καταφέρετε πράγματα με λέξεις», ο Austin(1962) παρουσίασε ένα νέο σχήμα ανάλυσης της σημασίας. Η έννοια της σημασίας περιγράφεται στο πλαίσιο των γλωσσολογικών συμβάσεων σχετικών με λέξεις ή προτάσεις, μία κατάσταση όπου ο ομιλητής εκφέρει κάτι προς τον ακροατή και ο ομιλητής έχει την πρόθεση να εκπέμψει ένα μήνυμα ή πληροφορία προς τον ακροατή. Η έννοια ότι η σημασία

συσχετίζεται με αυτές τις σχέσεις και συμβάσεις απεικονίζεται ξεκάθαρα από την έννοια των *πράξεων*. Η ανάλυση του Austin(1962) διαφέρει από άλλες γιατί η σημασία δεν επεξηγείται μέσω φόρμας μειώσεων όπως κάνει, για παράδειγμα, ο Russel, ο οποίος χρησιμοποιούσε μαθηματικά και λογικά μοντέλα και υποστήριζε ότι η σημασία μίας πρότασης είναι ένα γεγονός στο οποίο αντιστοιχίζεται η πρόταση (μείωση της σημασίας).

Με άλλα λόγια, ο Austin(1962) ήθελε να περιγράψει «τη συνολική έλλογη πράξη σε μία συλλογική έλλογη κατάσταση». Στο πλαίσιο των λεκτικών πράξεων και των συνθηκών ευτυχίας, ο Austin(1962) υποστήριζε ότι για να εκφραστεί μία επιτελεστική πρόταση πρέπει να λάβουν υπόψη του τους όρους του συμβατού, τη πραγματικότητα και της προθετικότητας. Για να το εξηγήσει αυτό, η εκφορά της πρότασης που σχετίζεται με έγκυρες συμβάσεις, η πραγματικά ακριβής εκφορά του ομιλητή προς τον ακροατή με την αντίστοιχη αντίδραση από τον ακροατή, καθώς και η συσχετιζόμενη πρόθεση του ομιλητή πρέπει να ληφθεί υπόψη. Ο Austin(1962) δήλωσε ότι με την εκφορά εκ μέρους του ομιλητή μίας επιτελεστικής πρότασης καταδεικνύει μία κατάσταση ομιλίας όπου υπάρχουν ορισμένες συμβάσεις, υπάρχουν συγκεκριμένοι συμμετέχοντες και συνθήκες, ο ομιλητής εκτελεί μία πράξη με ένα συγκεκριμένο τρόπο, ο ομιλητής επίσης ανταποκρίνεται με ορισμένο τρόπο επίσης, ο ομιλητής έχει σκέψεις, αισθήματα, προθέσεις και υποτίθεται ότι θα ενεργήσει με ορισμένο τρόπο στο κοντινό μέλλον.

Μετά ο Austin(1962) όρισε την έννοια της επιτελεστικότητας ως εννοιακότητα στην «εκτέλεση μίας πράξης». Κατηγοριοποίησε την λεκτικές πράξεις σε τρεις ξεχωριστές κατηγορίες: **Απολεκτικές**(*illocutionary*), **ελληλεκτικές**(*perlocutionary*) και **εκφωνητικές** πράξεις(*locutionary*). Ο Austin(1962) διακρίνει προσεκτικά την έννοια για ελληλεκτικές(*perlocutionary*) πράξεις από απολεκτικές(*illocutionary*) και εκφωνητικές(*locutionary*). Με τον όρο εκφωνητική πράξη αναφέρεται σε μια πράξη οικοδόμησης φράσης ακολουθώντας γραμματικές, συντακτικούς κανόνες και τρόπους άρθρωσης μιας φράσης. Οι εκφωνητικές πράξεις περιλαμβάνουν **φωνητικές πράξεις**- *phonetic acts* (πράξεις για προφορά ήχων), **απολεκτικές**

πράξεις - *phatic acts* (πράξεις έκφρασης λέξεων και προτάσεων , λαμβάνοντας υπόψη τους φωνολογικούς και συντακτικούς κανόνες της γλώσσας στην οποία ανήκουν) και **ρητορικές πράξεις**- *rhetic acts* (πράξεις έκφρασης μιας φράσης , όπου η αίσθηση είναι ότι αναφέρεται σε ένα περισσότερο ή λιγότερο καθορισμένο τρόπο) . Οι ελλεκτικές πράξεις αναφέρονται στην επίδραση που έχει το να προφέρεις μια φράση. Τέλος, επικεντρώθηκε περισσότερο στην ανάλυση των απολεκτικών πράξεων . Η διαφορά μεταξύ της ελλεκτικής και απολεκτικής πράξης είναι ότι η πρώτη είναι «συμβατική» και η δεύτερη είναι « μη συμβατική», σύμφωνα με την προσέγγιση του Austin(1962) . Με άλλα λόγια , αυτό σημαίνει ότι ένας ομιλητής που εκτελεί ελλεκτική πράξη στηρίζεται στις κοινωνικά αποδεκτές συμβάσεις και ως εκ τούτου ο ομιλητής αποκτά στην έκφρασή του μια ξεχωριστή δύναμη. . Από την άλλη πλευρά, οι απολεκτικές πράξεις είναι οι επιπτώσεις των ελλεκτική πράξεων . Σύμφωνα με τις διακρίσεις αυτές , οι ελλεκτικές , σε αντίθεση με τις απολεκτικές πράξεις , έχουν μια ιδιαίτερη δύναμη , καθώς επίσης οι ελλεκτικές πράξεις διατηρούν την ισχύ και την πληρότητα τους, χωρίς να θεωρούνται ως το αποτέλεσμά τους

. Ο Austin(1962) πρότεινε μια κατηγοριοποίηση των απολεκτικών πράξεων(*illocutionary acts*) σε πέντε τύπους . Χώρισε τις απολεκτικές πράξεις σε: **verdictives** (πράξεις που εκφράζουν απόφαση) , **exercitives** (προτρεπτικές) , **commissives** (επιτελεστικές) , **behatives** (συμπεριφορικές) και **expositives** (επικοινωνιακές) . Ο Austin(1962) ήταν από τους πρώτους που εισήγαγε τη θεωρία των προσλεκτικών πράξεων(*speech act theory*) . Παρουσίασε τις *felicity conditions*(προϋποθέσεις επιτυχίας της επικοινωνιακής πράξης) , τη διάκριση των απολεκτικών, ελλεκτικών και εκφωνητικών πράξεων, και την κατηγοριοποίηση των ελλεκτικών πράξεων σε επιμέρους κατηγορίες.

2.1.2 Searle : ΘΕΩΡΙΑ ΠΡΟΣΛΕΚΤΙΚΩΝ ΠΡΑΞΕΩΝ

Με σπουδές δίπλα στο δάσκαλο του τον Austin(1962) , ο Searle , ένας από τους ιδρυτές αυτής της θεωρίας , είχε επηρεαστεί πολύ από αυτόν. Προερχόμενος από μια φιλοσοφική περιοχή , προσπάθησε να διερευνήσει και να συλλάβει τη σχέση μεταξύ της κοινωνικής επικοινωνίας και της γλωσσολογίας . Η έκφραση , ως ένα φυσικό γεγονός, αποκτά την έννοια μόνο εάν βρίσκεται σε ένα πραγματικό πλαίσιο και συνδέεται με ρεαλιστικά φαινόμενα. " Σημασία κατά τη χρήση », όπως ο Leech δηλώνει στις Αρχές Πραγματολογίας (1983).

Ο Searle προτείνει μία ταξινόμησή των ελλεκτικών πράξεων(illocutionary acts) προχωρώντας στην επίλυση του μεθοδολογικού αυτού προβλήματος . Η εύρεση δώδεκα ειδών κριτηρίων εξηγεί το θεωρητικό πλαίσιο πάνω στο οποίο δημιουργεί την ταξινόμηση του. Η ταξινόμησή του βασίζεται στις συνθήκες επιτυχίας(felicity conditions) ως συστατικό κανόνων: το **προτασιακό περιεχόμενο**(*propositional content*,) , τις **προπαρασκευαστικές συνθήκες**(*preparatory conditions*) , τους **όρους ειλικρίνειας**(*sincerity conditions*) και τις **βασικές προϋποθέσεις**(*essential conditions*) . Διέκρινε την έννοια του ελλεκτικού σημείου και της ελλεκτικής δύναμης. Το ίδιο ελλεκτικό σημείο (δηλαδή το γεγονός που ωθεί τον αποδέκτη να κάνει μια πράξη) έχει πολλές κλίμακες ισχύος . Το τρίτο κεφάλαιο, που παρέχει ο Searle , είναι η εκπεφρασμένη ψυχολογική κατάσταση . Εξέφρασε ότι η ψυχολογική κατάσταση μπορεί να διαφέρει από τις προθέσεις ή τους σκοπούς ή την αρχική αλήθεια του ομιλητή . Κάποια άλλα κριτήρια είναι η θέση και η κατάσταση του ομιλητή . Είναι ό, τι , οι Brown and Levinson (1978) καλούν κοινωνικές παραμέτρους , στρατηγικές ευγένειας και συγκεκριμένα την κοινωνική απόσταση μεταξύ του ομιλητή και του ακροατή. Σύμφωνα με τον Searle , μια σημαντική πτυχή είναι το πώς το εκφώνημα του ομιλητή σχετίζεται με τα ενδιαφέροντα του ακροατή . Επίσης, ο ίδιος διαχωρίζει τα ρήματα σε **επιτελεστικά**(*performatives*) και σε **πιθανόν επιτελεστικά**(*potential performatives*). Η

εναλλακτική ταξινόμησή του, όπως ο ίδιος την ονόμασε (Searle : 1976) , βασίζεται σε αυτά τα κριτήρια .

Ο ίδιος διακρίνει τις προσλεκτικές πράξεις σε πέντε κατηγορίες . Σε αυτό το σημείο θα πρέπει να καθοριστεί η διαφορά μεταξύ των πράξεων ομιλίας και των ρημάτων των λεκτική πράξεων . Στην **προτρεπτική κατηγορία**(directive class) , ο ορισμός είναι ότι αποτελούν προσπάθειες – διαφορετικής βαρύτητας - και τα κοινά τους είναι ότι υπάρχει ένας ομιλητής που προσπαθεί να κάνει τον ακροατή να κάνει κάτι , να κάνει μια ενέργεια , μια πράξη . Στην **αντιπροσωπευτική κατηγορία**(representative class) , σκοπός της είναι να δεσμεύσει τον ομιλητή για την αλήθεια της εκπεφρασμένης πρόθεσης. Η κατηγορία αυτή βασίζεται στην διάσταση αληθείας και ψεύδους. **Δεσμευτικές**(commisives) είναι οι πράξεις των οποίων σκοπός είναι να δεσμεύσουν τον ομιλητή σε κάποια μελλοντική δράση . Οι **εκφραστικές**(expressives) αντικατοπτρίζουν την ψυχολογική κατάσταση του ομιλητή . Σε αυτή την κατηγορία δεν υπάρχει ένας συγκεκριμένος προσανατολισμός. Οι **δηλωτικές** (declarations) χαρακτηρίζονται από πλήρη αντιστοιχία μεταξύ του προθεσιακού περιεχομένου και της πραγματικότητας στην περίπτωση που υπάρχει επιτυχημένη έκφραση. Ο Searle δείχνει να αποδίδει , επίσης, το ρόλο της συντακτικής δομής του ρήματος και το πώς αυτό επηρεάζει την κατηγοριοποίησή του . Για ορισμένα ρήματα , η δομή είναι ένα τεκμήριο και ένδειξη για την κατηγορία στην οποία ανήκει.

Εν κατακλείδι , ο Searle σε " μια ταξινόμηση των ελλεκτική πράξεων" επικρίνει " την «αδυναμία " του Austin(1962) και παρέχει κάποια θεωρητικά ζητήματα που πρέπει να εξετάσουμε . Πρώτα απ 'όλα η κατάταξη, πρέπει να βασίζεται στην ακρίβεια. Επίσης, μια σημαντική γνώση της έρευνάς του είναι ότι η δομή ορισμένων ρημάτων είναι η απόδειξη για την κατηγορία τους. Μέχρι σήμερα, όλη η επιστημονική κοινότητα σέβεται και ακολουθεί την κατηγοριοποίηση του , παρόλα τα προβλήματα από το φιλοσοφικό παρελθόν του .

2.2 Νεότερες προσεγγίσεις για το Intention Analysis

Πολλοί επιστήμονες κάνουν αξιοσημείωτες προσπάθειες σε αυτόν τον τομέα της ανάλυσης της θεωρίας της γλωσσικής πράξης και της εφαρμογής της σε γραπτά είδη κειμένων (επεξηγηματικές ή διαδικτυακές πηγές). Αρκετές προσπάθειες επικεντρώνονται κυρίως στην ταξινόμηση του ηλεκτρονικού ταχυδρομείου (email).

Οι *Qadir and Riloff* (2011) προσπάθησαν να διακρίνουν μεταξύ των επεξηγηματικών προτάσεων και αυτών των γλωσσικών πράξεων σε αναρτήσεις σε πίνακες ανακοινώσεων και να κατατάξουν τις προτάσεις των γλωσσικών πράξεων σε τέσσερα είδη: επιτελεστικές (commissives), προτρεπτικές (directives), εκφραστικές (expressives) και αντιπροσωπευτικές (representatives) ακολουθώντας την κατηγοριοποίηση του Searle (1976). Καθόρισαν ξεκάθαρα κάθε κατηγορία και αργότερα σχεδίασαν μια ποικιλία από λεξικά, συντακτικά και σημασιολογικά χαρακτηριστικά κι επίσης ενσωμάτωσαν λίστες από λέξεις γλωσσικών πράξεων που συγκέντρωσαν από εξωτερικές πηγές. Τα χαρακτηριστικά αυτά χωρίζονται σε τρεις ομάδες: *Λεξικά και Συντακτικά Χαρακτηριστικά*, *Στοιχεία Γλωσσικών Πράξεων* και *Σημασιολογικά Χαρακτηριστικά*. Η πρώτη ομάδα περιλαμβάνει προσωπικές αντωνυμίες, χρόνους, χρόνους και πρόσωπα, επιρρήματα, ρηματικές φράσεις, σχεδιαστικές φράσεις, προτάσεις με πρώιμη στίξη, προτάσεις οι οποίες ξεκινούν με επίρρημα/ρήμα, γειτονική ερώτηση, θέση της πρότασης και αριθμός των ρημάτων. Η δεύτερη ομάδα περιλαμβάνει λίστα από λέξεις γλωσσικών πράξεων με λέξεις κλειδιά από τους Searle και Wierzbicka (ρήματα). Η τελευταία ομάδα περιλαμβάνει ένα σημασιολογικό λεξικό για κτηνιατρική ιατρική που χρησιμοποιεί τον αλγόριθμο του Basilisk το οποίο βασίστηκε σε ένα σώμα από αναρτήσεις σε πίνακα ανακοινώσεων κτηνιατρικής ώστε μετά με όλα τα διαγνωστικά τεστ που έκαναν να κατηγοριοποιήσουν τις προτάσεις ως γλωσσικές πράξεις σε αναρτήσεις πινάκων ανακοινώσεων.

Επίσης ο *Cohen* (2004) προσπάθησε να εισάγει την έννοια των 'γλωσσικών πράξεων ηλεκτρονικού ταχυδρομείου' ('email speech acts'). Σε αυτήν την έννοια η ταξινόμηση καθορίζεται από την άποψη ενός συγκεκριμένου ζεύγους ρήματος-ουσιαστικού σύμφωνα με μια προσχεδιασμένη οντολογία. Σε αυτή την οντολογία υπάρχουν δυο βασικά μέρη. Στο πρώτο θεωρείται σημαντικό η οντολογία να αντικατοπτρίζει παρατηρούμενη γλωσσολογική συμπεριφορά ώστε να αντανakλάται οποιαδήποτε θεωρητική, φιλοσοφική ή αφηρημένη άποψη του χώρου των πιθανών γλωσσικών πράξεων και στο δεύτερο η

ταξινόμηση πρέπει να αντανακλά κοινές μη γλωσσολογικές χρήσεις του ηλεκτρονικού ταχυδρομείου (αίτηση, πρόταση, τροποποίηση, διάπραξη, παράδοση, άρνηση, υπενθύμιση).

Οι *Goldstein* και *Sabin* (2006) υιοθέτησαν αυτήν την έννοια των πράξεων ηλεκτρονικού ταχυδρομείου αλλά έκαναν ένα βήμα μπροστά χρησιμοποιώντας λεξικά ρημάτων για να τις ταξινομήσουν.

Οι *Cohen* και *Carvalho* (2005) πρότειναν ένα νέο τρόπο ταξινόμησης των γλωσσικών πράξεων ηλεκτρονικού ταχυδρομείου εκμεταλλευόμενοι μια διαδοχική συσχέτιση μεταξύ μηνυμάτων ηλεκτρονικού ταχυδρομείου στο ίδιο νήμα ή σε γενικότερο πλαίσιο. Εφηύραν ένα αλγόριθμο ταξινόμησης κειμένων βασισμένο σε μια συλλογική μέθοδο ταξινόμησης.

Οι *Cohen* και *Carvalho* (2006) χρησιμοποίησαν χαρακτηριστικά αλληλουχίας n-gram ώστε να καθορίσουν και να αποφασίσουν ποια χαρακτηριστικά των n-gram έχουν μεγαλύτερη συσχέτιση με διαφορετικές γλωσσικές πράξεις ηλεκτρονικού ταχυδρομείου με βάση τις προθέσεις του συγγραφέα. Ο τελικός στόχος ήταν να βελτιώσουν και να βελτιστοποιήσουν την προηγούμενη ταξινόμηση ηλεκτρονικού ταχυδρομείου.

Οι *Twitchell and Jr.* (2004) και *Twitchell et al.* (2004) πρότειναν μια μέθοδο ονομαζόμενη ‘κατασκευή προφίλ γλωσσικών πράξεων’ (*speech act profiling*). Απεικόνισαν δυνητικές κατηγορίες διαλόγου σε ένα γράφημα ραντάρ ώστε να κατηγοριοποιήσουν συζητήσεις σε άμεσες ανταλλαγές μηνυμάτων και chat rooms. Η κατασκευή προφίλ γλωσσικών πράξεων αποτελεί μια μέθοδο η οποία αναλύει και απεικονίζει συζητήσεις και τους ομιλητές τους και είναι ένας συνδυασμός από ασαφή πληκτρολόγηση και μοντελοποίηση γλωσσικών πράξεων ή κατηγοριοποίηση. Ο επιστημονικός τομέας της ανάκτησης πληροφοριών μπορεί εύκολα να εκμεταλλευτεί την αυτόματη ταξινόμηση των γλωσσικών πράξεων. Με αυτόν τον τρόπο, οι συζητήσεις μπορούν να αποδίδονται με αφοσιωμένα προφίλ που υποδεικνύουν την πρόθεση της όλης συζήτησης. Εκτός από αυτό, οι συζητήσεις μπορούν εύκολα να προσδιοριστούν και να ανακτηθούν σύμφωνα με την απατηλή πρόθεση (η εξαπάτηση είναι μια απολεκτική πράξη της οποίας ο στόχος είναι να δημιουργήσει μια ψευδή πρόθεση στον ακροατή). Μια απλή ταξινόμηση είναι μεταξύ αυτών που μοιάζουν με συνεντεύξεις και αυτών που δε μοιάζουν. Το σύνολο ετικετών SWBD-DAMSL για πράξεις διαλόγων που χρησιμοποιείται από εκείνους βασίζεται ξεκάθαρα στην ταξινόμηση του Searle για τη θεωρία των γλωσσικών

πράξεων. Μια τελευταία παρατήρηση είναι ότι η κατασκευή προφίλ των γλωσσικών πράξεων χωρίζεται σε δυο κύριες κατηγορίες. Η πρώτη είναι ο προσδιορισμός δυνητικών πιθανοτήτων των γλωσσικών πράξεων χρησιμοποιώντας γνώση και ανατροφοδότηση από τη γλώσσα και μετρώντας γλωσσικές πράξεις σε υποσημειωμένα σώματα κειμένων. Η δεύτερη είναι η δημιουργία προφίλ γλωσσικών πράξεων για μια ομάδα εκφράσεων που περιλαμβάνει κάθε δυνητικές γλωσσικές πράξεις ευρισκόμενες σε αυτές τις εκφράσεις με τις πιθανότητές τους.

Το ενδιαφέρον στην επεξεργασία της φυσικής γλώσσας έγινε εργαλείο για εταιρείες να αποσπάσουν όχι μόνο πληροφορίες και καταστάσεις για γεγονότα αλλά κυρίως να στηρίζουν προθέσεις σε αυτά τα γεγονότα και αυτά τα αντικείμενα ώστε να εκμεταλλευτούν τα αποτελέσματά τους. Χρησιμοποιώντας τη θεωρία του Austin (1962) τα εργαστήρια Aiaίοο πάνε ένα βήμα παρακάτω και δημιουργούν ένα εργαλείο το οποίο βρίσκει διάφορες περιπτώσεις προθέσεων (όπως αγορά, ερώτηση, πρόταση, παραίτηση κ.ά.) και τους σημασιολογικούς τους ρόλους: αυτός που ενεργεί και εκφράζει την πρόθεση, αυτός που δέχεται την ενέργεια και ίσως συμμετέχει και το αντικείμενο στην πρόθεση. Αποσπούν αναλύσεις προθέσεων όπως εξηγείται πιο πάνω. Επίσης, χρησιμοποιούν ανάλυση συναισθημάτων για να εξάγουν πληροφορίες για το θετικό ή αρνητικό προσανατολισμό των προτάσεων και ανάλυση γεγονότων που είναι χρήσιμη για την ανάλυση της γνώσης. Το πρώτο βήμα είναι η διαφοροποίηση όλων των μορφών κειμένων σε μια δομική μορφή. Όταν το κείμενο είναι πληροφοριακό χρησιμοποιούν ανάλυση γεγονότων και όταν το κείμενο είναι μια γλωσσική πράξη χρησιμοποιούν ανάλυση προθέσεων. Μια εμπειρία από μια άλλη εταιρεία είναι η αναγνώριση των προθέσεων σε εισερχόμενα μηνύματα πελατών ή σε αντίγραφα ομιλιών από τηλεφωνικά κέντρα.

Ο *Jawad* στο άρθρο του περί προσλεκτικής πράξης στα Αραβικά σημειώνει ότι οι πέντε ταξινομήσεις του Searle δεν καλύπτουν όλες τις γλωσσικές πράξεις. Συμφωνεί με τον Dijk (1977b) ότι υπάρχουν 'βοηθητικές' γλωσσικές πράξεις. Εξετάζει τις άμεσες και τις έμμεσες γλωσσικές πράξεις και χρησιμοποιεί τις συνθήκες επιτυχίας του Searle. Όντας ένα κείμενο με θρησκευτικό προσανατολισμό και θρησκευτικούς υπαινιγμούς ο αριθμός των διευθυντικών γλωσσικών πράξεων είναι υψηλός. Είναι σημαντικό να αναφέρουμε τις πτυχές αυτών των πράξεων. Διαταγές και απαγορεύσεις, συμβουλές, προειδοποιήσεις και προσευχές είναι οι πέντε αυτές πτυχές. Σημειώνει τη σχέση μεταξύ της

δομής μιας δεδομένης γλωσσικής πράξης και της λειτουργίας της. Σύμφωνα με την ανάλυσή του, υπάρχει ένα λογικό επιχείρημα ότι ένα θρησκευτικό κείμενο έχει αυτά τα αποτελέσματα.

Όπως επίσημα υποστηρίζουν ‘η ευχή είναι μια επιθυμία ή ελπίδα να γίνει κάτι’. Επομένως αναφέρονται στην πρόθεση των ανθρώπων να κάνουν κάτι στο μέλλον. Πήραν πρωτοβουλία από τις ευχές που άνθρωποι ανά τον κόσμο προσέφεραν το Δεκέμβρη του 2007. Οι ευχές αυτές είχαν τυπωθεί σε κονφέτι κι έπεσαν από τον ουρανό κατά τη διάρκεια της διάσημης γιορτής της Παραμονής της Πρωτοχρονιάς στην Times Square της Νέας Υόρκης. Προσπάθησαν να αναλύσουν αυτές τις ευχές και να κτίσουν γενικούς ανιχνευτές ευχών για κείμενα φυσικής γλώσσας. Η ανίχνευση ευχής συμπληρώνει τις παραδοσιακές μελέτες για την ανάλυση συναισθημάτων και είναι πολύτιμη για την παγκόσμια συλλογή προθέσεων, επιθυμιών, ευχών κι ελπίδων. Τα στοιχεία που χρησιμοποίησαν προέρχονταν από κριτικές προϊόντων και διαδικτυακές πολιτικές συζητήσεις. Οι ευχές διαφέρουν από τις αναλύσεις συναισθημάτων και τις εξορύξεις απόψεων διότι αποκαλύπτουν ρητά τί θέλουν οι άνθρωποι να συμβεί στο μέλλον. Επίσης, οι ευχές αποκαλύπτουν τα αισθήματα των ανθρώπων, τις αντιλήψεις τους και τις επιθυμίες τους για το μέλλον. Στο επόμενο επίπεδο διαχώρισαν τις ευχές ανάλογα με το θέμα τους και την έκταση και σύγκριναν τις ευχές των Αμερικάνων με τις ευχές άλλων ανθρώπων σε διαφορετικές τοποθεσίες με τους ίδιους όρους της ανάλυσης. Δημιούργησαν δύο απλούς ανιχνευτές ευχών: 1) **χειρωνακτικούς** (χρησιμοποιώντας δύο φυσικούς ομιλητές της Αγγλικής για να βρουν πρότυπα που ίσως εκφράζουν ευχές) και 2) **Λεκτικό** (εξασκώντας ένα σταθερό ταξινομητή κειμένων βασισμένο σε λέξεις). Επίσης εφηύραν ένα τρόπο να ανακαλύπτουν αυτόματα πρότυπα επιθυμιών από σώματα ευχών εκμεταλλευόμενοι τους πλεονασμούς στο πώς η ίδια ευχή εκφράζεται. Τελικώς χρησιμοποίησαν αυτά τα πρότυπα ευχών ως χαρακτηριστικά για τη μαθησιακή διαδικασία μέσω αυτών των προτύπων. Δύο τύποι αντιστοιχιών είναι πιθανοί: η αυστηρή αντιστοιχία, η οποία απαιτεί το πρότυπο να αντιστοιχεί με μια ολόκληρη πρόταση από την αρχή μέχρι το τέλος και η μη αυστηρή αντιστοιχία, η οποία απαιτεί το πρότυπο να αντιστοιχεί μόνο κάπου μέσα σε μια πρόταση.

2.3 Η προσέγγιση του Manski

ο Charles Manski (1989) έχει κάνει μια πρωταρχική και γενική έρευνα πάνω στην πρόθεση και πρόβλεψη της ανθρώπινης συμπεριφοράς. Αυτή η εργασία επικεντρώνεται στη σχέση μεταξύ της δεδηλωμένης πρόθεσης και της μεταγενέστερης συμπεριφοράς κάτω από την υπόθεση ότι οι άνθρωποι έχουν συγκεκριμένες λογικές προσδοκίες και ότι οι απαντήσεις τους στις προσδοκίες αυτές είναι οι καλύτεροι δείκτες μελλοντικής συμπεριφοράς. Τα δεδομένα για τις προθέσεις φανερώνουν την πιθανότητα ενός ανθρώπου να συμπεριφερθεί με έναν συγκεκριμένο τρόπο στο μέλλον. Το θεωρητικό υπόβαθρο βασίζεται στην κοινωνική ψυχολογία και τις θεωρίες οικονομίας. Οι κοινωνικοί ψυχολόγοι δίνουν μεγάλη σημασία στις προθέσεις. Υποθέτουν ότι η πρόθεση συνιστά μια ψυχική κατάσταση η οποία προηγείται της συμπεριφοράς. Η συμπεριφορική πρόθεση ενός ανθρώπου είναι μια υποκειμενική πιθανότητα ότι η συμπεριφορά ή το ενδιαφέρον θα πραγματοποιηθούν πράγματι στο μέλλον. Οι οικονομολόγοι, από την άλλη πλευρά, αγνοούν τα δεδομένα των προθέσεων παρατηρώντας την απόκλιση μεταξύ των προθέσεων και της μεταγενέστερης συμπεριφοράς. Δείχνουν να πιστεύουν μόνο τις πράξεις και όχι τα λόγια των ανθρώπων. Οπότε οι οικονομολόγοι δε συνήθιζαν να παίρνουν υπόψη τα δεδομένα των προθέσεων για να προβλέψουν συμπεριφορές. Έτσι άφησαν την έρευνα της αγοράς προθέσεων και πιθανοτήτων των καταναλωτών στους αναλυτές αγορών διότι δεν έβρισκαν κανένα ενδιαφέρον στην ανάλυση των δεδομένων των προθέσεων.

Έχοντας λοιπόν ως βάση όλες αυτές τις κατηγοριοποιήσεις που προτείνονται από Austin, Searle καθώς και νεότερους ερευνητές σκοπός της εργασίας είναι η επικέντρωση στα commissive speech acts και η σημασιολογική διαφοροποίηση των ειδών της πρόθεσης ανάλογα με το εννοιολογικό βάρος που φέρουν.

ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ

ΜΕΘΟΔΟΛΟΓΙΑ-ΥΛΟΠΟΙΗΣΗ

3.1 Προσέγγιση θέματος-Στόχος εργασίας

Όπως διατυπώθηκε και στο κείμενο της εισαγωγής η εργασία έχει ως θέμα την **ανάλυση πρόθεσης(intention analysis)** του ομιλητή ή του γράφοντος και συγκεκριμένα την **εκπεφρασμένη πρόθεση(stated intention)**. Το ερευνητικό ενδιαφέρον εστιάζεται στην πρόθεση για μελλοντικές ενέργειες. Βάση της μεθοδολογίας που χρησιμοποιείται είναι το **speech act theory(θεωρία προσλεκτικών πράξεων)** όπως διατυπώθηκε από τους Austin(1962) αρχικά και Searle στη συνέχεια υιοθετώντας εν μέρει την κατηγοριοποίηση που πρότειναν.

Σημείο ενδιαφέροντος είναι τα **commissive speech acts** δηλαδή όλες εκείνες οι δομές οι οποίες εκφράζουν την υποχρέωση ή την πρόθεση του ομιλητή ή του γράφοντος να κάνει κάτι στο μέλλον. Διακρίνονται τρεις βασικές πτυχές (insights) με βάση την πιθανότητα να συμβεί αυτό το οποίο εκφράζει ο ομιλητής ή ο γράφων και θα αναλυθούν διεξοδικά στην συνέχεια της παρουσίασης της μεθοδολογίας:

- **επιθυμία(desire)**, δηλαδή την επιθυμία του ομιλητή να κάνει κάτι ή την ευχή του για κάτι μελλοντικό(low probability),
- **σκοπό(purpose)**, δηλαδή τον σκοπό του ομιλητή να κάνει κάτι (medium probability) και
- **δέσμευση(commitment)**, δηλαδή την δέσμευση του ομιλητή για μία μελλοντική πράξη(high probability).

Πηγή άντλησης της πρόθεσης θα αποτελέσουν ελεύθερα κείμενα ή προτάσεις τα οποία αντλούνται από το διαδίκτυο και συγκεκριμένα από blogs, webpages και μέσα κοινωνικής δικτύωσης όπως είναι το twitter. Περιλαμβάνονται προτάσεις οι οποίες εκφράζουν άμεση πρόθεση(explicit

intention) π.χ. I want to buy an I phone, αλλά και σε έμμεσες μορφές έκφρασης της πρόθεσης(implicit intention) π.χ. I want an Iphone.

3.2 Περιγραφή insights(πτυχών της πρόθεσης)

Για τον διαχωρισμό των πτυχών του intention αντλήθηκαν πληροφορίες από το **framenet**¹. Το framenet είναι μια πλατφόρμα, ένα εγχείρημα που χτίζει μία λεξιλογική βάση της Αγγλικής γλώσσας που είναι αναγνώσιμο και από ανθρώπους αλλά και μηχανές και βασίζεται σε επισημειωμένα παραδείγματα σχετικά με την χρήση των λέξεων σε πραγματικά κείμενα. Αποτελεί ένα λεξικό το οποίο περιλαμβάνει πάνω από 10,000 λεξικές σημασίες, οι περισσότερες από τις οποίες αποτελούν επισημειωμένα παραδείγματα τα οποία αποτυπώνουν την σημασία και την χρήση των λέξεων. Στην επεξεργασία φυσικού λόγου αποτελεί ένα μοναδικό σύνολο δεδομένων εκμάθησης για σημασιολογική κατηγοριοποίηση με εφαρμογή σε πολλούς τομείς όπως η εξόρυξη πληροφορίας, η μηχανική μετάφραση, η αναγνώριση γεγονότων, η ανάλυση συναισθήματος και άλλα. Η θεωρία πάνω στην οποία βασίζεται είναι ότι οι σημασίες των περισσότερων λέξεων μπορούν να κατανοηθούν στο πλαίσιο σημασιολογικών πεδίων(semantic frames) τα οποία περιλαμβάνουν την περιγραφή ενός είδους γεγονότος, τις σχέσεις που ενυπάρχουν και τους συμμετέχοντες σε αυτό. Έτσι οι λέξεις ανάλογα με αυτό που δηλώνουν εντάσσονται στην ανάλογη σημασιολογική κατηγορία.

Επίσης πληροφορία αντλήθηκε από το **wordnet**² το οποίο αποτελεί μία τεράστια λεξιλογική βάση της Αγγλικής γλώσσας που περιλαμβάνει ουσιαστικά, ρήματα, επίθετα, επιρρήματα οργανωμένα σε σύνολα συνωνύμων κάθε ένα από τα οποία εκφράζουν μία διακριτή έννοια.

Μία ακόμα πηγή από την οποία εκμαιεύθηκαν χρήσιμες πληροφορίες ήταν το **verbnet**³ το οποίο αποτελεί μία τεράστια διαδικτυακή βάση ρημάτων για τα Αγγλικά. Είναι ένα ιεραρχικό, ανεξάρτητο συγκεκριμένου πεδίου, λεξικό ρημάτων με παραπομπές και συνδέσεις σε άλλες βάσεις δεδομένων όπως το *wordnet* και το *framenet*. Κάθε κατηγορία ρημάτων συνοδεύεται από πληροφορίες για θεματικούς ρόλους, πλαίσια με περιγραφές που

-
1. <https://framenet.icsi.berkeley.edu/fndrupal/>
 2. <http://wordnet.princeton.edu/>
 3. <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

περιλαμβάνουν συντακτικές και σημασιολογικές πληροφορίες, βασιζόμενο στις θεωρίες των Moens and Steedman (1998) και στην επέκταση και βελτίωση της θεωρίας της Levin(1993) με την προσθήκη υποκατηγοριών για την επίτευξη συντακτικής και σημασιολογικής συνοχής ανάμεσα στα μέλη κάθε κατηγορίας.

Τέλος χρησιμοποιήθηκαν και κάποια λεξικά όπως τα **oxforddictionaries**¹, **learners dictionaries of Merriam Webster**², **wordnik**³ και **freedictionaries**⁴.

Όλα αυτά συνετέλεσαν στην τελική διαμόρφωση των τριών insights των οποίων η αναλυτική παρουσίαση ακολουθεί.

DESIRE(επιθυμία)

Η συγκεκριμένη κατηγορία αποτυπώνει την επιθυμία του ομιλητή ή του γράφοντος να κάνει κάτι στο μέλλον. Είναι η κατηγορία στην οποία αποδίδεται **χαμηλή πιθανότητα(low probability)** για μία μελλοντική ενέργεια καθώς η επιθυμία του ατόμου από τη μία αποκαλύπτει την πρόθεσή του να κάνει ή να μην κάνει κάτι αλλά δεν εμπεριέχει μεγάλη δόση σιγουριάς. Στη κατηγορία της επιθυμίας εντάσσω και την έννοια της ευχής(wish) του ατόμου για κάτι μελλοντικό και μόνο καθώς τα όρια της ευχής για το μέλλον και της επιθυμίας είναι δυσδιάκριτα και αλληλοεπικαλυπτόμενα. Δεν περιλαμβάνω παραδείγματα προτάσεων που εκφράζουν ευχή για κάτι που έγινε ή δεν έγινε ή δεν έπρεπε να γίνει στο παρελθόν διότι αυτό δεν ανήκει στο πλαίσιο της συγκεκριμένης έρευνας. Τα λεξιλογικά στοιχεία (ρήματα, επίθετα, εκφράσεις) που εντάσσονται εκφράζουν την επιθυμία ή την ευχή να γίνει ή να μην γίνει κάτι στο μέλλον χωρίς όμως να εκφράζεται η πεποίθηση ότι σίγουρα θα γίνει αυτό που εκφράζει η πρόταση. Ρήματα όπως want, need, plead, implore, wish, επίθετα όπως excited, athirst και εκφράσεις όπως I have a dream toκ.ά. αποτελούν ενδεικτικά παραδείγματα που μέσα από τη χρήση τους και το σημασιολογικό τους βάρος δικαιολογούν την ένταξή τους στη συγκεκριμένη κατηγορία.

Παραδείγματα:

- I want to buy a new car.
- I need to find a new job.
- I wish I will buy a new car soon.

1 <http://www.oxforddictionaries.com/>

2 <http://www.merriam-webster.com/>

3 <https://www.wordnik.com/>

4 <http://www.thefreedictionary.com/>

Τέλος, στο πλαίσιο της επιθυμίας, συμπεριλαμβάνεται και η ευχή διότι οι επιθυμίες (και οι εκφράσεις ευχής) αναφέρονται σε μια μελλοντική πράξη βασιζόμενος στη δουλειά των Goldberg-Fillmore-Andrzejewski-Xu-Gibson-Zhu με τίτλο ‘*Μακάρι όλες σου οι ευχές να βγουν αληθινές: Μια μελέτη για τις ευχές και πώς να τις αναγνωρίζετε*’.

PURPOSE(σκοπός)

Η συγκεκριμένη κατηγορία αποτυπώνει την πρόθεση του ομιλητή ή του γράφοντος η οποία δεν έχει την μορφή επιθυμίας αλλά έχει τη μορφή του σκοπού, δηλαδή ότι υπάρχει ένας στόχος, ένα σχέδιο, το οποίο είναι προϊόν σκέψης. Και ως προϊόν σκέψης έχει μεγαλύτερη πιθανότητα να πραγματοποιηθεί στο μέλλον άρα εντάσσεται σε ένα ενδιάμεσο επίπεδο πιθανότητας, την **μέτρια πιθανότητα(medium probability)**. Στην κατηγορία του σκοπού(purpose) εντάσσονται επίσης και κάποια άλλα insights που για λόγους οικονομίας και καλύτερης επεξεργασίας των δεδομένων εντάχθηκαν στην ευρύτερη κατηγορία του σκοπού καθώς και αυτά εμπεριέχουν την ίδια έννοια στη χρήση τους σε πραγματικά κείμενα. Αυτά τα συγκεκριμένα insights είναι τα:

deciding/choosing(αυτά τα δύο έχουν δυσδιάκριτα όρια γιατί και τα δύο εκφράζουν την έννοια της επιλογής), **cogitation**(εκφράζει την σκέψη του ατόμου να κάνει κάτι), **seeking**(εκφράζει την πρόθεση του ατόμου να ζητήσει ή να αναζητήσει κάτι) και το **attempt**(εκφράζει την προσπάθεια που καταβάλλει το άτομο για κάτι). Όπως είναι φανερό όλα εμπεριέχουν την έννοια του σκοπού και για αυτό το λόγο εντάσσονται στην εν λόγω υπερκείμενη κατηγορία.

Ρήματα όπως think,cogitate, plan, επίθετα όπως focused, able και εκφράσεις όπως είναι My plan/target is to...είναι ενδεικτικά της λογικής της κατηγοριοποίησης.

Παραδείγματα:

- I am planning on buying a new phone.
- I am thinking of getting the new Iphone.
- I am interested in reading this book.
- My plan is to make a trip to England.

COMMITMENT(δέσμευση)

Η τελευταία κατηγορία περιλαμβάνει τα εκφραστικά μέσα εκείνα τα οποία κάνουν έκδηλη, την δέσμευση, την υποχρέωση του ατόμου να κάνει ή να μην κάνει κάτι στο μέλλον. Η δέσμευση συνεπάγεται όπως είναι φυσικό και την μεγαλύτερη πιθανότητα να συμβεί ή να μην συμβεί κάτι στο μέλλον συγκριτικά με τις δύο παραπάνω κατηγορίες. Για αυτό το λόγο της αποδίδεται η **μεγάλη/υψηλή πιθανότητα(high probability)**. Όταν το άτομο δεσμεύεται για κάτι, αυτόματα συνεπάγεται η πεποίθηση ότι η αυτό το οποίο λέγεται θα γίνει και συνεπώς να εξαχθούν τα πιο ασφαλή συμπεράσματα για την πρόθεση του ατόμου για μια μελλοντική πράξη. Ρήματα όπως swear, commit, επίθετα όπως, obliged, destined, dedicated, sticked, sworn, εκφράσεις όπως I have the obligation/determination to..., I commit myself to... αποτελούν ενδεικτικά παραδείγματα της τελευταίας κατηγορίας.

Παραδείγματα:

- I sworn to secrecy.
- I am obligated to buy a new car.
- I have the determination for something better.
- We will dedicate our lives to keep up the discipline.

3.3 Λεξικό-lexical resources

Για την κατασκευή των κανόνων χρειάστηκε προηγουμένως να διενεργηθεί η κατασκευή ενός λεξικού βάση του οποίου θα κατασκευάζονταν και οι κανόνες. Στην πρώτη –και αποτυχημένη- απόπειρα έγινε μία προσπάθεια να γίνει μια αμιγώς σημασιολογική κατηγοριοποίηση όλων των entries που εκφράζουν «πρόθεση» σύμφωνα με τα δεδομένα και την μελέτη που εκπονήθηκε. Στην συνέχεια όμως αποφασίστηκε να γίνει μία κατηγοριοποίηση χρησιμοποιώντας συντακτικά κριτήρια δηλαδή ανάλογα με το τι ακολουθεί το entry και την συντακτική του λειτουργία στην πρόταση. Παράλληλα έγινε και ο διαχωρισμός και η ένταξη του κάθε entry στο ανάλογο insight. Σε όλη αυτή την διαδικασία και σε ό, τι περιγραφεί στη συνέχεια βασιστήκαμε σε wordnet, λεξικά αγγλικής γλώσσας(oxforddictionary, freedictionary) και σε δεδομένα από τα σώματα κειμένων που συλλέχθηκαν και μελετήθηκαν για αυτόν ακριβώς τον σκοπό..

Το λεξικό περιέχει ρήματα, επίθετα και εκφράσεις που δηλώνουν την πρόθεση κάποιου να κάνει κάτι. Κάθε ένα από αυτά περιέχει πληροφορία για το πώς συντάσσεται και σε ποιο insight ανήκει.

Η λογική που ακολουθήθηκε ήταν ότι εκκινώντας από έναν συγκεκριμένο αριθμό όρων εκκίνησης (seed terms) εμπλουτίστηκαν οι κατηγορίες των ρημάτων και των επιθέτων με σημασιολογικά συγγενείς καταχωρήσεις βασιζόμενοι πάντα στις πληροφορίες που αντλούνταν από τα λεξικά που χρησιμοποιήθηκαν, από το framenet ώστε να τα κατατάξουμε στο κατάλληλο insight, και από τη χρήση τους σε πραγματικό κείμενο. Στη συνέχεια εντοπίσαμε τους διάφορους τρόπους σύνταξης της κάθε μιας καταχώρησης ξεχωριστά και ομαδοποιήθηκαν προκειμένου αυτό το είδος της μοντελοποίησης να περαστεί στους κανόνες.

Όσον αφορά τις εκφράσεις χρησιμοποιήθηκε πληροφορία τόσο από λεξικά όσο και από την παρατήρηση χρήσης τους στα σώματα κειμένων που αντλήθηκαν από το Διαδίκτυο για την εκπόνηση της εργασίας.

✓ Ρήματα

Συντακτικές κατηγορίες	Επεξήγηση	Παραδείγματα
<i>VB11a</i>	<i>Μετά το ρήμα ακολουθεί ουσιαστικό (NN) που μπροστά μπορεί να παίρνει προαιρετικά άλλους όρους και μέρη του λόγου ως προσδιοριστικά.</i>	I am starving for a coffee.
<i>VB11b</i>	<i>ρήμα συντάσσεται μετά με to+ ρήμα.</i>	I am about to change my life
<i>VB11c</i>	<i>Το ρήμα ακολουθείται από σκέτο ρηματικό τύπο.</i>	
<i>VB11ab</i>	<i>Το ρήμα μπορεί να εμφανίζει και την σύνταξη όπου μετά ακολουθεί ουσιαστικό αλλά και την σύνταξη to+ ρήμα.</i>	I want to buy a new phone. I want a new phone.
<i>VB12a</i>	<i>ρήμα μπορεί να ακολουθείται από</i>	I wish for the

	<i>for?+ουσιαστικό ή από PRP+ρήμα.</i>	best. I wish I will buy a new car.
<i>VBI3a</i>	<i>Το ρήμα ακολουθείται από for+ουσιαστικό και to+ρήμα.</i>	I opt to buy a new home. I opt for a new car.
<i>VBI4</i>	<i>Ειδική κατηγορία που περιλαμβάνει τύπους που λειτουργούν ως βοηθητικά στο ρήμα.</i>	
<i>VBI5</i>	<i>Το ρήμα συντάσσεται με to?+ ουσιαστικό ή με PRP+ ρήμα.</i>	I swear to my life. I swear I will buy a new phone.
<i>VBI6</i>	<i>Το ρήμα συντάσσεται μόνο με PRP+ ρήμα.</i>	I guess I will find a new job.
<i>VBI7</i>	<i>Το ρήμα μπορεί να εμφανίζει και την σύνταξη όπου μετά ακολουθεί ουσιαστικό αλλά και την σύνταξη to+ ρήμα + NN? με την διαφορά ότι περιορίζουμε το πρόσωπο του υποκειμένου μόνο στο πρώτο ενικό μόνο για την συγκεκριμένη κατηγορία(PRS2).</i>	I need a phone. I need to buy a new phone.

- ❖ Στη συνέχεια τα ρήματα που συλλέχθηκαν εκτός από την ταξινόμησή τους στις ανάλογες κατηγορίες ανάλογα με τον τρόπο που συντάσσονταν κατανεμήθηκαν και στα ανάλογα insights βασιζόμενη στο σημασιολογικό βάρος που κατείχαν σύμφωνα με το framenet και το verbnet και σύμφωνα με την παρατήρηση των σωμάτων κειμένων που κατασκευάστηκαν. Στη συνέχεια αυτή η πληροφορία αποτυπώθηκε σε ένα αρχείο excel.

-Ένα μικρό δείγμα αυτής της πληροφορίας:

ENTRIES	POS	VB_type	insight
want	VB	VBI1ab	intention/desire
aim	VB	VBI3a	intention/purpose
aim at	PVB	VBI1a	intention/purpose
aspire	VB	VBI1b	intention/desire

be about	VP	VBI1b	intention/purpose
be into	VP	VBI1a	intention/purpose
be on	VP	VBI1b	intention/purpose
be up	VP	VBI1b	intention/purpose
can't/cannot wait	VP	VBI3a	intention/desire
cogitate on	VP	VBI1c	intention/purpose
cogitate on	VP	VBI1a	intention/purpose
consider	VB	VBI1c	intention/purpose
contemplate	VB	VBI1c	intention/purpose
covet	VB	VBI3a	intention/desire
crave	VB	VBI3a	intention/desire
decide	VB	VBI1b	intention/purpose

Το σύνολο των εγγραφών ανέρχεται σε 174.

✓ Επίθετα

Συντακτικές κατηγορίες	Επεξήγηση	Παραδείγματα
JJ1a	<i>Το επίθετο συντάσσεται με to+ ρήμα/ουσιαστικό.</i>	I am willing to buy a new car. I am stucked to this plan.
JJ1b	<i>Το επίθετο συντάσσεται με to+ρήμα και for+ουσιαστικό.</i>	I am ready to get this shirt. I am ready for a new start.
JJ1c	<i>επίθετο συντάσσεται με on/υρον+ρήμα και με on/υρον+ ουσιαστικό.</i>	I am focused on buying this car. I am keen on this project.
JJ1d	<i>Το επίθετο συντάσσεται με in+ρήμα/ ουσιαστικό.</i>	I am interested in buying a new home. I am interested in this album.
JJ1e	<i>Το επίθετο συντάσσεται με about/for/of/with+ουσιαστικό.</i>	I am crazy for this new dress.

- ❖ Η ίδια διαδικασία που ακολουθήθηκε για τα ρήματα εφαρμόστηκε και για τα επίθετα με την μόνη διαφορά ότι σε κάθε μία κατηγορία επιθέτων δημιουργήθηκε άλλη μία αντίστοιχη με ίδια σύνταξη και ίδια σημασιολογική κατανομή που περιλαμβάνει και ενσωματώνει την πληροφορία της άρνησης(negation) η οποία θα αναλυθεί εκτενέστερα στο κεφάλαιο όπου περιγράφονται οι κανόνες.

-Το σύνολο των εγγραφών ανέρχεται σε 116.

- Ένα μικρό δείγμα αποτελεί το παρακάτω:

ENTRIES	POS	JJ_type	insights
determined	JJ	JJ1a	intention/commitment
dictated	JJ	JJ1a	intention/commitment
willing	JJ	JJ1a	intention/desire
disposed	JJ	JJ1a	intention/commitment
predisposed	JJ	JJ1a	intention/commitment
resolved	JJ	JJ1a	intention/commitment
obliged	JJ	JJ1a	intention/commitment
inclined	JJ	JJ1a	intention/commitment
devoted	JJ	JJ1a	intention/commitment
pledged	JJ	JJ1a	intention/commitment
meant	JJ	JJ1b	intention/purpose
aspired	JJ	JJ1a	intention/desire
set	JJ	JJ1a	intention/commitment
driven	JJ	JJ1a	intention/purpose

✓ Εκφράσεις

Συντακτικές κατηγορίες	Επεξήγηση	Παραδείγματα
<i>EXP1a</i>	<i>Η έκφραση ακολουθείται από to+ ρήμα/ουσιαστικό.</i>	I will make a commitment to get this desktop soon.
<i>EXP1b</i>	<i>Η έκφραση ακολουθείται από ρήμα/ουσιαστικό.</i>	I consider the possibility of buying a new car.
<i>EXP1c</i>	<i>Η έκφραση ακολουθείται από</i>	I am armed and ready to

	<i>to+ρήμα ή for+ουσιαστικό.</i>	start a new life. I am ready and willing for a new beginning.
<i>EXP1d</i>	<i>Η έκφραση ακολουθείται από PRP+ρήμα.</i>	It's about time we buy a new washing machine.

- ❖ Στις εκφράσεις η διαδικασία είναι πανομοιότυπη με την διαδικασία που ακολουθήθηκε σε επίθετα και επιρρήματα με την διαφορά ότι :
 - Διαμορφώθηκαν για τις τρεις πρώτες κατηγορίες οι αντίστοιχές τους EXP2a, EXP2b και EXP2c που αναφέρονται σε εκφράσεις που είναι παγιωμένες και δεν παίρνουν υποκείμενο I/we. Π.χ There is a possibility to buy a new home.
 - Εντάχθηκε στους κανόνες και η πληροφορία για την άρνηση (negation) όπου θα αναφερθώ εκτενέστερα στο κεφάλαιο περιγραφής των κανόνων και της κατασκευής της Intention Grammar.

ENTRIES	EXP_ty	insight
make/have a commitment/promise	EXP1a	intention/commitment
have the(JJ)(a/an) /intention/purpose/plans....	EXP1c	intention/purpose
make an (JJ)attempt/effort	EXP1c	intention/purpose
It's (about) time	EXP1c	intention/commitment
consider the possibility/theory/chance of	EXP1b	intention/purpose
My (JJ)plan/ambition/purpose/goal/target/purpose/ is	EXP1a	intention/purpose
consider the possibility/theory/chance of	EXP1b	intention/purpose

Οι εγγραφές στις εκφράσεις ξεπερνούν τις 70.

3.4 Intention Grammar

Το στάδιο αυτό περιλαμβάνει την δημιουργία των κανόνων με βάση τους οποίους θα γίνει η εξαγωγή της πρόθεσης του γράφοντος ή του ομιλούντος προσώπου. Η κατασκευή των λεξικών πόρων είναι ένα ζωτικής σημασίας κομμάτι για τον εντοπισμό και την εξόρυξη του Intention και κατ' επέκταση των πτυχών του αλλά δεν είναι δυνατόν να εξαχθούν αξιόπιστα αποτελέσματα χωρίς την δημιουργία συντακτικών κανόνων (Intention Rules) που θα λειτουργήσουν και θα υποβοηθηθούν παράλληλα από τα Lexical Resources η δημιουργία των οποίων προηγήθηκε. Η δημιουργία των κανόνων και ο εμπλουτισμός των λεξικών λειτουργεί παράλληλα με την εξαγωγή της πρόθεσης.

Οι συντακτικοί κανόνες αποτελούν την αποτύπωση της επιφανειακής συντακτικής δομής των προτάσεων η μορφή των οποίων αποκρυσταλλώθηκε μέσα από την προσεκτική μελέτη και παρατήρηση των παραδειγμάτων που συλλέχθηκαν και οργανώθηκαν σε σώματα κειμένων για τον σκοπό της διπλωματικής.

Στους κανόνες αυτούς γίνεται χρήση του taggset του gate αλλά και των λεξικών από τα οποία αντλούνται τα ρήματα, τα επίθετα και οι εκφράσεις οι οποίες είναι φορείς πρόθεσης μαζί με την πληροφορία για τα insights. Επίσης οι κανόνες είναι δομημένοι σύμφωνα με τις συντακτικές κατηγορίες που ορίστηκαν στο λεξικό σχετικά με τον τρόπο σύνταξης όλων αυτών των λεξικών στοιχείων και τι έπεται αυτών.

Τέσσερα είναι τα βασικά σημεία στα οποία οι κανόνες καλούνται να φέρουν αποτέλεσμα.

- ✚ Το πρώτο εξ' αυτών είναι ο εντοπισμός του **intention_predicate**, δηλαδή του **κατηγορήματος** το οποίο είναι φορέας της πρόθεσης καθώς και το είδος της πρόθεσης σύμφωνα με το insight στο οποίο έχει καταχωρηθεί στο λεξικό το κάθε ένα λεξικό στοιχείο (ρήμα, επίθετο, έκφραση). Όλη αυτή η πληροφορία εμπεριέχεται στα λεξικά τα οποία τρέχουν παράλληλα με τους κανόνες ανάλογα με το λεξικό στοιχείο το οποίο κάθε φορά συναντάται.

- Παράδειγμα:
 - I want.....,
 - I need.....,
 - I am determined.....,

○ I have the determination.....

✚ Η εμφάνιση της συντακτικής κατηγορίας στην οποία ανήκει το κατηγορήμα.

✚ Σε ένα επόμενο επίπεδο ανάλυσης στους κανόνες αναζητάμε το **object_of_intention** στο βαθμό που αυτό είναι δυνατό. Το αντικείμενο της πρόθεσης είναι το μέρος του λόγου εκείνο στο οποίο μεταφέρεται η πρόθεση του ομιλούντος ή του γράφοντος, δηλαδή το αποτέλεσμα της μελλοντικής του ενέργειας ή πράξης. Αυτό είναι ένα θέμα το οποίο χρήζει ιδιαίτερης προσοχής λόγω της ιδιαιτερότητάς του και θα συμπεριληφθεί σε αυτά που μπορούμε να κάνουμε στο μέλλον για βελτίωση του υπάρχοντος εργαλείου.

Παράδειγμα:

I want to buy an I phone..
↑ ↑ ↑
predicate Obj 1 Obj2

✚ Τέλος εστίαση υπάρχει στο φαινόμενο της **άρνησης(negation)** τόσο στη δομή των κανόνων για χάρη του οποίου δημιουργήθηκε ξεχωριστό λεξικό με κατάλογο λέξεων και δομών που λειτουργούν ως negators, όσο και στην έκφραση της άρνησης εσωτερικά στα λεξικά τόσο στα επίθετα(unable, unsworn), στο επίπεδο των εκφράσεων(I have no plans for tonight, There is no chance of buying the new Iphone etc.), όσο και στο επίπεδο της πρότασης όπου με βάση την παρατήρηση των παραδειγμάτων προβλέφθηκαν στους κανόνες όλες οι πιθανές θέσεις εμφάνισης της άρνησης.

Για το λόγο αυτό φτιάχτηκε ένα ειδικό λεξικό το οποίο ονομάστηκε **negators** και το οποίο περιέχει μέσα σε μορφή json τα είδη των αρνήσεων τα οποία και ήταν σε συχνότερη χρήση.

3.5 Intention Rules

Η δημιουργία των κανόνων αποτέλεσαν την κατάληξη του συνόλου της εκπόνησης της εργασίας και ήταν απόρροια της παρατήρησης των κειμένων και του τρόπου με τον οποίο οι χρήστες του διαδικτύου εξέφραζαν την πρόθεσή τους να κάνουν κάτι στο μέλλον.

Πριν όμως από την τελική διαμόρφωση των κανόνων έγιναν κάποιες δοκιμές με πιο απλές δομές, τις λεγόμενες πρότυπες δομές(patterns) που αποτύπωναν την πολύ βασική δομή των προτάσεων που αναζητούσαμε ώστε να διασφαλιστεί η σωστή πορεία της έρευνας.

Για παράδειγμα στα ρήματα ανάλογα με το πώς συντάσσονταν ζητούσαμε τις πολύ βασικές δομές όπως είχαν οριστεί στις κατηγορίες των lexical resources έπειτα από την παρατήρηση των σωμάτων κειμένων. Το ίδιο έγινε και στα επίθετα αλλά και στις εκφράσεις. Με αυτό τον τρόπο κατέστη εφικτό να εμπλουτίσουμε τους κανόνες με περισσότερα χαρακτηριστικά ώστε να γίνουν όσο το δυνατόν πιο ακριβείς γίνεται με βάση την ποσότητα των δεδομένων που υπήρχαν στη διάθεσή μας.

Συνολικά τα patterns είναι 19 και περιέχονται αναλυτικά στο φάκελο jape που έχει φτιαχτεί με το **όνομα bootstrapping¹ patterns** τα οποία βοήθησαν στον εμπλουτισμό των κανόνων και στην διαμόρφωση της τελικής δομής τους.

Παραδείγματα patterns από ρήματα και επίθετα:

PRS(I/We)+ VB(!Intentionlexicon)+ DT?+JJ?+NN:

- Στο συγκεκριμένο κανόνα ζητάω να μου φέρει πίσω οποιαδήποτε δομή που η σειρά των όρων να έχει ως υποκείμενο το I/We, ρήμα που να μην ανήκει στο λεξικό και μετά ουσιαστικό υποχρεωτικά σαν αντικείμενο με τους όρους που προηγούνται να είναι προαιρετικοί.

PRS(I/We)+ VB(!Intentionlexicon)+ TO|IN?+VB + DT?+JJ?+NN?

- Στο συγκεκριμένο κανόνα ζητάω να μου φέρει πίσω οποιαδήποτε δομή που η σειρά των όρων να έχει ως υποκείμενο το I/We, ρήμα που να μην ανήκει στο λεξικό , έπειτα το ή άλλη πρόθεση με ρηματικό τύπο που θα είναι το αντικείμενο της πρόθεσης και προαιρετικά ουσιαστικό το οποίο θα είναι δεύτερο αντικείμενο της πρόθεσης.

1. Η τεχνική του bootstrap στη στατιστική αποτελεί ένα σύγχρονο τρόπο λήψης δειγμάτων. Bootstrapping είναι η μέτρηση των ιδιοτήτων ενός εκτιμητή που δειγματοληπτούνται από μία κατά προσέγγιση κατανομή.

PRS(I/We)+ be+ JJ(!Intentionlexicon)+ TO|IN?+VB + DT?+JJ?+NN?

- Στο συγκεκριμένο κανόνα ζητάω να μου φέρει πίσω οποιαδήποτε δομή που η σειρά των όρων να έχει ως υποκείμενο το I/We, επίθετο που να μην ανήκει στο λεξικό , έπειτα το ή άλλη πρόθεση με ρηματικό τύπο που θα είναι το αντικείμενο της πρόθεσης και προαιρετικά ουσιαστικό το οποίο θα είναι δεύτερο αντικείμενο της πρόθεσης.

Τα συγκεκριμένα patterns μας βοήθησαν να εντοπίσουμε πιθανές συντακτικές δομές που δεν είχαν προβλεφθεί στους κανόνες που ήδη είχαν κατασκευαστεί ώστε αν εντοπιζόταν κάτι να περιλαμβανόταν στο ήδη δημιουργημένο σύνολο των υπαρχόντων κανόνων.

Συνολικά δημιουργήθηκαν **59 κανόνες** οι οποίοι είχαν τις εξής λειτουργίες και έφεραν πίσω τα εξής αποτελέσματα:

- Υπήρχε αυστηρός περιορισμός στα πρόσωπα τα οποία λειτουργούσαν ως υποκείμενα στο κατηγορημα της πρότασης (ρήμα). Τα πρόσωπα αυτά ήταν το πρώτο ενικό και πληθυντικό(I and We). Τα υπόλοιπα θεωρήθηκε ότι δεν εξέφραζαν την άμεση πρόθεση του γράφοντος αλλά απλά μετέφεραν την πρόθεση κάποιου τρίτου ατόμου για την οποία δεν υπήρχε η βεβαιότητα ότι ίσχυε και στην πραγματικότητα. Επομένως εστίαζε σε προτάσεις που είχαν ως υποκείμενα ή το I ή το We.
- Βρίσκουν το ρήμα(predicate), το επισημαίνουν και επιστρέφει την πληροφορία σχετικά με την κατηγορία που ανήκει το ρήμα ανάλογα με:
 - A) την συντακτική κατηγορία στην οποία ανήκει ανάλογα με τον τρόπο που συντάσσεται το ρήμα και
 - B) την σημασιολογική κατηγορία στην οποία ανήκει, δηλαδή τι είναι αυτό που εκφράζει εννοιολογικά (desire, purpose, commitment).Η συγκεκριμένη πληροφορία έχει περασθεί στα λεξικά σε μορφή jape.

- Τέλος βρίσκει και επισημαίνει το αντικείμενο ή τα αντικείμενα του intention δηλαδή την οντότητα ή τις οντότητες στην οποία ή στις οποίες μεταβαίνει η ενέργεια του intention. Στο object of intention όπως καθορίστηκε να ονομάζεται δεν έχει προστεθεί πληροφορία ούτε κάποια κατηγοριοποίηση(ex. Purchase, sell, quit etc.) ανάλογα με αυτό που εκφράζει η πρόταση πράγμα που θα ήταν επιθυμητό να γίνει στο μέλλον σε βελτίωση της παρούσας δουλειάς (εκτενέστερη αναφορά στο κεφάλαιο με τις μελλοντικές προσεγγίσεις).

Χρήσιμο θα ήταν να αναφερθεί ότι ο συμβολισμός των διαφόρων μερών του λόγου ώστε να δομηθούν οι συντακτικοί κανόνες έγινε με βάση με βάση το Penn Treebank tag set (Marcus et al.1993) που αποτελείται από 45 κατηγορίες και είναι οι εξής:

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential there
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP\$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	to
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP\$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Γραμμική αναπαράσταση των κανόνων:

Rule: intexp1

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)?
(VBI1ab|VBI1a) (CD)? (IN)?(DT) ? (PRP)? (RB)? (JJ)? (NN)

I want a new phone.

I really want a new phone.

I may not need a phone.

I don't want a new phone.

I never ever want a phone.

I would like a car.

I am searching for a new home.

Rule: intexp2

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)?
(VBI1ab|VBI1b) (neg)? (to) (RB)? (VB) (TO/IN)? (CD)? (IN)?(DT) ? (PRP)? (RB)? (JJ)? (NN)

I'd love to come.

I have to buy a new phone.

I 've got to get a new car.

I am going to buy a phone.

I don't have to buy a new phone.

I may not have to buy a new phone.

I will have to buy a new phone.

I am going to buy a phone.

Rule: intexp3

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI1c) (neg)? (VB) (RB)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN/VBN)

I am looking forward to get a new bike.

I' m planning on buying a car.

I am thinking of getting married.

I consider moving abroad.

I contemplate moving abroad.

I may consider moving abroad.

I may not consider moving abroad.

I am not thinking of getting married.

I am thinking of not getting married.

I don't consider going out.

I don't contemplate going out.

Rule: intexp4

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI2a) (for)?(be)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)

I may hope for a new job.

I may not consider for a new job.

I wish for a car.

I don't wish a new car.

Rule: intexp5

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI2a)
 (TO)? (PRP)? (that)? (PRS1) (punct)? (neg)? (VBI4)? (could)?(neg)? (be)?
 (RB)? (neg)? (VB) (RB)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)?
 (NN|PRP)

. I wish I will buy a new phone.

I think I may buy a new phone.

I wish I could move abroad.

Rule: intexp6

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI3a) (for)
 (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)

I am not opting for a coffee.

I can't wait for the new phone.

I don't crave for mercy.

I' m thinking for a new start.

Rule: intexp7

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI3a) (neg)?
 (to) (RB)? (VB) (RB)? (be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)?
 (JJ)? (NN|PRP)

. I am craving to buy a new phone.

I' m thinking to buy a car.

I don't opt to get a phone.

Rule: intexp8

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VB!Lexicon)
(RB)? (neg)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)?
(NN|PRP)

I gotta buy a new car.

I wanna shop around.

I don't wanna get married.

Rule: intexp9

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (JJ1a) (be)?
(neg)? (to) (RB)? (VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)?
(RB)? (JJ)? (NN|PRP)?

I am determined to do it.

I am determined not to do it.

I' m not determined to do it.

I am decided to buy a new house.

I am bound to buy a new car.

Rule: intexp10

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB|VBG)? (neg)? (JJ1a)
(neg)? (to) (be)?(neg)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)?
(NN|PRP)?

I am devoted to this plan.

I am not devoted to this plan.

I may not be stuck to this project.

Rule: intexp11

(PRS1) (punct)? (VBI4)? (be)? (RB|VBG)? (JJ1aN) (to) (RB)? (VB) (RB)? (be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I am unbound to buy a new car.

I am unable to drive a car.

Rule: intexp12

(PRS1) (punct)? (VBI4)? (be)? (RB|VBG)? (JJ1aN) (to) (be)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I am undevoted to this plan.

I am undetermined to this project.

Rule: intexp13

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB|VBG)? (neg)? (JJ1b) (be) ?(neg)? (to) (RB)? (VB) (RB)? (be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)

I am eager to buy a new car.

I' m ready to start a new life.

Rule: intexp14

(PRS1) (punct)? (VBI4)? (be)? (RB|VBG)? (JJ1bN) (to) (RB)? (VB) (RB)? (be)?(TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I am unavailable to come.

Rule: intexp15

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB|VBG)? (neg)? (JJ1b) (be) (neg)? (for) (be)? (neg)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I am ready for a new trip.

I am not ready for a new trip.

I may not be ready for a new trip.

I am not eager for a new trip.

Rule: intexp16

(PRS1) (punct)? (VBI4)? (be)? (RB|VBG)? (JJ1bN) (be)? (for) (be)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I am unprepared for a new job.

I am unavailable for a trip.

Rule: intexp17

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB|VBG)? (neg)? (JJ1c) (be)? (neg)? (on|upon) (be)? (neg)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I am focused on this plan.

I am keen on golf.

I am not focused upon this plan.

Rule: intexp18

(PRS1) (punct)? (VBI4)? (be)? (RB|VBG)? (JJ1cN) (be)? (on|upon) (be)? (CD)? (IN)? (DT) ? (PRP)? (JJ)? (RB)? (NN|PRP)?

I am unfocused on this plan.

I am unkeen on golf.

Rule: intexp19

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB|VBG)? (neg)? (JJ1c)
(be)? (neg)? (on|upon) (be)? (RB)? (VB) (RB)? (TO/IN)? (CD)? (IN)? (DT)
? (PRP)? (RB)? (JJ)? (NN|PRP)

I may be focused on buying a new car.

I may not be focused on buying a new car.

Rule: intexp20

(PRS1) (punct)? (VBI4)? (be)? (RB|VBG)? (JJ1cN) (be) (on|upon)(be)?
(RB)? (VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)?
(NN|PRP)?

I am unfocused on buying a new car.

I am unkeen on buying a new car.

Rule: intexp21

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB|VBG)? (neg)? (JJ1d)
(be)? (neg)? (in) (be)? (neg)? (CD)? (IN)? (DT) ? (PRP)? (JJ)? (RB)?
(NN|PRP)?

I am interested in golf.

I am not interested in golf.

Rule: intexp22

(PRS1) (punct)? (VBI4)? (be)? (RB|VBG)? (JJ1dN) (in) (be)? (CD)?
(IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I am uninterested in golf.

Rule: intexp23

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB|VBG)? (neg)? (JJ1d)
(neg)? (in) (be) (neg)? (RB)? (VB) (RB)?(neg)? (TO/IN)? (CD)? (IN)?
(DT) ? (PRP)? (RB)? (JJ)? (NN)

. I am interested in buying a new car.

I may not be interested in getting a new phone.

Rule: intexp24

(PRS1) (punct)? (VBI4)? (be)? (RB|VBG)? (JJ1dN) (in) (be)? (RB)?
(VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (JJ)? (RB)?
(NN|PRP)?

I may be uninterested in getting a new phone.

Rule: intexp25

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB|VBG)? (neg)? (JJ1e)
(neg)? (about|for|of|with) (be)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)?
(NN|PRP)?

I am ravenous for a new start.

I may not be esurient for a new start.

I am covetous for a burger.

I am excited with this movie.

Rule: intexp26

(PRS1) (punct)? (VBI4)? (be)? (RB|VBG)? (JJ1eN) (about|for|of|with)
(be)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I am unambitious for this plan.

Rule: intexp27

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (EXP1a)
(neg)? (to) (RB)? (VB) (RB)?(be)? (neg)? (TO/IN)? (CD)? (IN)? (DT) ? (
PRP)? (RB)? (JJ)? (NN|PRP)?

I have the determination to sell my house.

It's my obligation to buy a new car.

My dream is to buy a new car.

My dream is not to buy a new car.

Rule: intexp28

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (EXP1a)
(neg)? (to) (be)?(neg)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)?
(NN|PRP)?

I have the determination to this plan.

Rule: intexp29

(PRS1) (punct)? (VBI4)? (be)? (RB)? (EXP1aN) (be)? (to) (RB)? (VB)
(RB)? (be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)?
(NN|PRP)?

I have no determination to sell my house.

It's not my obligation to help you.

Rule: intexp30

(PRS1) (punct)? (VBI4)? (be)? (RB)? (EXP1aN) (to) (be)? (CD)? (IN)?
(DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I have no determination to this plan.

Rule: intexp31

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (EXP1b)(be)? (neg)? (CD)? (IN)? (DT)? (PRP)? (RB)? (JJ)? (NN|PRP)?

I examine the possibility of a new job.

I may not examine the possibility of a new trip.

Rule: intexp32

(PRS1) (punct)? (VBI4)? (be)? (RB)? (EXP1bN?) (be)? (CD)? (IN)? (DT)? (PRP)? (RB)? (JJ)? (NN|PRP)?

There is no chance of a new trip.

Rule: intexp33

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (EXP1b) (neg)?(PRP)? (RB)? (VB) (RB)? (TO/IN)? (CD)? (IN)? (DT)? (PRP)? (RB)? (JJ)? (NN|PRP)?

I consider the possibility of buying a new house.

I may not examine the possibility of buying a new car.

I am likely to acquire a new phone.

I'd rather buy a new car.

I'd better buy a new car.

I would rather not buy a new car.

I'd better not miss the festival.

Rule: intexp34

(PRS1) (punct)? (VBI4)? (be)? (RB)? (EXP1bN?) (VB) (TO/IN)? (CD)? (IN)? (DT)? (PRP)? (RB)? (JJ)? (NN|PRP)?

There is no chance of getting a new phone.

Rule: intexp35

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (EXP1c) (be) (neg)? (to) (RB)? (VB) (RB)? (be)? (neg)?(TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I have plans to buy a new home.

I am armed and ready to get my phone.

I may not be armed and ready for a new beginning.

I have the strong intention to get a new home.

Rule: intexp36

(PRS1) (punct)? (VBI4)? (be)? (RB)? (EXP1cN) (to) (RB)? (VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

It's not time to buy a new car.

I've got no plans to sell my home.

Rule: intexp37

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (EXP1c)(be)? (neg)? (for)(be)? (neg)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I've got plans for a new trip.

I don't have plans for something new.

It's about time for a new start.

Rule: intexp38

(PRS1) (punct)? (VBI4)? (be)? (RB)? (EXP1cN?) (be)?(for)(be)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I've got no plans for a new trip.

It's not time for a new start.

It's not time to buy a new home.

It's not about time to sell our home.

Rule: intexp39

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (EXP1d)
(PRS1) (punct)? (neg)? (VBI4)? (could)? (neg)? (be)? (RB)? (neg)? (VB)
(RB)? (RB)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

It's time we start a new job.

It's about time I buy a new car.

It's high time we sell our house.

Rule: intexp40

(PRS1) (punct)? (VBI4)? (be)? (RB)? (EXP1d) (PRS1) (punct)? (neg)?
(VBI4)?(could)? (neg)? (be)? (RB)? (neg)? (VB) (RB)? (CD)? (IN)? (DT)
? (PRP)? (RB)? (JJ)? (NN|PRP)?

It's not time we start a new job.

It's not about time I buy a new car.

It's not high time we sell our house.

Rule: intexp41

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI5) (to)?
(CD)? (IN)? (DT) ? (PRP)? (JJ)? (NN|PRP)?

I swear to my life.

I swear to my children.

Rule: intexp42

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI5)
(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VB) (RB)?
(CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I swear I will find a new job.

I promise I will start a new life.

I may not swear to find a new job.

I promise to start a new life.

Rule: intexp43

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI6)
(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VB) (RB)?
(CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I guess i may start a new life.

I don't guess i may start a new life.

Rule: intexp44

(PRS1) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (JJ1e) (neg)?
(about|for|of|with)(be)?(neg)? (RB)? (VB)? (RB)? (be)?(neg)?(TO/IN)?
(CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

I am capable of buying a new car.

I am not capable of buying a new car.

Rule: intexp45

(PRS1) (punct)? (VBI4)? (be)? (RB)? (JJ1eN) (about|for|of|with)
(be)? (RB)? (VB)? (RB)? (be)?(TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)?
(JJ)? (NN|PRP)?

I am incapable of buying a new car.

Rule: intexp46

(EXP2a) (to) (RB)? (VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

It's my dream to buy a new car.

Rule: intexp47

(EXP2a) (to) (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

Rule: intexp48

(EXP2aN) (to) (RB)? (VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

It may not be my dream to get a new computer.

Rule: intexp49

(EXP2aN) (to) (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

Rule: intexp50

(EXP2b) (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

It may be my choice for this new start.

Rule: intexp51

(EXP2bN) (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

It may not be my choice for this new start.

Rule: intexp52

(EXP2b) (PRP)? (RB)? (VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

There is chance of buying a new car.

Rule: intexp53

(EXP2bN) (PRP)? (RB)? (VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

There may not be a possibility of buying a good phone.

Rule: intexp54

(EXP2c) (to) (RB)? (VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

There is a good chance to buy one of these phones.

Rule: intexp55

(EXP2cN) (to) (RB)? (VB) (RB)?(be)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

There is not a good chance to buy one of these phones.

Rule: intexp56

(EXP2c) (for) (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

There is a good chance for a brand new start.

Rule: intexp57

(EXP2cN) (for) (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

There is not a good chance for a brand new start.

Rule: intexp58

(PRS2) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI7) (be)? (neg)? (be)?(neg)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

All I need is a brand new life.

Rule: intexp59

(PRS2) (punct)? (neg)? (VBI4)? (neg)? (be)? (RB)? (neg)? (VBI7) (be)? (neg)? (to) (RB)? (VB) (RB)? (be)?(neg)? (TO/IN)? (CD)? (IN)? (DT) ? (PRP)? (RB)? (JJ)? (NN|PRP)?

All I need is to buy a brand new car.

3.6 ΣΥΝΤΟΜΗ ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ GATE

Το **Gate**¹ αποτελεί την πλατφόρμα που χρησιμοποιήθηκε για την διεξαγωγή των πειραμάτων που απαιτούνταν για την εκπόνηση και την ολοκλήρωση της διπλωματικής. Το Gate αποτελεί ένα πολύ χρήσιμο εργαλείο καθώς χρησιμοποιείται πάνω από 15 έτη σε όλες τις υπολογιστικές εργασίες συμπεριλαμβανομένης και της επεξεργασίας του ανθρώπινου λόγου. Το Gate εξειδικεύεται στην ανάλυση κειμένων ανεξαρτήτως μορφής και μεγέθους τους. Το μεγάλο του πλεονέκτημα είναι ό, τι διατίθεται δωρεάν προς εγκατάσταση και χρήση μέσω διαδικτύου σε όσους το έχουν ανάγκη και βελτιώνεται συνεχώς αποτελώντας έναν πολύτιμο σύμμαχο σε όσους επιχειρήσουν να ασχοληθούν με συναφή θέματα.

Από το Gate χρησιμοποιήθηκαν τα παρακάτω εργαλεία για την προεπεξεργασία των κειμένων που χρησιμοποιήσα κατά την διάρκεια των πειραμάτων που διενεργήθησαν καθ' όλη την διάρκεια της διπλωματικής:

- Τον **Tokenizer** ο οποίος χωρίζει το κείμενο σε tokens (αριθμοί , σημεία στίξης , λέξεις διαφόρων κατηγοριών).
- Τον **Sentence Splitter** ο οποίος διαχωρίζει το κείμενο σε προτάσεις.
- Τον **Part of Speech Tagger** που διακρίνει τις λέξεις ανάλογα με το μέρος του λόγου στο οποίο ανήκει η κάθε μία.
- Τον **Gazetteer** που χρησιμοποιείται για να επισημανθούν οι ονοματικές οντότητες.
- Τον **Morphological Analyzer** για τη διάκριση των μορφολογικών κατηγοριών.

Όλη αυτή η διαδικασία ήταν απαραίτητη προκειμένου τα κείμενα να είναι έτοιμα ώστε να αναλάβει το ρόλο του το JAPE TRANSDUCER το οποίο παίρνοντας το αρχείο των κανόνων(rules) το οποίο ήταν γραμμένο σε μορφή **jape file** να δημιουργήσει μία μηχανή αυτόματης εξόρυξης της πρόθεσης(intention) από τα κείμενα που είχαν ήδη υποστεί την απαιτούμενη επεξεργασία.

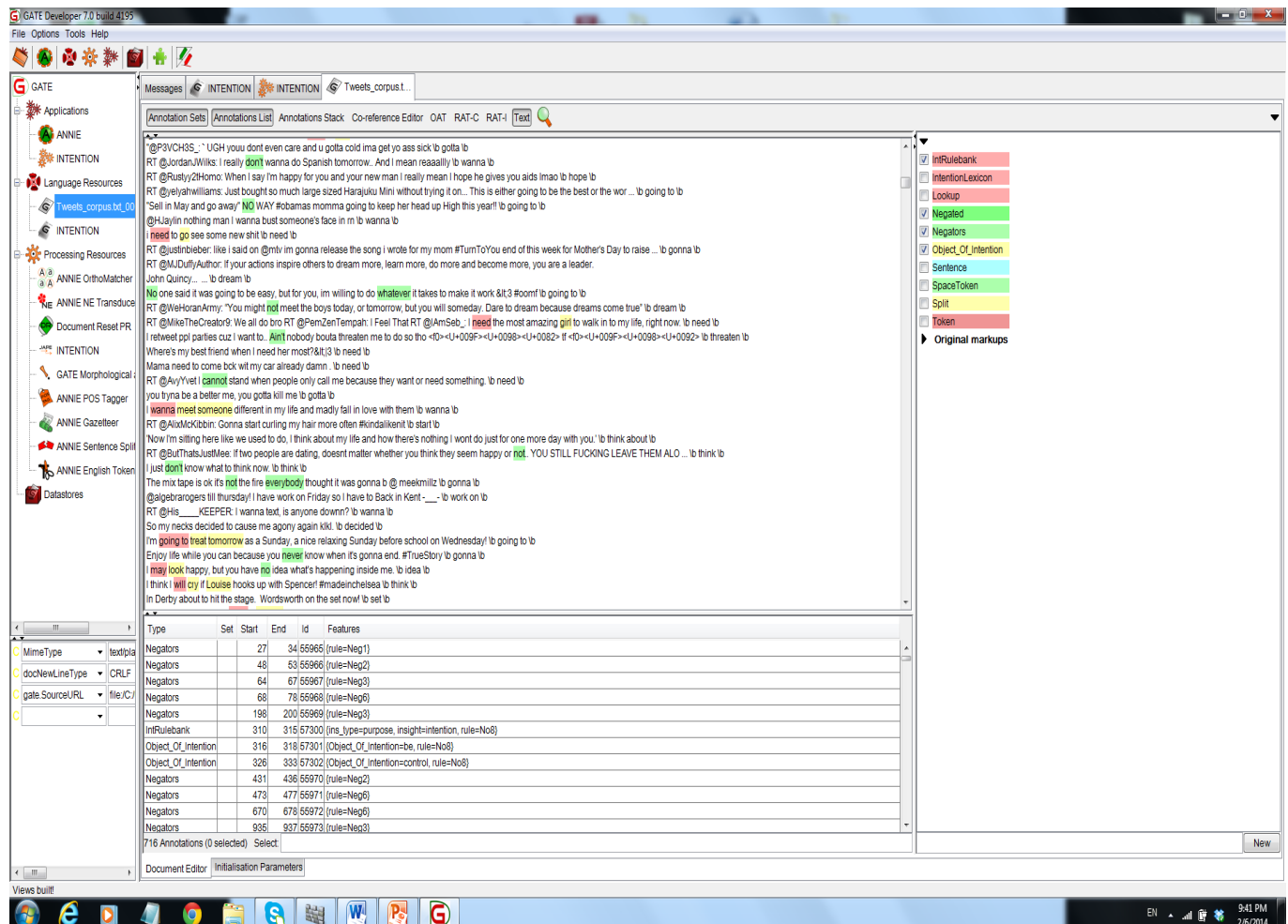
1. <http://gate.ac.uk/gatewiki/cow/doc/gatewiki.html>

Η **main.jape**(**βασική συνάρτηση**) καλεί κάθε φορά όλους τους κανόνες που έχουν γραφτεί σε μορφή jape και αυτό αποτελεί την γραμματική-μηχανή αυτόματης εξόρυξης του **INTENTION(πρόθεσης)** που πιθανόν να υπάρχει μέσα στα κείμενα και γυρίζει επισημειωμένα κάποια αποτελέσματα.

```
/*
* main.jape
*/
```

MultiPhase: IntentionExtraction

- Phases:
- Negators
- IntentionVB2
- IntentionJJ2
- IntentionEXP2
- IntentionRules



ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ

ΠΕΡΙΓΡΑΦΗ ΣΩΜΑΤΩΝ ΚΕΙΜΕΝΩΝ

4.1 Development corpora

Έγινε η επιλογή από τα ήδη υπάρχοντα λεξικά τα οποία αναφέρθηκαν σε προηγούμενο κεφάλαιο των λημμάτων τα οποία αποτελούν το Intention Lexicon και χρησιμοποιήθηκαν ως λέξεις κλειδιά(key words) για την άντληση παραδειγμάτων προς έρευνα από διάφορα open sources(tweets from twitter) και από το internet(blogs, sites etc.). Τα παραδείγματα αυτά χρησιμοποιήθηκαν για την ορθότερη κατηγοριοποίηση των λημμάτων στους διάφορους τύπους intention και στην δημιουργία patterns πάνω σε απτά παραδείγματα έχοντας ως κατάληξη την δημιουργία της Intention Grammar.

Από το twitter δημιουργήθηκαν σταδιακά corpora παραδειγμάτων τα οποία χρησιμοποιήθηκαν στην πορεία της έρευνας. Ως λέξεις κλειδιά τέθηκαν τα λήμματα τα οποία αποτελούν το Intention Lexicon.

CORPORA	SOURCE	KEYWORD	NUMBER OF SENTENCES
intentions	Twitter+online lexica	Λήμματα Intention lexicon	205
Intention_corpora v 0.1	Twitter+online lexica	Λήμματα Intention lexicon	450
Tweet_corpus	Twitter+online lexica	Λήμματα Intention lexicon	1085

Δημιουργήθηκαν στην αρχή και μικρότερα επιμέρους corpora τα οποία συγχωνεύθηκαν σε μεγαλύτερα. Επιπλέον συγκεντρώθηκε από blogs και sites ένας αριθμός άρθρων και κειμένων τα οποία συνέβαλαν και αυτά σε περαιτέρω διαπιστώσεις και βελτιώσεις της Intention Grammar.

CORPORA	SOURCE	KEYWORD	NUMBER OF ARTICLES
blogtexts	Blogs and sites	Λήμματα Intention lexicon	21
FOXNEWS	FOXNEWS	Λήμματα Intention lexicon	24
WEB	Blogs and sites	Λήμματα Intention lexicon	30

4.2 Evaluation corpora

Δημιουργήθηκε ένα σώμα από 350 προτάσεις από πέντε διαφορετικούς τομείς στις οποίες έτρεξε η γραμματική που δημιουργήθηκε. Στα αποτελέσματα της γραμματικής θα βασιστούν και οι παρατηρήσεις που θα αναλυθούν αναλυτικότερα στο κεφάλαιο των συμπερασμάτων και των μελλοντικών εξελίξεων της εργασίας.

Οι τομείς από τους οποίους αντλήθηκαν αυτά τα δεδομένα είναι οι:

- Product_Reviews
- Politics
- Web
- Blogs
- Social Media

Σε επόμενο στάδιο έγινε μια προσπάθεια να πραγματοποιηθεί στο Gate χειροκίνητη επισημείωση με βάση δοσμένο Annotation Schema:

Intention Annotation Schema

```
<?xml version="1.0"?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
  <!-- XSchema definition for Intention-->
  <element name="Intention">
    <complexType>
      <attribute name="ins_type" use="required">
        <simpleType>
          <restriction base="string">
            <enumeration value="desire"/>
            <enumeration value="commitment"/>
            <enumeration value="purpose"/>
          </restriction>
        </simpleType>
      </attribute>
    </complexType>
  </element>
</schema>
```

```

</attribute>
  <attribute name="Object_Of_Intention" use="required">
    <simpleType>
      <restriction base="string">
        <enumeration value="Object_Of_Intention"/>
      </restriction>
    </simpleType>
  </attribute>
  <attribute name="Negated" use="required">
    <simpleType>
      <restriction base="string">
        <enumeration value="Negated"/>
      </restriction>
    </simpleType>
  </attribute>
  <attribute name="Annotator" use="required" default="thanassis">
    <simpleType>
      <restriction base="string">
        <enumeration value="other"/>
        <enumeration value="thanassis"/>
      </restriction>
    </simpleType>
  </attribute>
</complexType>

```

</element>

</schema>

Με βάση όλα αυτά θα γίνει η αξιολόγηση του Intention Extraction σχετικά με την ακρίβεια των αποτελεσμάτων που φέρνει η Intention Grammar πάνω στα κείμενα στα οποία τρέχει ώστε να προσδιοριστούν και τα σφάλματα τα οποία σε μελλοντική δουλειά θα αποτελέσουν στόχοι για περαιτέρω βελτίωση του αρχικού προγράμματος.

ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ

ΣΥΜΠΕΡΑΣΜΑΤΑ-ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ

5.1 ΣΥΜΠΕΡΑΣΜΑΤΑ

Όπως αναφέρθηκε στην εισαγωγή της εργασίας συνοπτικά αλλά και εκτενέστερα στο τρίτο κεφάλαιο όπου περιγράφεται αναλυτικά η μεθοδολογία που υιοθετήθηκε σκοπός της προσπάθειας αυτής είναι η αυτόματη εξαγωγή της πρόθεσης και κυρίως της *direct intention* από κείμενα, η κατηγοριοποίησή της στα διάφορα είδη της πρόθεσης όπως ορίστηκαν από την μελέτη και ο εντοπισμός του αντικειμένου της πρόθεσης (*object of intention*).

Για αυτό το λόγο δημιουργήθηκαν τα λεξικά και η γραμματική η οποία έτρεξε μέσω της πλατφόρμας του *gate* πάνω στα δεδομένα. Με βάση την παρατήρηση των αποτελεσμάτων των δεδομένων υπάρχει η δυνατότητα να εξαχθούν ορισμένα συμπεράσματα σχετικά με την αποτελεσματικότητα της γραμματικής αλλά και τις αδυναμίες της που θα αποτελέσουν τη βάση για μελλοντική βελτίωση της εργασίας. Τέλος θα γίνει μία σύγκριση με το αποτέλεσμα της *aiaioo.com* στα ίδια δεδομένα.

Πιο συγκεκριμένα:

- Στον εντοπισμό των ρημάτων, επιθέτων και εκφράσεων που εκφράζουν πρόθεση και τα οποία έχουν ενταχθεί στο λεξικό και σε ανάλογη κατηγορία πρόθεσης υπάρχει αρκετό ποσοστό επιτυχίας καθώς επισημειώνονται επάνω στο *output* της πλατφόρμας.
 - Σε μερικές περιπτώσεις όμως δεν ενεργοποιούνται οι κανόνες καθώς ή ο *tagger* δεν έχει ως κατηγορία *VB* το ρήμα αλλά ως *NN* όπως στην περίπτωση του ρήματος *need* στο οποίο παρατηρούνται οι πιο πολλές αποκλίσεις.
 - Επίσης λείπουν οι προσωπικές αντωνυμίες από την αρχή της πρότασης που δηλώνουν το πρώτο πρόσωπο ενικό και πληθυντικό άμεσο στοιχείο για την δήλωση της άμεσης πρόθεσης (*direct intention*) δηλαδή οι αντωνυμίες *I* και *We*.
 - Κάποιες φορές δεν ενεργοποιούνται οι κανόνες λόγω του ότι κάποια ρήματα, επίθετα ή εκφράσεις ανήκουν σε παραπάνω από μία συντακτικές κατηγορίες στο λεξικό και δημιουργείται πρόβλημα με την προτεραιότητα (*priority*) των κανόνων δηλαδή με την σειρά που ενεργοποιούνται. Αν ένα ρήμα, επίθετο ή έκφραση

ανήκει σε παραπάνω από μία κατηγορίες συντακτικές στη γραμματική ενεργοποιείται ο κανόνας ο πρώτος που συναντιέται χωρίς να υπάρχει η δυνατότητα να μπει το πρόγραμμα και στους υπόλοιπους κανόνες και έτσι χάνονται προτάσεις που θα μπορούσαν να έχουν αναγνωριστεί καθώς η δομή τους είχε ήδη προβλεφθεί στη γραμματική.

- Στις φράσεις του τύπου I wanna buy a phone, πολλές φορές το wanna το βλέπει ως ρήμα και όχι ως βοηθητικό και το ρήμα που έχει μετά το πιάνει ως NN (ουσιαστικό) πράγμα το οποίο έρχεται σε ασυμφωνία με τις προβλεφθείσες δομές με συνέπεια να μην ανιχνεύεται από την γραμματική.
- Ορθογραφικά λάθη και συντομογραφίες του τύπου Im αντί για I'm δημιουργούν επιπρόσθετα προβλήματα.
- Περιπτώσεις μη ενεργοποίησης κανόνων λόγω απουσίας καταχώρησης ρημάτων, επιθέτων ή εκφράσεων δεν υπήρχαν. Σε περίπτωση χρήσης σε μεγαλύτερο όγκο δεδομένων από διαφορετικές θεματικές περιοχές υπάρχει ανάγκη εμπλουτισμού σε μελλοντική βελτίωση.
- Κάποια προβλήματα υπάρχουν στο φαινόμενο της άρνησης(negation) καθώς ο tagger αντιμετωπίζει δυσκολία αναγνώρισης του **cannot** αλλά και μερικών εκφράσεων του τύπου I have no idea παρόλο που έχουν προβλεφθεί οι ανάλογες δομές και στα λεξικά και στη γραμματική.
- Τέλος το μεγαλύτερο πρόβλημα είναι η χαμηλή ακρίβεια στην πρόβλεψη του αντικειμένου της πρόθεσης(object of intention). Τα phrasal verbs(break up etc.) αλλά και η ύπαρξη συνθετότερης συντακτικής δομής πρότασης με πολλούς παρεμβαλλόμενους όρους ανάμεσα στο κατηγορήμα της πρότασης και στο αντικείμενο δημιουργούν την ανάγκη για βαθύτερη συντακτική ανάλυση όπως και στο φαινόμενο της άρνησης και ενασχόληση με το dependency parsing.

5.2 ΣΥΓΚΡΙΣΗ ΜΕ ΑΙΑΙΟΟ¹

Σε αυτό το στάδιο γίνεται μία σύγκριση με μία πλατφόρμα Intention Analysis της Aiaioo.com. Οι ερευνητές της aiaioo¹ προτείνουν ένα μοντέλο ανάλυσης πρόθεσης, βασιζόμενοι και αυτοί στις θεωρίες του Austin και του Searle καθώς και στη θεωρία του discourse analysis, όπου υπάρχουν 10(δέκα) κατηγορίες πρόθεσης: wish, purchase, inquire, compare, praise, criticize, complain, quit, sell, other. Είναι σαφές ότι η κατηγοριοποίηση που προτείνεται από την aiaioo και την κατηγοριοποίηση της παρούσας εργασίας διαφέρει στην προσέγγιση στο θεματικό πυρήνα του Intention analysis. Συγκρίνοντας όμως το output της Intention Grammar και της πλατφόρμας ανάλυσης της aiaioo μπορούν να εξαχθούν κάποια χρήσιμα συμπεράσματα. Οι περιπτώσεις στις οποίες η γραμματική που προτείνεται έχει συμφωνία στο output με το αντίστοιχο της aiaioo και στους δύο άξονες σύγκρισης που θέσαμε εξ αρχής δηλαδή το είδος του intention και το object of intention είναι πολύ λίγες. Ο λόγος που συμβαίνει αυτό είναι στην διαφορετική αντίληψη του τι θεωρείται intention και ποια κατηγορήματα φέρουν εννοιολογικά το σημασιολογικό φορτίο της πρόθεσης. Σε πολλές περιπτώσεις ενώ η γραμματική που προτείνεται αναγνωρίζει την πρόθεση σύμφωνα με την κατηγοριοποίηση που προτάθηκε και αναλύθηκε στο τρίτο κεφάλαιο της εργασίας στο αντίστοιχο output της aiaioo είτε κατηγοριοποιείται σε άλλο είδος πρόθεσης είτε δεν αναγνωρίζεται καθόλου ως πρόθεση. Τέλος στο επίπεδο του object η γραμματική που παρουσιάζεται στην εργασία σε αυτό το επίπεδο αναγνωρίζει συγκεκριμένες δομές και όχι phrasal verbs ή ολόκληρες προτάσεις όπως της aiaioo λόγω του ότι οι κανόνες δεν έχουν εμβαθύνει τόσο στο επίπεδο του dependency parsing πράγμα το οποίο αναφέρεται ως μελλοντική δουλειά.

1. <http://www.aiaioo.com/>

5.3 ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ

Παρόλη την εποικοδομητική προσπάθεια και έρευνα που διενεργήθηκε καθ' όλη την διάρκεια εκπόνησης της διπλωματικής εργασίας υπάρχουν αρκετά ζητήματα που δεν έχουν επιλυθεί ακόμα στον συγκεκριμένο χώρο της επεξεργασίας φυσικής γλώσσας(Natural Language Processing) και τα οποία σε μελλοντική επέκταση και βελτίωση της συγκεκριμένης διπλωματικής, ίσως σε διδακτορικό επίπεδο, θα γίνει η προσπάθεια αν όχι να επιλυθούν οριστικά να διενεργηθεί μια ενδελεχέστερη μελέτη ώστε να υπάρξει αν καταστεί δυνατόν ένα ποσοστό προόδου.

Αυτά τα θέματα μπορούν να συνοψιστούν στα εξής:

- ❖ Έντονο είναι το ενδιαφέρον και για τον έμμεσο τρόπο έκφρασης της πρόθεσης(**indirect intention**) μέσω της χρήσης υποθετικού λόγου καθώς και η επίτευξη μεγαλύτερων ποσοστών επιτυχίας στην πρόβλεψη της πρόθεσης σε προτασιακό επίπεδο. Π.χ. If only I could buy a new home.
- ❖ Ενδιαφέρουσα πρόκληση επίσης είναι η μελέτη των περιπτώσεων όπου το intention(πρόθεση) γίνεται request όπου η μελλοντική πρόθεση του ατόμου να κάνει κάτι προϋποθέτει κάτι άλλο ώστε η πρόθεση του ατόμου να αποτελεί αποτέλεσμα αυτής της προϋπόθεσης.
- ❖ Ένα άλλο θέμα είναι η βελτίωση του precision στη ανίχνευση του object of intention(αντικείμενο της πρόθεσης), που σημαίνει βελτίωση της εργασίας στο επίπεδο του dependency parsing για το αντικείμενο της πρόθεσης . Για παράδειγμα ένα πρόβλημα είναι ότι δεν πιάνει τα phrasal verbs ως αντικείμενο όλη την φράση αλλά μόνο την πρώτη λέξη. Π. χ I want to break up. Το πρόγραμμα σταματάει στο break και δεν πιάνει και το up όπως θα ήταν το ορθό. Ίσως μία πιθανή λύση σε αυτό το τεχνικό πρόβλημα που δεν κατέστη δυνατό να επιλυθεί με τις παρούσες γνώσεις κατασκευής κανόνων jape να ήταν η δημιουργία λιστών με τα phrasal verbs τα οποία να τρέχουν παράλληλα με το κυρίως πρόγραμμα.
- ❖ Τα λεγόμενα «θορυβώδη κείμενα»(noisy text) τα οποία είναι συχνά εμφανιζόμενα σε ιστοτόπους κοινωνικής δικτύωσης οι οποίοι χρησιμοποιούνται κυρίως από νέους σε ηλικία ανθρώπους τα οποία

διαθέτουν ορθογραφικά λάθη, απουσία σημείων στίξης ή προβληματική στίξη, συντακτικά λάθη, τα οποία χρησιμοποιούν αργκό ή ιδιάζοντες τρόπους έκφρασης αποτελούν πάντα πρόκληση για τους ερευνητές που αναπτύσσουν ανάλογα συστήματα.

Μέσα λοιπόν σε μια περιοχή που συνεχώς αναπτύσσεται και βελτιώνεται οι μελλοντικοί στόχοι θα είναι οι εξής:

- Χρήση μεγαλύτερου όγκου δεδομένων και στροφή προς τα big data προκειμένου να υπάρξει αξιοσημείωτη βελτίωση και σαφέστερα αποτελέσματα σε δεδομένα από διαφορετικές θεματικές περιοχές.
- Υιοθέτηση τεχνικών μηχανικής μάθησης(machine learning) πάντα σε συνδυασμό με τον ανθρώπινο παράγοντα και την ανθρώπινη επίβλεψη και αξιολόγηση για αύξηση του recall .
- Κατασκευή λεξικών πόρων και σωμάτων κειμένων με μηχανές αναζήτησης οι οποίες θα μπορέσουν να εμπλουτίσουν με περισσότερα παραδείγματα τα ήδη υπάρχοντα.
- Επέκταση σε διδακτορικό επίπεδο της εργασίας στο χώρο της πρόληψης αυτοκτονιών και γενικά σε θέματα που άπτονται του ηλεκτρονικού εγκλήματος ή για την κατασκευή εμπορικών εφαρμογών για αξιολόγηση προϊόντων ώστε πλέον η δουλειά αυτή να έχει και πρακτική εφαρμογή φέρνοντας απτά και χρήσιμα αποτελέσματα στους χρήστες πράγμα που αποτελεί και προσωπική μου επιθυμία.

ΚΕΦΑΛΑΙΟ ΕΚΤΟ

ΒΙΒΛΙΟΓΡΑΦΙΑ

Andrew B. Goldberg, Nathanael Fillmore, David Andrzejewski, Zhiting Xu, Bryan Gibson and Xiaojin Zhu.

May All Your Wishes Come True: A Study of Wishes and How to Recognize Them.

Annual Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2009).

Brown, P. & S. C. Levinson, (1987), *Politeness: Some Universals In Language Usage*, Cambridge: Cambridge University Press

Carlos, C. S. , *Intention Analysis- Business Use Case*, Aiaioo Labs, Bangalore India

Carlos, C. S. *Whitepaper on a 360 Degree Strategy for Text Analysis*, Aiaioo Labs, Benson Town, Bangalore, India

Charles Manski, 1989. The use of intentions data to predict behavior: A best case analysis, Department of Economics and Institute for Research on Poverty, University of Wisconsin-Madison, p.1-27 .

Dontcheva – Navratilova, O. (2005), *Politeness Strategies in Institutional Speech Acts in Discourse and Interaction* 1. Brno Seminar on Linguistic Studies in English: Proceedings

Douglas P. Twitchell and Jay F. Nunamaker Jr. (2004). Speech act profiling: a probabilistic method for analyzing persistent conversations and their participants. In System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on, pages 1–10, January.

Douglas P. Twitchell, Mark Adkins, Jay F. Nunamaker Jr., and Judee K. Burgoon. 2004. Using speech act theory to model conversations for automated classification and retrieval. In Proceedings of the International Working Conference Language Action Perspective CommunicationModelling (LAP 2004), pages 121–130.

Etsuko O. (2006), Austin(1962)'s Speech Act Theory and the Speech Situation pages 1-14

Grice, H. P. (1996). Meaning. In A. P. Martinich (Ed.), *The philosophy of language* (3rd ed.) (pp.85-91). New York: Oxford University Press.

Jade Goldstein and Roberta Evans Sabin. 2006. Using speech acts to categorize email and identify email genres. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences - Volume 03*, pages 50.2–, Washington, DC, USA. IEEE Computer Society. analyzing persistent conversations and their participants. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 1–10, January.

Jawad, F. A., *A pragmatic Analysis of Illocutionary Speech Acts in Standard Arabic with a Special Reference to Al- Ashet' s "Epistole"*, Babylon University/ College of Education/ Safi yil Deen Alhilli

Kasper, G (2004) Speech acts in (inter) action: Repeated questions, *Intercultural Pragmatics* 1:1, 125-133.

Leech, G. N. (1983), *Principles of Pragmatics*, London: Longman

Qadir A. and Riloff E. (2011), *Classifying Sentences as Speech Acts in Message Board Posts*, In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP-2011)* pages 1-11.

Robert M. Harnish (2005), COMMITMENTS AND SPEECH ACTS, *Philosophica* 75 (2005) pp. 11-41

Searle, J. R. (1969), *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press

Searle, J. R. (1976), A classification of illocutionary acts: *Language in Society* 5, 1-23

Searle, J. R. (1991), *Meaning, intentionality, and speech acts*, in E. Lepore & R. van Gulick (επιμ.), *John Searle and His Critics*, Oxford: Basic Blackwell, 81-102

Strawson, P. F. (1969). Intention and Convention in Speech Acts. In K. T. Fann (Ed.),

Symposium on J. L. Austin(1962) (pp. 380-400). New York: Humanities Press

Thomas, J (1995), *Meaning in interaction: An Introduction to Pragmatics*, London: Longman

Vitor R. Carvalho and William W. Cohen. (2005). *On the collective classification of email "speech acts"*. In SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pages 345–352, New York, NY, USA. ACM Press.

Vitor R. Carvalho and William W. Cohen. (2006). *Improving "email speech acts" analysis via n-gram selection*. In Proceedings of the HLT-NAACL 2006 Workshop on Analyzing Conversations in Text and Speech, ACTS '09, pages 35–41, Stroudsburg, PA, USA. Association for Computational Linguistics.

Wierzbicka, A., (1987). *English speech act verbs: a semantic dictionary*. Academic Press, Sydney, Orlando

William W. Cohen, Vitor R. Carvalho, and Tom M. Mitchell. (2004), *Learning to classify email into "speech acts"*. In EMNLP, pages 309–316. ACL.

Yoshitake Masaki (2004), *Critique of J.L. Austin(1962)'s Speech Act Theory: Decentralization of the Speaker-Centered Meaning in Communication*, Communication Association of Japan Kyushu Chapter, Kyushu Communication Studies, 2004. pp2:27-43

Κανάκης, Κ. (2007), *Εισαγωγή στην πραγματολογία: Γνωστικές και κοινωνικές όψεις της γλωσσικής χρήσης*, Αθήνα: Εκδόσεις του Εικοστού Πρώτου

