

“Συντακτικός αναλυτής ελληνικής γλώσσας και εφαρμογές ανάληψης πληροφορίας”

Σπάχος Πέτρος



Πολυτεχνείο Κρήτης

Τμήμα Ηλεκτρονικών Μηχανικών & Μηχανικών Υπολογιστών

Εξεταστική Επιτροπή

Καθηγητής Σ. Χριστοδουλάκης (*Επιβλέπων*)

Αναπ.Καθηγητής Α.Ποταμιάνος

Αναπ.Καθηγητής Ε. Πετράκης

Χανιά, Ιούλιος 2008

Αφιέρωση

*Στους γονείς μου,
Κλεομένη και Ελευθερία
και στην αγαπημένη μου αδερφή Δήμητρα,
που είναι πάντα δίπλα μου
και με στηρίζουν*

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις ιδιαίτερες ευχαριστίες μου στον επιβλέποντα καθηγητή μου κ. Σταύρο Χριστοδουλάκη για την επίβλεψη και καθοδήγησή του στην εκπόνηση αυτής της διπλωματικής εργασίας, καθώς και για τις σημαντικές εμπειρίες που μου προσέφερε κατά την διάρκεια της εργασίας μου στο Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων και Εφαρμογών.

Θα ήθελα επίσης να ευχαριστήσω τους καθηγητές κκ. Ποταμιάνο Αλέξανδρο και Πετράκη Ευριπίδη για το χρόνο που θα αφιερώσουν στην ανάγνωση του κειμένου και για τις εποικοδομητικές παρατηρήσεις τους.

Επίσης, οφείλω ένα ιδιαίτερο ευχαριστώ στη Φώτη Καζάση, για την συνεργασία και την αρωγή που μου προσέφερε στην υλοποίηση της εργασίας αυτής καθώς και τις συμβουλές του σε τεχνικά θέματα.

Ακόμη θα ήθελα να ευχαριστήσω ιδιαίτερα τον Κοτόπουλο Γιάννη που με βοήθησε, με συμβούλευσε και με στήριξε καθόλη τη διάρκεια τη εργασίας μου.

Ιδιαίτερες ευχαριστίες οφείλω στον καθηγητή κ.Ζερβάκη Μιχάλη για την αρωγή του στην υλοποίηση των μελλοντικών σχεδίων μου αλλά και στην Ζερβάνου Κέλλυ για την πολύτιμη αρωγή της σε τεχνικά θέματα.

Ένα μεγάλο ευχαριστώ σε όλα τα μέλη του εργαστηρίου και ιδιαίτερα στους Μουμουτζη Νεκτάριο και Μυλωνάκη Μανώλη για τη βοήθεια και συμπαράσταση που μου έδειξαν όλο το καιρό που ήμουνα στο εργαστήριο.

Κλείνοντας, θα ήθελα να ευχαριστήσω τους φίλους μου ,Δημήτρη και Τάσο ,που ήταν πάντα δίπλα μου στις επιλογές μου αλλά και που με άντεξαν όλα αυτά τα χρόνια (και ήταν πολλά...) και έκαναν την παραμονή μου στα Χανιά ευχάριστη και ξεχωριστή. Ευχαριστώ επίσης το Σύνδεσμο Διαιτητών Χανίων ο οποίος με γέμισε εμπειρίες.Τέλος θα ήθελα να ευχαριστήσω όλα τα μπαρ των Χανίων που έδιναν κατα καιρούς διέξοδο στα αδιέξοδά μου.

Περίληψη

Η φυσική γλώσσα χρησιμοποιείται ως μέσο επικοινωνίας μεταξύ ανθρώπων. Λέγεται έτσι για να τη διαχωρίσουμε από την τεχνητή γλώσσα η οποία χρησιμοποιείται για επικοινωνία ανθρώπου με υπολογιστή και υπολογιστή με υπολογιστή. Τα τελευταία χρόνια γίνονται προσπάθειες ώστε η φυσική γλώσσα να μπορεί να αναγνωρισθεί και να αναλυθεί από τον υπολογιστή. Για να γίνει κάτι τέτοιο, πρέπει να κατασκευάσουμε εφαρμογές που είναι προσαρμοσμένες σε φυσικές γλώσσες. Αυτό σημαίνει πως αυτές οι εφαρμογές θα απευθύνονται σε συγκεκριμένες γλώσσες έτσι ώστε να λαμβάνουν υπόψη τους τις ιδιαιτερότητες κάθε γλώσσας. Η ελληνική γλώσσα είναι ιδιαίτερα πλούσια τόσο συντακτικά όσο και σημασιολογικά.

Στα πλαίσια της παρούσας διπλωματικής εργασίας παρουσιάζεται ο σχεδιασμός του μοντέλου και η υλοποίηση της αρχιτεκτονικής ενός συντακτικού αναλυτή για τη Νέα ελληνική γλώσσα. Επίσης παρουσιάστηκε μία γραφική διεπαφή η οποία πέρα από συντακτική ανάλυση, έχει τη δυνατότητα να χρησιμοποιηθεί και ως μηχανή αναζήτησης με συντακτικά κριτήρια. Παρουσιάζεται επίσης η αρχιτεκτονική που πρέπει να ακολουθηθεί προκειμένου να ληφθούν υπόψη οι ιδιαιτερότητες της γλώσσας και να γίνει σωστά η συντακτική ανάλυση.

Τα τμήματα λογισμικού του συντακτικού αναλυτή εξετάζουν ομοιόμορφα μια σειρά προβλημάτων στην ανάλυση της πρότασης τα οποία απαιτούν χωριστούς μηχανισμούς. Μια ενιαία αρχιτεκτονική χειρίζεται τη συντακτική και σημασιολογική ανάλυση, αλλά και τις ασάφειες. Παράλληλα, το πλαίσιο έχει κατασκευαστεί με τέτοιο τρόπο έτσι ώστε να αποφεύγονται οι εξαρτήσεις με συγκεκριμένες κατηγορίες ερωτήσεων ή κείμενα συγκεκριμένων γνωστικών πεδίων με αποτέλεσμα να είναι επαναχρησιμοποιήσιμο σε διαφορετικές εφαρμογές με διαφορετική σημασιολογία περιοχών. Υλοποιείται επίσης μια γραφική διεπαφή αναζήτησης ερωτήσεων αλλά και συντακτικής ανάλυσης κειμένων.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΕΧΟΜΕΝΑ.....	V
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ.....	VII
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ.....	VIII
ΚΕΦΑΛΑΙΟ 1.....	1
1.1 ΕΙΣΑΓΩΓΗ – ΑΝΑΓΚΑΙΟΤΗΤΑ.....	1
1.2 ΓΛΩΣΣΑ ΚΑΙ ΑΣΑΦΕΙΑ.....	3
1.3 ΣΤΟΧΟΙ ΔΙΠΛΩΜΑΤΙΚΗΣ.....	5
1.4 ΣΥΝΕΙΣΦΟΡΑ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ.....	6
1.5 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ.....	7
ΚΕΦΑΛΑΙΟ 2.....	9
2.1 ΦΥΣΙΚΗ ΓΛΩΣΣΑ.....	9
2.1.1 Προκαταρκτικά.....	9
2.1.2 Συντακτική ανάλυση και ασάφεια.....	10
2.1.3 Σημασιολογική αποσαφήνιση.....	11
2.2 ΣΥΝΤΑΚΤΙΚΟ ΕΛΛΗΝΙΚΗΣ ΓΛΩΣΣΑΣ.....	12
2.2.1 Πρόταση.....	13
2.2.2 Όροι πρότασης.....	13
2.2.2.1 Υποκείμενο.....	14
2.2.2.2 Κατηγορήμα.....	15
2.2.3 Είδη προτάσεων.....	15
2.2.3.1 Περιεχόμενο.....	16
2.2.3.2 Ποιότητα.....	17
2.2.3.3 Σχέση τους με τις άλλες.....	18
2.2.3.4 Ως προς τη δομή τους.....	18
2.2.4 Σύνδεση προτάσεων.....	20
2.2.5 Σχήματα λόγου.....	24
2.2.6 Σύνολα λέξεων.....	26
2.2.6.1 Ονοματικό σύνολο.....	27
2.2.6.2 Ρηματικό σύνολο.....	30
2.2.7 Γλωσσικό μοντέλο.....	31
2.3 ΚΑΝΟΝΙΚΕΣ ΕΚΦΡΑΣΕΙΣ.....	33
2.3.1 Ορισμός κανονικών εκφράσεων.....	33
2.3.2 Η γλώσσα Tcl.....	34
2.4 ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΗ ΑΝΑΛΥΣΗ.....	35
2.4.1 Γενικά.....	35
2.4.2 Μορφοσυντακτική ανάλυση.....	36
2.4.3 Μορφοσυντακτική αποσαφήνιση στη Νέα Ελληνική.....	38
2.4.4 Μορφοσυντακτικός σχολιαστής.....	41
ΚΕΦΑΛΑΙΟ 3.....	43
3.1 ΤΟ ΜΟΝΤΕΛΟ ΤΟΥ ΣΥΝΤΑΚΤΙΚΟΥ ΑΝΑΛΥΤΗ.....	44
3.2 ΔΟΜΙΚΕΣ ΜΟΝΑΔΕΣ.....	45
3.2.1 Λεκτικές μονάδες.....	47
3.2.2 Σημεία στίξης.....	47
3.2.3 Δεσμευμένες λέξεις.....	49
3.2.4 Συντακτικοί κανόνες.....	50
3.2.5 Συντακτικοί όροι.....	55
3.2.6 Σύνδεσμοι.....	56
3.3 ΤΟ ΡΗΜΑ.....	58

3.3.1	<i>Εύρεση προτάσεων</i>	58
3.3.2	<i>Προσδιορισμός υποκειμένων</i>	61
3.3.3	<i>Προσδιορισμός αντικειμένων</i>	62
3.4	Ο ΟΡΟΣ ΥΠΟΚΕΙΜΕΝΟ	63
3.4.1	<i>Κανόνες εύρεσης</i>	64
3.4.2	<i>Σύνδεση υποκειμένων</i>	66
3.5	Ο ΟΡΟΣ ΑΝΤΙΚΕΙΜΕΝΟ	67
3.5.1	<i>Κατηγορίες αντικειμένων</i>	68
3.5.1.1	<i>Απλό αντικείμενο</i>	69
3.5.1.2	<i>Άμεσο και έμμεσο αντικείμενο</i>	70
3.5.1.3	<i>Κατηγορούμενο</i>	72
3.5.2	<i>Σύνδεση όρων αντικείμενα</i>	74
3.6	ΟΝΟΜΑΤΙΚΟΙ ΠΡΟΣΔΙΟΡΙΣΜΟΙ	75
3.6.1	<i>Ουσιαστικό με άρθρο</i>	76
3.6.2	<i>Ουσιαστικό και αντωνυμία</i>	78
3.6.3	<i>Ουσιαστικό και ουσιαστικό</i>	78
3.6.3.1	<i>Ομοιόπρωτος προσδιορισμός</i>	79
3.6.3.2	<i>Ετερόπρωτος προσδιορισμός</i>	80
3.6.4	<i>Επίθετο και ουσιαστικό</i>	80
3.6.4.1	<i>Επιθετικός προσδιορισμός</i>	81
ΚΕΦΑΛΑΙΟ 4		83
4.1	ΕΙΣΑΓΩΓΗ	83
4.1	ΓΕΝΙΚΗ ΑΡΧΙΤΕΚΤΟΝΙΚΗ	84
4.2	Η ΜΟΝΑΔΑ ΤΟΥ ΜΟΡΦΟΣΥΝΤΑΚΤΙΚΟΥ ΣΧΟΛΙΑΣΤΗ	85
4.3	Η ΜΟΝΑΔΑ ΣΥΝΔΕΣΗΣ – ΑΠΟΜΟΝΩΣΗΣ	88
4.3.1	<i>Η υπομονάδα αφαίρεσης σημείων στίξης (stopword)</i>	89
4.3.2	<i>Η υπομονάδα εύρεσης και διαχωρισμού προτάσεων (sentence splitter)</i>	90
4.4	Η ΜΟΝΑΔΑ ΤΟΥ ΣΥΝΤΑΚΤΙΚΟΥ ΣΧΟΛΙΑΣΤΗ	91
4.4.1	<i>Η υπομονάδα γραμματικής αποσαφήνισης (clarifications)</i>	92
4.4.2	<i>Η υπομονάδα συντακτικής αποσαφήνισης (syntactic anaysis)</i>	94
4.4.3	<i>Η υπομονάδα κανόνων και κανονικών εκφράσεων (rules and regular expressions)</i>	95
4.5	ΟΙ ΠΡΟΤΑΣΕΙΣ ΠΟΥ ΥΠΟΣΤΗΡΙΖΟΝΤΑΙ	97
4.6	SYSTEM FLOW	99
ΚΕΦΑΛΑΙΟ 5		106
5.1	ΕΙΣΑΓΩΓΗ	106
5.2	ΣΕΝΑΡΙΑ ΧΡΗΣΗΣ	106
5.3	ΛΕΙΤΟΥΡΓΙΚΟΤΗΤΑ	109
5.4	ΠΑΡΑΔΕΙΓΜΑΤΑ ΧΡΗΣΗΣ	110
ΚΕΦΑΛΑΙΟ 6		115
6.1	ΑΞΙΟΛΟΓΗΣΗ ΟΡΘΟΤΗΤΑΣ ΚΑΙ ΚΑΤΑΛΛΗΛΟΤΗΤΑΣ	115
6.2	ΣΥΜΠΕΡΑΣΜΑΤΑ	118
ΚΕΦΑΛΑΙΟ 7		120
7.1	ΑΝΑΚΕΦΑΛΑΙΩΣΗ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ	120
7.2	ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ	121

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Εικόνα 1: Τα είδη των προτάσεων.....	20
Εικόνα 2: Η σύνδεση των προτάσεων	24
Εικόνα 3: Οι κύριοι όροι απλής πρότασης.....	25
Εικόνα 4: Οι κύριοι όροι απλής πρότασης(2).....	26
Εικόνα 5: Τα σύνολα των λέξεων.....	31
Εικόνα 6: Το γλωσσικό μοντέλο που προκύπτει από τη συντακτική ανάλυση.....	32
Εικόνα 7: Αρχιτεκτονική του μορφοσυντακτικού σχολιαστή.....	41
Εικόνα 8: Το μοντέλο του συντακτικού αναλυτή.....	44
Εικόνα 9: Οι δομικές μονάδες του συντακτικού μας αναλυτή.....	46
Εικόνα 10: Οι περιπτώσεις του αντικειμένου.....	71
Εικόνα 11: Η αρχιτεκτονική του συντακτικού αναλυτή.....	84
Εικόνα 12: Η μονάδα του μορφοσυντακτικού αναλυτή.....	86
Εικόνα 13: Η μονάδα σύνδεσης – απομόνωσης.....	89
Εικόνα 14: Ο συντακτικός σχολιαστής.....	92
Εικόνα 15: Το αρχικό κείμενο πριν την εισαγωγή στον αναλυτή.....	99
Εικόνα 16: Το xml κείμενο όπως προκύπτει από το μορφοσυντακτικό αναλυτή.....	101
Εικόνα 17: Το xml κείμενο έπειτα από τη μονάδα σύνδεσης – απομόνωσης.....	103
Εικόνα 18: Η συντακτική ανάλυση του κειμένου.....	105
Εικόνα 19: Screen shot της ενέργειας της αναζήτησης.....	107
Εικόνα 20: Screen shot της ενέργειας της ανάλυσης.....	108
Εικόνα 21: Παράδειγμα επιτυχημένης ερώτησης(1), ο χρήστης εισάγει την ερώτησή του στο αντίστοιχο πεδίο και πατώντας επάνω στο κουμπί «Αναζήτηση», γίνεται αναζήτηση της ερώτησης. Στη συνέχεια η απάντηση της ερώτησης εμφανίζεται στο αντίστοιχο πεδίο ενώ εάν αυτή είναι θετική, εμφανίζεται και το κείμενο από το οποίο πρέκυψε η απάντηση.....	111
Εικόνα 22: Παράδειγμα επιτυχημένης ερώτησης(2), ο χρήστης πατώντας το κουμπί «Σύνταξη» μπορεί να δει τη σύνταξη του κειμένου απάντησης. Επίσης έχει τη δυνατότητα να επιλέξει να δει συγκεκριμένους συντακτικούς όρους, επιλέγοντας τα αντίστοιχα πεδία.....	112
Εικόνα 23: Παράδειγμα ανάλυσης κειμένου(1). Ο χρήστης έχει τη δυνατότητα να επιλέξει ένα συγκεκριμένο συντακτικό όρο και αυτός, εάν υπάρχει, θα τονισθεί μέσα στο κείμενο. Κάτω από το κείμενο θα ακολουθήσει η γραμματική ανάλυση του όρου.....	113
Εικόνα 24: Παράδειγμα ανάλυσης κειμένου(2). Παρόμοιο με το παράδειγμα της εικόνας που προηγήθηκε, αλλάζει μόνο ο συντακτικός όρος που αναζητάται.....	114

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας 1: Οι περιπτώσεις του ονοματικού συνόλου	27
Πίνακας 2: Οι περιπτώσεις του ρηματικού συνόλου	31
Πίνακας 3: Στοιχεία για το μορφολογικό λεξικό της EM2 του ΙΤΥ	39
Πίνακας 4: Σύμβολα κανονικών εκφράσεων	51
Πίνακας 5: Σύμβολα που χρησιμοποιούνται	51
Πίνακας 6: Οι σύνδεσμοι	57
Πίνακας 7: Αποτελέσματα evaluation με κείμενα εφημερίδων	115
Πίνακας 8: Στην πρώτη στήλη του πίνακα εμφανίζονται οι ερωτήσεις. Στη δεύτερη εμφανίζεται ο αναμενόμενος αριθμός rss που απαντούν την ερώτηση χωρίς τη χρήση του συντακτικού αναλητή, με απλό pattern match των keyword. Στην τρίτη και τελευταία στήλη εμφανίζονται τα αποτελέσματα με χρήση του συντακτικού μας αναλυτή.	117

Κεφάλαιο 1

Εισαγωγή

1.1 Εισαγωγή – Αναγκαιότητα

Η γλώσσα αποτελεί για τον άνθρωπο το κύριο μέσο ανταλλαγής και καταγραφής πληροφοριών. Επίσης, η γλώσσα αποτελεί μέσο έκφρασης συναισθημάτων. Από την πολυπλοκότητα και το εύρος των σκέψεων, των ιδεών και των συναισθημάτων που μπορούν να εκφραστούν μέσω της γλώσσας, συμπεραίνουμε ότι πρόκειται για ένα μέσο αρκετά σύνθετο και πολύπλοκο. Σύνηθες χαρακτηριστικό της γλωσσικής επικοινωνίας μεταξύ των ανθρώπων είναι η παρερμηνεία και η παρεξήγηση, γιατί η ερμηνεία των λεγόμενων προϋποθέτει γνώση τόσο για το ίδιο το γλωσσικό σύστημα όσο και για τον κόσμο και τις έννοιες που εκφράζονται μέσω της γλώσσας. Ο τρόπος που ένας άνθρωπος χειρίζεται τη γλώσσα αποτελεί βασικό κοινωνικό δείκτη. Η ίδια η γλώσσα είναι ο κυριότερος φορέας πολιτισμικών χαρακτηριστικών.

Εφόσον το μεγαλύτερο κομμάτι της ανθρώπινης γνώσης είναι καταγραμμένο μέσω της φυσικής γλώσσας¹ (Natural Language Processing ή αλλιώς NLP) , η πρόσβαση σε αυτή τη γνώση θα ήταν εξαιρετικά εύκολη, ιδιαίτερα τώρα που υπάρχει το διαδίκτυο, αν η φυσική γλώσσα ήταν κατανοητή από τον υπολογιστή. Αρκεί να φανταστούμε τον υπολογιστή:

- να απαντά ερωτήσεις στο τηλέφωνο
- να επεξεργάζεται το κείμενο μιας ιστοσελίδας και να στέλνει μια περίληψή της σε πιθανούς ενδιαφερόμενους
- να μεταφράζει μια καθημερινή εφημερίδα από τα Αγγλικά στα Ελληνικά
- να εξάγει γνώση από περιοδικά/ βιβλία/ ιστοσελίδες και να δημιουργεί ένα έμπειρο σύστημα που να καθοδηγεί τη λήψη αποφάσεων

Ο στόχος της επεξεργασίας της φυσικής γλώσσας είναι να σχεδιάσει και να κατασκευάσει λογισμικό που θα αναλύει και θα καταλαβαίνει γλώσσες που οι

¹ ο όρος *φυσική γλώσσα* αναφέρεται στη γλώσσα ως μέσο επικοινωνίας των ανθρώπων, σε αντιδιαστολή προς τον όρο *τεχνητή γλώσσα* ως μέσο επικοινωνίας ανθρώπου-υπολογιστή ή υπολογιστή-υπολογιστή

άνθρωποι χρησιμοποιούν φυσικά, έτσι ώστε τελικά ο χρήστης να μπορεί να απευθυνθεί στον υπολογιστή του με τον ίδιο τρόπο που θα απευθυνόταν σε ένα άλλο άτομο.

Αυτός ο στόχος βέβαια δεν είναι εύκολο να επιτευχθεί. Το να καταλαβαίνει κάποιος τη γλώσσα σημαίνει μεταξύ άλλων, να γνωρίζει ποιες έννοιες μία λέξη ή φράση αντιπροσωπεύουν και να ξέρει πως να συνδέσει αυτές τις έννοιες μαζί με τρόπο τέτοιο ώστε το τελικό αποτέλεσμα να έχει νόημα. Αποτελεί σχήμα οξύμωρο το γεγονός ότι η φυσική γλώσσα, το σύστημα συμβόλων που είναι το ευκολότερο για τους ανθρώπους να καταλάβουν, είναι το πιο δύσκολο αντίστοιχα για τους υπολογιστές.

Υπάρχουν πολλές απόψεις σχετικά με την ικανότητα των υπολογιστών να επεξεργαστούν και να κατανοήσουν τη φυσική γλώσσα. Πολλοί τροφοδοτούν την υπερβολή, προβλέποντας ότι σε λίγα χρόνια θα υπάρχουν υπολογιστές που θα επικοινωνούν με τους ανθρώπους σε φυσική γλώσσα, κάτι που μέχρι σήμερα υποστηριζόταν μόνο από σενάρια επιστημονικής φαντασίας. Το 1968 η ταινία “2001: A Space Odyssey” των Kubrick και Clarke παρουσίαζε έναν υπολογιστή, τον Hal, που συνδιαλεγόταν με τους αστροναύτες: αναγνώριση, κατανόηση και παραγωγή ομιλίας για την παροχή συμβουλών ψυχολογικού περιεχομένου από έναν υπολογιστή-μέντορα. Οι μικροεπεξεργαστές, οι οποίοι ανακαλύφθηκαν πέντε χρόνια μετά την ταινία, μπορούν σήμερα και ελέγχουν τη λειτουργία των κινητήρων των αυτοκινήτων χιλιάδες φορές το δευτερόλεπτο, δεν μπορούν όμως να προσδώσουν στους προσωπικούς υπολογιστές την ικανότητα να συνδιαλέγονται με τους χρήστες τους, ούτε καν στοιχειωδώς.

Για την επικοινωνία των υπολογιστών μεταξύ τους, ή για την επικοινωνία ανθρώπου-υπολογιστή, επινοήθηκαν και επινοούνται διάφορες γλώσσες, οι λεγόμενες *τυπικές* ή *τεχνητές γλώσσες*. Η κυριότερη διάκριση των τεχνητών γλωσσών είναι σε γλώσσες *χαμηλού* και σε γλώσσες *υψηλού επιπέδου*. Οι γλώσσες χαμηλού επιπέδου είναι πιο “κοντά” στον τρόπο που αντιλαμβάνεται ο υπολογιστής την πληροφορία, με αριθμούς, σύμβολα και αριθμητικές /συμβολικές εκφράσεις, ενώ οι γλώσσες υψηλού επιπέδου προσπαθούν να προσεγγίσουν τον άνθρωπο και την ανθρώπινη λογική, με δομημένους τύπους δεδομένων και πολύπλοκες εκφράσεις. Ακόμη όμως και οι τεχνητές γλώσσες υψηλού επιπέδου έχουν τόση εκφραστική δυνατότητα όση τους επιτρέπει να μεταφράζονται μονοσήμαντα σε αριθμητικά δεδομένα και αριθμητικές εκφράσεις, γεγονός που τις φέρνει πιο “κοντά” στον υπολογιστή παρά στον άνθρωπο.

Παρ' όλα αυτά, ο άνθρωπος χρησιμοποιεί εκτεταμένα τους υπολογιστές για να παραγάγει, να αποθηκεύσει, να οργανώσει και να μεταδώσει πληροφορίες κωδικοποιημένες σε φυσική γλώσσα, υπό τη μορφή ηλεκτρονικών κειμένων ή ψηφιοποιημένης ομιλίας. Αυτό βέβαια δεν υπονοεί κάποια μορφή επικοινωνίας του ανθρώπου με τον υπολογιστή σε φυσική γλώσσα. Πηγή και αποδέκτης της πληροφορίας σε φυσική γλώσσα είναι ο άνθρωπος.

Η επιστήμη της Πληροφορικής μελετά την αναπαράσταση, επεξεργασία και ανταλλαγή πληροφορίας σε φυσικά και τεχνητά συστήματα. Έχει υπολογιστικό, γνωστικό και κοινωνικό προσανατολισμό. Κεντρική έννοια είναι η μετάπλαση της πληροφορίας — είτε υπολογιστικά είτε επικοινωνιακά, είτε από φυσικούς οργανισμούς είτε από τεχνουργήματα. Στα φυσικά και τεχνητά συστήματα, η πληροφορία οργανώνεται σε διάφορα επίπεδα, π.χ. στα βιολογικά μόρια και στα ηλεκτρονικά κυκλώματα, στα νευρικά συστήματα και στους ηλεκτρονικούς υπολογιστές, στις ανθρώπινες κοινωνίες και στα κατανεμημένα υπολογιστικά συστήματα. Η Πληροφορική έρχεται να συμπληρώσει και να προαγάγει όλες τις υπόλοιπες επιστήμες (που παραδοσιακά μελετούν τη φύση, τον άνθρωπο και τα κατασκευάσματα του ανθρώπου): κοινός παρονομαστής όλων είναι η *πληροφορία*. Η φυσική γλώσσα, το βασικό συστατικό των ανθρώπινων πληροφοριακών συστημάτων, δεν μπορεί παρά να αποτελεί αντικείμενο μελέτης της Πληροφορικής, και πιο συγκεκριμένα της Υπολογιστικής Γλωσσολογίας, ενός κλάδου που συνδυάζει την Πληροφορική με τη Γλωσσολογία.

1.2 Γλώσσα και ασάφεια

Γιατί είναι τόσο δύσκολη η επεξεργασία της φυσικής γλώσσας από τον υπολογιστή; Επειδή η φυσική γλώσσα είναι εκ φύσεως ασαφής: σχεδόν κάθε λέξη έχει πολλαπλές ερμηνείες και χρήσεις. Αυτή η δυσκολία γίνεται ακόμη μεγαλύτερη στην περίπτωση της ελληνικής γλώσσας, μία γλώσσα με πολλές ιδιαιτερότητες τόσο στον προφορικό όσο και στον γραπτό λόγο.

Όταν οι άνθρωποι επεξεργάζονται τη γλώσσα, κάνουν διαρκώς επιλογή μεταξύ διαφορετικών ερμηνειών χρησιμοποιώντας εκτενώς τη γνώση τους για τον

κόσμο και τον πολιτισμό. Για να μπορέσει ο υπολογιστής να κατανοήσει και να επιλύσει την ασάφεια, πρέπει να κάνει κτήμα του αυτή τη γνώση, η οποία είναι τόσο μεγάλη σε όγκο και τόσο πολύπλοκη, που η υπάρχουσα τεχνολογία δεν μπορεί να τη χειριστεί.

Οι άνθρωποι χρησιμοποιούν τη γλώσσα μέσα σε συγκεκριμένες περιστάσεις επικοινωνίας. Η γνώση της περίπτωσης μπορεί να παράσχει μέχρι και το 90% της πληροφορίας που είναι απαραίτητη για την επικοινωνία και επιτρέπει στα λεγόμενα που αναφέρονται στα πλαίσια της περίπτωσης να είναι σαφή. Αν επισκεφτούμε ένα φίλο μας μια ζεστή μέρα και τον ρωτήσουμε «Έχει νερό το ψυγείο;», αυτός θα σκεφτεί αν θυμήθηκε να βάλει ένα μπουκάλι με νερό στο ψυγείο. Είναι απίθανο να σκεφτεί αν θυμήθηκε να γεμίσει με νερό το ψυγείο του αυτοκινήτου του. Αν βέβαια βρούμε το φίλο μας στο γκαράζ να ασχολείται με το αυτοκίνητό του, τότε είναι πιθανότερο η ίδια ερώτηση να αναφέρεται στο ψυγείο του αυτοκινήτου.

Γενικά αναγνωρίζονται πέντε επίπεδα ανάλυσης του γλωσσικού συστήματος [Lyons 1981, Φιλιππάκη-Warburton 1992, Allen 1987, Grishman 1987]:

- *Φωνητική-Φωνολογία*. Η Φωνητική ασχολείται με την προφορά των λέξεων, καθεμιάς χωριστά ή συνδυασμένων μεταξύ τους μέσα σε προτάσεις. Η Φωνολογία ασχολείται με τη λειτουργία των φθόγγων μέσα σε ένα συγκεκριμένο γλωσσικό σύστημα.
- *Μορφολογία*. Η Μορφολογία ασχολείται με την εσωτερική δομή των λέξεων και το σχηματισμό τους από τα συστατικά τους στοιχεία, τα μορφήματα.
- *Σύνταξη*. Η Σύνταξη ασχολείται με τους τρόπους και τους κανόνες σύμφωνα με τους οποίους οι λέξεις συνδυάζονται μέσα σε μεγαλύτερες ενότητες, σχηματίζοντας φράσεις και προτάσεις.
- *Σημασιολογία*. Η Σημασιολογία ασχολείται με τα νοήματα που εκφράζουν οι λέξεις και τον τρόπο που αυτά τα νοήματα αλληλεπιδρούν για τη διαμόρφωση του νοήματος των προτάσεων.
- *Πραγματολογία*. Η Πραγματολογία ασχολείται με την εξάρτηση του νοήματος μιας πρότασης από τα ευρύτερα (γλωσσικά και εξωγλωσσικά) συμφραζόμενα μέσα στα οποία εμφανίζεται.

Η γλωσσική ασάφεια οφείλεται στο γεγονός ότι σε κάποιο από τα επίπεδα ανάλυσης του γλωσσικού συστήματος η ίδια οντότητα αντιστοιχεί σε διαφορετικές

οντότητες κάποιου άλλου επιπέδου. Για παράδειγμα στην ελληνική γλώσσα, η λέξη *λόγος* είναι συγκεκριμένη μορφολογική οντότητα, η οποία αντιστοιχεί σε περισσότερες από μια έννοιες στο σημασιολογικό επίπεδο (*ομιλία* || *λεγόμενα* || *αιτία* || *αιτιολογία* || *υπόσχεση* κτλ.), και η φράση *χτύπησε τη γυναίκα με τα λουλούδια* αποτελείται από σαφώς διακεκριμένες μορφολογικές και σημασιολογικές οντότητες αλλά επιδέχεται περισσότερες από μία

δομικές αναλύσεις στο συντακτικό επίπεδο, που διαφοροποιούν τη σημασία σε επίπεδο πρότασης (*χτύπησε [τη γυναίκα με τα λουλούδια]*, *χτύπησε [τη γυναίκα] [με τα λουλούδια]*).

Γλωσσική ασάφεια μπορεί επίσης να προκύψει κατά τη μεταφορά του λόγου από το φωνητικό μέσο (=ομιλία) στο γραφικό μέσο (=κείμενο), ή το αντίστροφο. Για παράδειγμα στην ελληνική γλώσσα, στο γραφικό μέσο η λέξη *ήπια* (< [ήπιος], [πίνω]) αναπαριστά δύο διαφορετικές λέξεις του φωνητικού μέσου, ενώ τα ζεύγη *κλίμα-κλήμα*, *είμαστε-ήμαστε*, *μάλλον-μ'άλλον*, που είναι διακεκριμένα στο γραφικό μέσο, είναι πανομοιότυπα στο φωνητικό μέσο. Επίσης, ο επιτονισμός λέξεων ή φράσεων στο φωνητικό μέσο, ο οποίος βοηθά σημαντικά στην αποσαφήνιση, χάνεται στο γραφικό μέσο.

1.3 Στόχοι διπλωματικής

Το αντικείμενο της παρούσας διπλωματικής εργασίας είναι η υλοποίηση ενός **συντακτικού αναλυτή** για την ελληνική φυσική γλώσσα και η οποία θέτει τους εξής στόχους:

1. Ο **συντακτικός αναλυτής** να είναι σε θέση να αναγνωρίσει τους βασικούς συντακτικούς όρους² της Νέας ελληνικής γλώσσας με πολύ μεγάλη ακρίβεια. Για να το καταφέρει αυτό πρέπει να διαθέτει ένα σύνολο υπολογιστικών μηχανισμών με τους οποίους να:
 - i. χειρίζεται τους κανόνες που διέπουν το ελληνικό συντακτικό
 - ii. χειρίζεται τις συντακτικές και σημασιολογικές ασάφειες

² Συντακτικοί όροι είναι οι όροι εκείνοι που μπορούν να συναντηθούν σε μία πρόταση, όταν αυτή αναλύεται συντακτικά. Για παράδειγμα, συντακτικοί όροι είναι το υποκείμενο, το ρήμα, το αντικείμενο κ.ά.

- iii. χρησιμοποιεί τη γνώση που του παρέχεται τόσο από παραδοχές αλλά και από άλλες εφαρμογές για την καλύτερη και πληρέστερη συντακτική ανάλυση
- 2. Ο **συντακτικός αναλυτής** να είναι σε θέση να κάνει συντακτική ανάλυση σε κείμενα και συλλογές κειμένων. Έτσι ο αναλυτής μπορεί να εφαρμοστεί και να αξιολογηθεί σε μεγάλες συλλογές κειμένων που περιέχουν το σύνολο των συντακτικών όρων που μπορούν να εμφανιστούν στην ελληνική γλώσσα.
- 3. Ο **συντακτικός αναλυτής** να χρησιμοποιείται σε αναζήτηση ερωτήσεων και να επιστρέφει αποτελέσματα με βάση συντακτικά κριτήρια. Οι ερωτήσεις μπορούν επίσης να αναλυθούν συντακτικά με μεγάλη ακρίβεια ενώ μπορούν να απευθύνονται σε συλλογές κειμένων και να αναζητούν συντακτικούς όρους(ποιος, τι, που). Γίνεται, δηλαδή ,ουσιαστικά και σημασιολογική αναζήτηση.
- 4. Ο **συντακτικός αναλυτής** να είναι ανεξάρτητος του γραμματικού αναλυτή που προηγείται της συντακτικής ανάλυσης. Μπορεί να εφαρμοστεί και να δουλέψει με επιτυχία με οποιονδήποτε γραμματικό αναλυτή με μικρές αλλαγές. Αλλαγές στην ακρίβεια των αποτελεσμάτων θα εμφανιστούν σε εκείνα τα αποτελέσματα τα οποία εξαρτώνται άμεσα από τα αποτελέσματα του γραμματικού αναλυτή και μόνο.

Για να υλοποιήσουμε όλους τους παραπάνω στόχους θα υλοποιήσουμε ένα συντακτικό αναλυτή σε μία διεπαφή φυσικής γλώσσας, για τη Νέα ελληνική. Η διεπαφή φυσικής γλώσσας (Natural Language Interface) χρησιμοποιείται σε ένα σύστημα ερωταπαντήσεων (question – answering) το οποίο έχει ως είσοδο το αίτημα του χρήστη σε φυσικό λόγο και μετά από διάφορα στάδια επεξεργασίας επιστρέφει αποτελέσματα σε αυτόν. Ο έλεγχος για πιθανές απαντήσεις γίνεται επάνω σε συλλογές κειμένων. Επίσης θα υλοποιηθεί και μία ακόμη διεπαφή στην οποία ο συντακτικός αναλυτής θα εφαρμόζεται πάνω σε ένα κείμενο και θα μας δίνει τη δυνατότητα να κάνουμε αναζήτηση συγκεκριμένων συντακτικών όρων.

1.4 Συνεισφορά της διπλωματικής

Μετά από επισκόπηση της σχετικής βιβλιογραφίας πάνω στο θέμα της συντακτικής ανάλυσης η παρούσα διπλωματική εργασία έρχεται να καλύψει με ολοκληρωμένο τρόπο το θέμα αυτό για τη Νέα Ελληνική, μια γλώσσα που δεν έχει τύχει ιδιαίτερου ενδιαφέροντος από τη διεθνή κοινότητα των υπολογιστικών γλωσσολόγων. Το απόσταγμα όσων παρουσιάζονται στην εργασία θα μπορούσε να εκφραστεί μέσω των παρακάτω ισχυρισμών:

Ισχυρισμός #1: Ένα εύρωστο σύστημα συντακτικής ανάλυσης αποτελεί το υπόβαθρο τόσο για την προαγωγή της έρευνας στην υπολογιστική γλωσσολογία όσο και για την ανάπτυξη συστημάτων προηγμένης γλωσσικής τεχνολογίας.

Ισχυρισμός #2: Το κομμάτι της (συντακτικής) γνώσης που είναι δύσκολο και χρονοβόρο να εξαχθεί από τον άνθρωπο μπορεί να εξαχθεί αυτόματα μέσω κατάλληλων εφαρμογών.

Ισχυρισμός #3: Η γνώση που εξάγεται **από** τον υπολογιστή **για** τον υπολογιστή μπορεί να είναι κατανοητή και από τον άνθρωπο.

Ισχυρισμός #4: Η συντακτική ανάλυση μιας γλώσσας με “πλούσια” μορφολογία, όπως είναι η Νέα Ελληνική, δεν είναι ευκολότερη ή δυσκολότερη από τη μορφοσυντακτική ανάλυση μιας γλώσσας με “φτωχή” μορφολογία, όπως είναι η Αγγλική. Απλά, πρέπει να προσεγγιστεί με διαφορετικό τρόπο.

Ισχυρισμός #5: Η συντακτική ανάλυση της Νέας Ελληνικής μπορεί να διεξαχθεί με υψηλά ποσοστά επιτυχίας με μικρή χρήση σημασιολογικής πληροφορίας.

1.5 Δομή της εργασίας

Μετά το τρέχον –πρώτο κεφάλαιο-, το οποίο αποτελεί και την εισαγωγή της διπλωματικής διατριβής, ακολουθούν πέντε ακόμα κεφάλαια.

Στο δεύτερο κεφάλαιο γίνεται μια αναλυτική παρουσίαση της θεωρίας που αποτέλεσε τη βάση της διπλωματικής εργασίας. Για την καλύτερη κατανόηση υπάρχουν και αρκετά παραδείγματα σε αυτό το κεφάλαιο.

Στο τρίτο κεφάλαιο παρουσιάζεται το μοντέλο του συντακτικού αναλυτή που αναπτύξαμε. Γίνεται αναλυτική περιγραφή της δομής του και του τρόπου λειτουργίας του.

Στο τέταρτο κεφάλαιο παρουσιάζεται η υλοποίηση του συντακτικού αναλυτή. Σε αυτό το κεφάλαιο γίνονται πολλές αντιστοιχίσεις με τη θεωρία του συντακτικού αναλυτή που παρουσιάστηκε στο προηγούμενο κεφάλαιο.

Στο πέμπτο κεφάλαιο γίνεται αναφορά στη γραφική διεπαφή που κατασκευάστηκε και παρουσιάζεται και η χρήση της.

Στο έκτο και τελευταίο κεφάλαιο, γίνεται αξιολόγηση της εφαρμογής ενώ προτείνονται και ορισμένες μελλοντικές επεκτάσεις.

Κεφάλαιο 2

Επισκόπηση σχετικής θεωρίας

2.1 Φυσική Γλώσσα

Στις ακόλουθες παραγράφους θα παρουσιαστεί το απαραίτητο θεωρητικό υπόβαθρο με σκοπό την ομαλή εισαγωγή μας στις έννοιες και λειτουργίες που σχετίζονται με τη φυσική γλώσσα.

2.1.1 Προκαταρκτικά

Προκειμένου να καταστεί το υπόλοιπο αυτής της διατριβής συνεπές και κατανοητό, θα πρέπει να διευκρινιστούν μερικές βασικές έννοιες. Αυτή η εργασία εξετάζει τις πτυχές της γλωσσικής τεχνολογίας, και όχι τη λεκτική τεχνολογία. Η λεκτική τεχνολογία αναφέρεται στην επεξεργασία γλωσσών στο επίπεδο της φωνολογίας και της φωνητικής και θεμελιώνεται στην Ακουστική, την Ηλεκτρομηχανική και την Πληροφορική.

Η γλωσσική τεχνολογία είναι η διαδικασία της ανάλυσης της εισαγωγής κειμένου από την άποψη της σύνταξης, της μορφολογίας, και της σημασιολογίας. Τα θεμέλια για αυτήν βρίσκονται στην υπολογιστική γλωσσολογία, την ψυχολογία και την πληροφορική. Τα συστήματα που περιγράφονται εδώ είναι συστήματα διαλόγου που περιορίζονται στην επεξεργασία της γραπτής - ή δακτυλογραφημένης - αλληλεπίδρασης [4].

Έκφραση (utterance) είναι ένας γενικός όρος που αναφέρεται σε οποιοδήποτε τρόπο επικοινωνίας. Δηλαδή, για παράδειγμα, μια έκφραση μπορεί να δακτυλογραφηθεί ή να υπαγορευτεί [19].

Η έννοια της φυσικής γλώσσας (natural language) αφορά στις ανθρώπινες γλώσσες που μπορούν να γραφτούν ή/ και να μιληθούν (π.χ. Αγγλικά, Ελληνικά).

2.1.2 Συντακτική ανάλυση και ασάφεια

Η σύνταξη μιας φυσικής γλώσσας αναπαρίσταται υπολογιστικά με τη βοήθεια κανόνων φραστικής δομής (*phrase structure rules*, βλ. [Chomsky 1965][5]). Σε πρώτο επίπεδο, οι κανόνες φραστικής δομής περιγράφουν το σχηματισμό φράσεων ή εκφράσεων από λέξεις. Ο συντακτικός ρόλος των φράσεων που παράγονται εξαρτάται από το συντακτικό ρόλο του κυριότερου από τις λέξεις-συστατικά, το οποίο καλείται *κεφαλή* (*head*). Το στοιχείο που προσδιορίζει το συντακτικό ρόλο μιας λέξης είναι το Μέρος του Λόγου (ΜτΛ), το οποίο αναφέρεται στη βιβλιογραφία και ως *major feature* [Pollard and Sag 1987, σελ. 59][6]. Όταν το ΜτΛ μιας λέξης είναι ασαφές, ο συντακτικός αναλυτής πρέπει να εξετάσει όλους τους πιθανούς συντακτικούς της ρόλους και στην περίπτωση που ενεργοποιούνται περισσότεροι από έναν κανόνες, να παραγάγει όλες τις φραστικές δομές που αυτοί υπαγορεύουν, με την ελπίδα ότι μόνο μία ανάλυση τελικά θα επιτύχει. Αυτό όμως προϋποθέτει την επιτυχία στην αναγνώριση γειτονικών δομών, γεγονός που είναι αμφίβολο, αν θεωρήσουμε ότι και αυτές μπορεί να περιέχουν ασαφή συστατικά.

Υπάρχουν πολλές περιπτώσεις φυσικού λόγου, στις οποίες παρουσιάζονται ασάφειες στα μέρη του λόγου που αποτελούν τη πρόταση, δηλαδή περιπτώσεις που ίδιες λέξεις αποτελούν διαφορετικό μέρος του λόγου ανάλογα με τη πρόταση στην οποία βρίσκονται. Για παράδειγμα, η λέξη “*κλειδώσεις*” ως ουσιαστικό μπορεί να παίζει το ρόλο κεφαλής³ σε μια ονοματική φράση⁴, ενώ ως ρήμα μπορεί να παίζει το ρόλο κεφαλής σε μια ρηματική φράση⁵. Ας θεωρήσουμε τη λέξη “*κλειδώσεις*” στις δύο παρακάτω περιπτώσεις.

1. Να θυμηθείς πριν φύγεις να *κλειδώσεις*
2. Οι *κλειδώσεις* του σώματος φθείρονται εύκολα

Παρατηρούμε λοιπόν ότι η λέξη “*κλειδώσεις*” είναι ρήμα στο πρώτο παράδειγμα ενώ ουσιαστικό στο δεύτερο. Αντιλαμβανόμαστε λοιπόν πόσο σημαντικό είναι να διακρίνουμε σωστά τα μέρη του λόγου, έτσι ώστε να αποσαφηνίσουμε αποδοτικά το νόημα της κάθε πρότασης. Η διάκριση αυτή είναι απαραίτητη εάν

³ Η λέξη που δεσπόζει στη φράση

⁴ Η κυρίαρχη λέξη είναι όνομα

⁵ Η κυρίαρχη λέξη είναι ρήμα

επιθυμούμε να αναλύσουμε τη πρόταση ,αφού ίδιες λέξεις αν ερμηνευτούν με διαφορετικό τρόπο, δίνουν τελείως διαφορετικό νόημα στη πρόταση.

Επιπλέον, είναι επιθυμητό η αποσαφήνιση του ρόλου της λέξης **“κλειδώσεις”** να μη βασιστεί στους κανόνες σχηματισμού⁶ ονοματικών και ρηματικών φράσεων. Με άλλα λόγια, είναι επιθυμητή η μορφοσυντακτική ασάφεια - κυρίως η ασάφεια ως προς το ΜτΛ -να έχει αρθεί προτού μπει σε εφαρμογή η διαδικασία της συντακτικής ανάλυσης, κάτι που προσφέρεται μέσω της μορφοσυντακτικής ανάλυσης. Επιμένοντας στο παράδειγμα, η λέξη **“κλειδώσεις”** ως ουσιαστικό μπορεί να έχει τη λειτουργία υποκειμένου όταν είναι σε ονομαστική ή τη λειτουργία αντικειμένου όταν είναι σε αιτιατική. Για παράδειγμά :

1. Οι **“κλειδώσεις”** φθείρονται εύκολα (υποκείμενο)
2. Ο Γιάννης πονάει στις **“κλειδώσεις”** του (αντικείμενο)

Και στις δύο όμως περιπτώσεις, αν ο συντακτικός αναλυτής γνωρίζει ότι πρόκειται για ουσιαστικό, θα εμπλέξει κανόνες σχηματισμού ονοματικών φράσεων και μόνο. Επιπλέον, αν προηγείται άρθρο, τότε ένας κανόνας της μορφής

Ονοματική Φράση → Άρθρο + Ουσιαστικό

θα αποδώσει στη λέξη **“κλειδώσεις”** την πτώση του άρθρου, αίροντας έτσι την ασάφεια ως προς την πτώση.

2.1.3 Σημασιολογική αποσαφήνιση

Σημασιολογικός σχολιασμός (semantic tagging) είναι η διαδικασία απόδοσης σημασιολογικής πληροφορίας στις λέξεις των ηλεκτρονικών κειμένων με τη βοήθεια ερμηνευτικού λεξικού. Το γεγονός ότι μια λέξη έχει κατά κανόνα περισσότερες από

⁶ Κανόνες που πρέπει να ακολουθηθούν για το σωστό ορισμό φράσεων

μία σημασίες - ένα φαινόμενο που ονομάζεται *πολυσημία* (polysemy, βλ. Kilgariff 1992)[7]) - εισάγει ένα πολύ δύσκολο πρόβλημα, που πρέπει να αντιμετωπιστεί από το σημασιολογικό σχολιασμό, το πρόβλημα της *σημασιολογικής αποσαφήνισης*: επιλογή εκείνης της σημασίας που ταιριάζει με το περιβάλλον εμφάνισης της λέξης. Οι Wilks και Stevenson [8] υποστηρίζουν ότι η σημασιολογική αποσαφήνιση και η μορφοσυντακτική αποσαφήνιση δεν είναι δυο ανεξάρτητες διαδικασίες, γιατί η υφή των γραμματικών κατηγοριών στις οποίες κατατάσσονται οι λέξεις βάσει των μορφοσυντακτικών τους χαρακτηριστικών είναι σημασιολογική. Μια λέξη που είναι ουσιαστικό αναφέρεται σε ένα πρόσωπο, τόπο, πράγμα ή σε μια αφηρημένη έννοια, ιδιότητα, ενέργεια ή κατάσταση, ενώ μια λέξη που είναι ρήμα εκφράζει μια ενέργεια ή μια κατάσταση Ένα ουσιαστικό σε ονομαστική μπορεί να έχει σχέση υποκειμένου με ένα ρήμα, ενώ ένα ουσιαστικό σε αιτιατική μπορεί να έχει σχέση αντικειμένου με ένα ρήμα (βλ. 2.1.2)· και οι δύο σχέσεις, αν και εξαρτώνται από το μορφοσυντακτικό χαρακτηριστικό Πτώση, ανήκουν στο συντακτικοσημασιολογικό επίπεδο. Έχοντας κατηγοριοποιήσει τις σημασίες μιας λέξης ανάλογα με τα ΜτΛ που αυτή μπορεί να λάβει, η επίλυση της ασάφειας ως προς το ΜτΛ αυτόματα περιορίζει και τις σημασίες που μπορεί να έχει η λέξη μέσα σε μια πρόταση.

Στη Νέα Ελληνική το φαινόμενο της ασάφειας ως προς το ΜτΛ (βλ. 5.1) δεν είναι τόσο έντονο. Παρ' όλα αυτά, η αποσαφήνιση του ΜτΛ σίγουρα είναι μια διαδικασία που βοηθά στη σημασιολογική αποσαφήνιση, π.χ. στην πρόταση:

Ο Βασίλης αισθάνθηκε το αίμα να χτυπάει με ορμή στις φλέβες του, μυρμηγκιάζοντας στις *κλειδώσεις*.

το γεγονός ότι η μορφοσυντακτική αποσαφήνιση αποφαίνεται ότι ο τύπος *κλειδώσεις* είναι ουσιαστικό και όχι ρήμα βοηθάει ένα σημασιολογικό σχολιαστή να του αποδώσει τις σημασίες του λεξήματος ⁷[*κλειδώση*] και όχι τις σημασίες του λεξήματος [*κλειδώνω*].

2.2 Συντακτικό ελληνικής γλώσσας

⁷ Λέξημα είναι το κυριότερο μέρος της λέξης που φέρει τη σημασία της

Στην παρούσα παράγραφο θα παρουσιαστούν τα βασικά μέρη του συντακτικού της ελληνικής γλώσσας με ιδιαίτερη έμφαση σε εκείνα τα συντακτικά φαινόμενα και κανόνες που αποτέλεσαν τη συντακτική βάση της υλοποίησής μας. Το σύνολο τόσο των φαινομένων όσο και των κανόνων προέκυψε από την ανάγνωση διαφορετικών συντακτικών και γραμματικών της νέας ελληνικής⁸ με στόχο αυτοί να είναι κατά το δυνατόν πληρέστεροι και κοινώς αποδεκτοί.

2.2.1 Πρόταση

Πρωταρχική δομική και λειτουργική μονάδα του λόγου είναι η πρόταση. Με τον όρο πρόταση εννοούμε το μικρότερο κομμάτι του λόγου, το οποίο αποδίδει ένα νόημα, είναι διαρθρωμένο σύμφωνα με τους συντακτικούς κανόνες και συμμετέχει στην απόδοση του συνολικού νοήματος του λόγου.

Η πρόταση πρέπει οπωσδήποτε να περιγράφει μια διαδικασία, μια ενέργεια, καθώς και όλους τους παράγοντες που καθορίζουν την ενέργεια αυτή. Τέτοιοι παράγοντες είναι το πρόσωπο το οποίο κάνει την ενέργεια (δράστης), η κατεύθυνση της ενέργειας και οι τρόποι, οι όροι και οι συνθήκες κάτω από τις οποίες πραγματοποιείται η ενέργεια αυτή.

2.2.2 Όροι πρότασης

Η πιο απλή διατύπωση που μπορεί να πάρει η πρόταση είναι εκείνη στην οποία υπάρχουν δύο απαραίτητοι όροι, που είναι οι πυρήνες γύρω από τους οποίους οργανώνονται τα άλλα μέλη της προτάσεως, για να δώσουν σε αυτή το νόημα που θέλουμε. Αυτοί οι όροι είναι:

α) το **υποκείμενο** , δηλαδή οι λέξεις ή η λέξη που φανερώνει για ποιόν γίνεται λόγος μέσα στην πρόταση και

⁸ «Συντακτικό της Νέας Ελληνικής», «Νεοελληνική Γραμματική», «Αναλυτική Γραμματική της Νέας Ελληνικής», «Γραμματική της ελληνικής γλώσσας»

β) το **κατηγορημα** , δηλαδή οι λέξεις ή η λέξη, η οποία φανερώνει εκείνο που λέγεται μέσα στην πρόταση για το υποκείμενο.

2.2.2.1 Υποκείμενο

Το υποκείμενο μιας πρότασης το βρίσκουμε απαντώντας στην ερώτηση ποιος, ποια, ποιο. Αυτό μπορεί να είναι ουσιαστικό (παρ. α, η, θ,) αντωνυμία (παρ. β), ουσιαστικοποιημένο επίθετο (παρ. γ), καθώς και κάθε άλλη λέξη, ομάδα λέξεων ή πρόταση με το άρθρο (σύνναρθρο) ή χωρίς άρθρο (άναρθρο), όταν παίρνουν τη θέση ουσιαστικού (παρ. δ, ε, στ, ζ, η1, θ1).

α. **Ο καθηγητής** γράφει

β. **Εκείνη** ήταν καλύτερη

γ. **Οι νέοι** διασκεδάζουν

δ. **Το αύριο** είναι άδηλο

ε. **Το πήγαινε έλα** το κούρασε

στ. **Το αν** είναι σύνδεσμος

ζ. **Το Α** είναι γράμμα

η. Απαγορεύεται **το κάπνισμα**

η1. Απαγορεύεται **να καπνίζετε**

θ. **Ο ερχομός σου** με ευχαρίστησε

θ1. **Το ότι ήρθες** με ευχαρίστησε

Το υποκείμενο είτε σύνναρθρο είτε άναρθρο **τοποθετείται ή εννοείται σε πτώση ονομαστική**

Όταν το υποκείμενο είναι πρώτου ή δεύτερου προσώπου, η αντωνυμία (*εγώ, εσύ – εμείς, εσείς*) συνήθως **παραλείπεται**, γιατί δηλώνεται από την κατάληξη του ρήματος (παρ. α, β). Όταν όμως θέλουμε να το αντιδιαστείλουμε από άλλο ή να το τονίσουμε, τότε μπαίνει (παρ. γ, δ).

α. Διαβάζεις; (*ποιος*; εννοείται *εσύ*)

β. Φεύγουμε για το χωριό. (*ποιοι*; εννοείται *εμείς*)

γ. *Εγώ* σας μιλώ και *εσείς* κοιμάστε.

δ. Το λόγο τώρα έχεις *εσύ*.

2.2.2.2 Κατηγορημα

Το κατηγορημα αποτελείται απαραίτητα από το **ρήμα** της πρότασης, συνοδευόμενο προαιρετικά από **κατηγορούμενο** ή **αντικείμενο**. Έτσι λοιπόν το κατηγορημα μιας πρότασης μπορεί:

α) να εκφράζεται μόνο με ρηματικό τύπο (παρ. α, β) και τότε λέγεται **κατηγορηματικό ρήμα**,

β) να αποτελείται από έναν τύπο του ρήματος *είμαι* ή άλλου συγγενικού και από ένα όνομα, επίθετο ή ουσιαστικό (παρ. γ-ζ). Το όνομα (επίθετο ή ουσιαστικό) που με τη μεσολάβησή του ρήματος *είμαι* ή άλλου συγγενικού, φανερώνει ποια ποιότητα ή ιδιότητα αποδίδεται στο υποκείμενο λέγεται **κατηγορούμενο** (προσεκτικός, μαθήτρια, καταγάλανος στρατηγός, δεμένοι) ενώ το ρήμα που μεσολαβεί μεταξύ του υποκειμένου και του κατηγορουμένου λέγεται **συνδετικό ρήμα**, γιατί συνδέει το κατηγορούμενο με το υποκείμενο,

γ) να αποτελείται από ένα οποιοδήποτε ρήμα και από ένα όνομα, επίθετο ή ουσιαστικό (παρ. η) στο οποίο αποδίδεται η έννοια του ρήματος (**μεταβατικό ρήμα**). Το όνομα (επίθετο ή ουσιαστικό) αυτό λέγεται **αντικείμενο**.

α. Ο ήλιος **λάμπει**

β. Εσύ **έλειπες**

γ. Ο Πέτρος **είναι προσεκτικός**

δ. Η Μαρία **είναι μαθήτρια**

ε. Ο ουρανός **ήταν καταγάλανος**

στ. Ο Αριστείδης **έγινε στρατηγός**

ζ. Οι αιχμάλωτοι **βρέθηκαν δεμένοι**

η. Ο Γιάννης πήγε **σχολείο**

2.2.3 Είδη προτάσεων

Κάθε πρόταση μπορούμε να την μελετήσουμε από τέσσερις διαφορετικές απόψεις ως προς: α) το περιεχόμενο, β) την ποιότητα, γ) τη σχέση της με τις άλλες και δ) την ιδιαίτερη δομή τους(τους όρους που τις αποτελούν).

2.2.3.1 Περιεχόμενο

Ως προς το περιεχόμενό τους, δηλαδή ως προς αυτό που θέλει να πει κάθε φορά ο ομιλητής, οι προτάσεις μπορούν να διαιρεθούν σε τέσσερις υποκατηγορίες.

i. Κρίσεως

Υπάρχουν προτάσεις με τις οποίες ο ομιλητής θέλει να πληροφορήσει, να δηλώσει κάτι, να διατυπώσει μια γνώμη ή να κρίνει κάτι:

- a. Πληροφορίας : *Χθες πήγα στο Πολυτεχνείο*
- b. Δήλωσης: *Δε θα κάνω διακοπές φέτος*
- c. Διατύπωση γνώμης: *Μου αρέσει πολύ η μπλούζα σου*

Οι προτάσεις αυτές ονομάζονται προτάσεις κρίσεως

ii. Επιθυμίας

Υπάρχουν προτάσεις με τις οποίες ο ομιλητής θέλει να εκφράσει παράκληση, προτροπή, προσταγή, ευχές κ.ά. και γενικά μια επιθυμία του:

- a. Επιθυμίας : *Κλείσε την πόρτα*
- b. Παράκλησης: *Δάνεισέ μου για λίγο το μολύβι σου*
- c. Προτροπής: *Φύγε αμέσως*
- d. Προσταγής: *Βγες έξω από την αίθουσα*
- e. Ευχής: *Με το καλό να γυρίσετε*

Αυτές οι προτάσεις ονομάζονται προτάσεις επιθυμίας

iii. Επιφωνηματικές

Υπάρχουν άλλες προτάσεις με τις οποίες εκφράζεται έκπληξη ή θαυμασμός και γενικά ένα έντονο συναίσθημα:

- a. Έκπληξης: *Γύρισε ολόκληρο τον κόσμο!*
- b. Θαυμασμού: *Είναι απίστευτο!*
- c. Κάθε έντονο συναίσθημα: *Τι φρικτό που ήταν!*

Αυτές οι προτάσεις στο γραπτό λόγο συνοδεύονται από θαυμαστικό (!), ενώ στον προφορικό διακρίνονται από τον τόνο της φωνής και ονομάζονται επιφωνηματικές προτάσεις.

iv. Ερωτηματικές

Υπάρχουν τέλος οι προτάσεις με τις οποίες ο ομιλητής διατυπώνει μια ερώτηση, για να πληροφορηθεί κάτι που δεν ξέρει, να επιβεβαιώσει μια πληροφορία του, να διαψεύσει κάτι, να προτρέψει να γίνει κάτι, να παρακαλέσει να γίνει κάτι :

- a. Άγνοιας: *Πώς ονομάζεσαι;*
- b. Επιβεβαίωσης: *Δε σου το 'χα πει;* (σου το είχα πει, ήμουν σίγουρος)
- c. Άρνησης: *Πώς είναι δυνατόν να έρθω κι εγώ;* (Δεν είναι δυνατόν να έρθω κι εγώ)
- d. Προτροπής: *Δεν πας καλύτερα μαζί του;* (Πήγαινε μαζί του)
- e. Παράκλησης: *Δε θα με βοηθήσεις;* (Βοήθησέ με)

Αυτές οι προτάσεις , που στο γραπτό λόγο συνοδεύονται από ερωτηματικό (;) και στον προφορικό με ανέβασμα του τόνου της φωνής, ονομάζονται ερωτηματικές προτάσεις.

2.2.3.2 Ποιότητα

Ως προς την ποιότητά τους , δηλαδή ως προς αυτό που εκφράζουν, οι προτάσεις μπορούν να διαιρεθούν σε δύο υποκατηγορίες.

i. Καταφατικές

Το ρήμα των προτάσεων αυτών είναι θετικό ή δε συνοδεύεται από αρνητικό μόριο ή από αρνητικό επίρρημα:

Επιστρέφω στο σπίτι

Μίλησέ μου για εσένα

ii. Αρνητικές

Το ρήμα των προτάσεων αυτών συνοδεύεται συνήθως από το αποφαιτικό μόριο δε(ν) :

Δε θα σου ξαναμιλήσω

Δεν τον ξαναείδε ποτέ

2.2.3.3 Σχέση τους με τις άλλες

Ως προς τη σχέση τους με τις άλλες προτάσεις που εμφανίζονται σε ένα κείμενο ή σε ένα διάλογο, οι προτάσεις μπορούν να κατηγοριοποιηθούν ως εξής:

i. Κύριες

Αυτές οι προτάσεις μπορούν να σταθούν μόνες τους στο λόγο και να εκφράζουν από μόνες τους ένα πλήρες νόημα:

Γύρισα σπίτι αργά τη νύχτα

Γιατί ήρθες τόσο αργά;

ii. Δευτερεύουσες

Αυτές οι προτάσεις δεν μπορούν να σταθούν μόνες τους στο λόγο και χρησιμεύουν για να προσδιορίσουν άλλη πρόταση ή κάποιο όρο άλλης πρότασης:

Δεν απάντησε στο τηλέφωνο, γιατί δεν ήταν σπίτι

Ο Γιώργος μου ζήτησε το βιβλίο ,που σου δάνεισα

2.2.3.4 Ως προς τη δομή τους

Ως προς τη δομή τους, δηλαδή ως προς τους όρους που τις αποτελούν (βλέπε επόμενη παράγραφο, όροι πρότασης), οι προτάσεις μπορούν να χωριστούν στις ακόλουθες τέσσερις κατηγορίες:

i. Απλές

Αποτελούνται μόνο από τους κύριους όρους της πρότασης(υποκείμενο - κατηγορήμα):

Ο Νίκος είναι ψηλός (υποκείμενο - κατηγορήμα)

ii. Σύνθετες

Διαθέτουν περισσότερα από ένα υποκείμενα ή / και κατηγορήματα:

Ο Νίκος είναι ψηλός ,ξανθός και λεπτός (υποκείμενο - πολλά κατηγορούμενα)

Ο Γιώργος και η Μαρία πήγαν σχολείο (πολλά υποκείμενα - κατηγορήμα)

Τα μήλα και τα πορτοκάλια είναι φρέσκα και γευστικά (πολλά υποκείμενα-πολλά κατηγορούμενα)

iii. Επαυξημένες

Οι κύριοι όροι των επαυξημένων προτάσεων συνοδεύονται από συμπληρώματα της έννοιάς τους που λέγονται προσδιορισμοί ή δευτερεύοντες όροι της πρότασης:

Ο Νίκος, ο αδερφός μου, είναι ψηλός (υποκείμενο – κατηγορήμα – δευτερεύοντες όροι)

iv. Ελλειπτικές

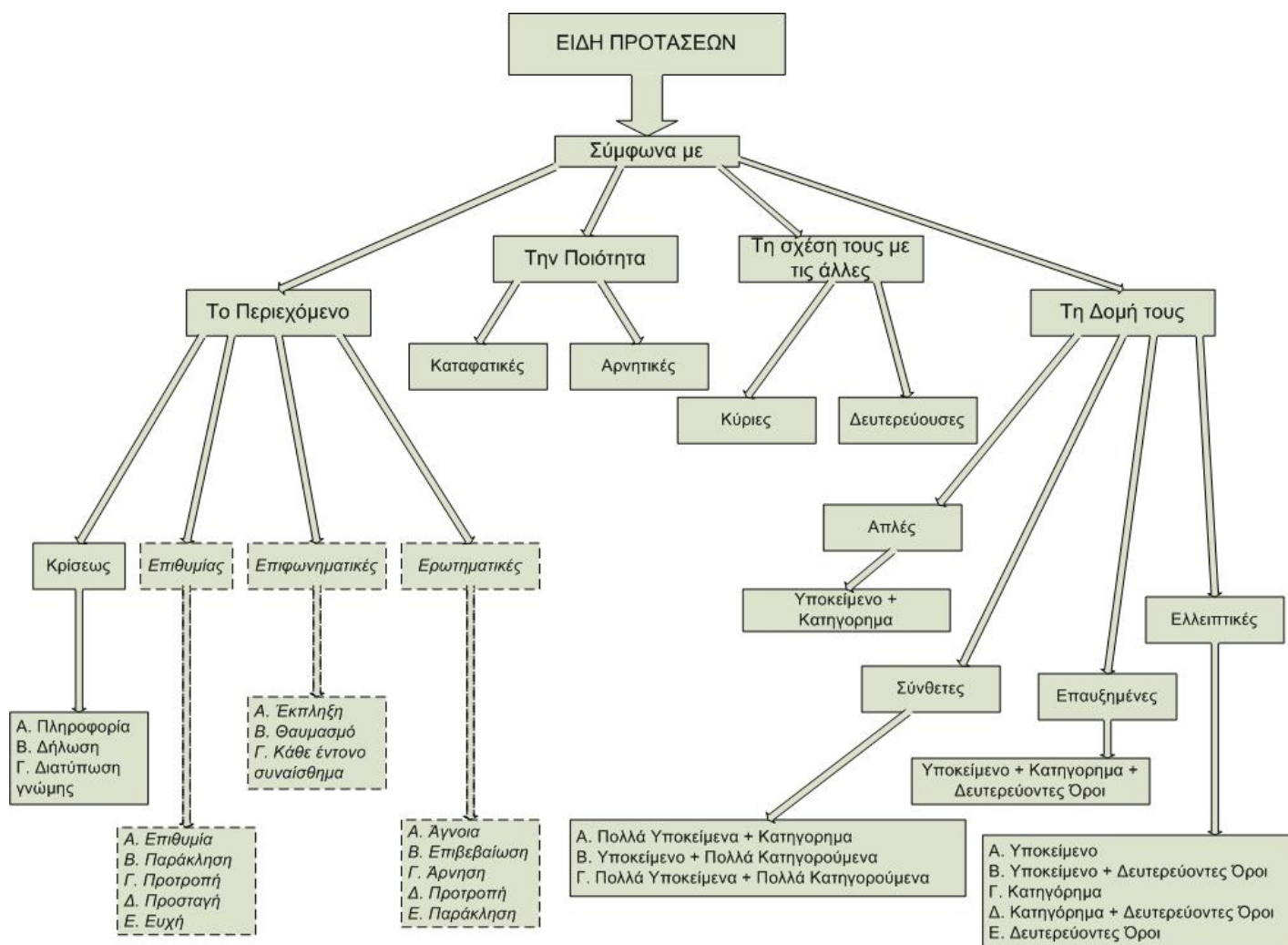
Πολύ συχνά στην καθημερινή μας ομιλία είναι δυνατόν να παραλείπουμε όρους της προτάσεως και να μιλάμε με τμήματά της, ακόμα και μονολεκτικά, τα οποία όμως εκφράζουν ολοκληρωμένο το νόημά της. Αυτές οι προτάσεις καλούνται ελλειπτικές:

Γεια χαρά (Σου εύχομαι υγεία και χαρά) (υποκείμενο)

Τι κρύο το χέρι σου! (Τι κρύο που είναι το χέρι σου) (υποκείμενο - δευτ. όροι)

Για που; (Για που πηγαίνεις;) (κατηγορήμα)

Στη συνέχεια παρουσιάζεται ένα σχεδιάγραμμα με όσα αναφέρθηκαν για τα είδη των προτάσεων σε αυτή την παράγραφο:



Εικόνα 1: Τα είδη των προτάσεων

2.2.4 Σύνδεση προτάσεων

Όπως είδαμε και παραπάνω (παρ. 2.2.3.3), οι προτάσεις είναι κύριες και δευτερεύουσες (εξαρτημένες). Όταν θέλουμε να εκφράσουμε ένα απλό νόημα, μας αρκεί μόνο μία κύρια πρόταση. Όταν όμως θέλουμε να εκφράσουμε ένα σύνθετο νόημα, έχουμε ανάγκη όχι μόνο από μία αλλά από δύο ή περισσότερες προτάσεις. Αυτές, ανάλογα με το νόημα που θέλουμε να διατυπώσουμε, μπορούν να συνδυαστούν ως εξής:

α) κύριες + κύριες,

β) κύριες + δευτερεύουσες,

γ) κύριες + δευτερεύουσες + δευτερεύουσες.

Όπως γίνεται εύκολα κατανοητό και από τον ορισμό τους (παρ. 2.2.3.3) δεν μπορούν να χρησιμοποιηθούν μόνο δευτερεύουσες προτάσεις για να έχουμε ένα ολοκληρωμένο νόημα.

Τα είδη της σύνδεσης των παραπάνω προτάσεων είναι δύο:

1. Παρατακτική σύνδεση ή παράταξη
2. Υποτακτική σύνδεση ή υπόταξη

Οι προτάσεις συνδέονται παρατακτικά ή υποτακτικά με τη χρησιμοποίηση των αντίστοιχων συνδέσμων (παρατακτικών ή υποτακτικών)

Μπορούμε να πούμε ότι με τον παρατακτικό τρόπο σύνδεσης συνδέονται **ισοδύναμες** προτάσεις μεταξύ τους (κύρια – κύρια, δευτερεύουσα - δευτερεύουσα), ενώ με τον υποτακτικό τρόπο σύνδεσης συνδέονται μία δευτερεύουσα με μια κύρια και μερικές φορές, όταν το συνολικό νόημα της περιόδου είναι πιο περίπλοκο και με την προϋπόθεση ότι στην περίοδο αυτή υπάρχει ήδη μια κύρια πρόταση, δευτερεύουσες μεταξύ τους.

1. Παράταξη:

Συναντάμε τέσσερις κατηγορίες συνδέσεων(με τους αντίστοιχους συνδέσμους):

A. Συμπλεκτικοί σύνδεσμοι:

και, δε(ν) – και δε(ν), και μη(ν), όχι μόνο – αλλά και, μήτε – μήτε, ούτε – ούτε, και να μη(ν), δε(ν) – ούτε

π.χ: «*Ούτε του μιλώ ούτε μου μιλάει*»

B. Αντιθετικοί σύνδεσμοι:

αλλά, όμως, ωστόσο, μολαταύτα, έπειτα, μονάχα που, εξάλλου, ναι μεν –
αλλά, μα, εντούτοις, αλλά όμως, έπειτα όμως, μονάχα όμως, εξάλλου όμως κτλ.

π.χ: *«Γκρίνιαζε όλη τη μέρα, αλλά κανείς δεν τον άκουγε»*

Γ. Διαζευκτικοί σύνδεσμοι:

ή, ή – ή, είτε – είτε, (κατά περίπτωση)θέλετε – θέλετε, θες – θες

π.χ: *«Η ηρέμησε ή φύγε»*

«Θες δεν ήθελε, θες δεν μπορούσε, δεν τα κατάφερε»

Δ. Συμπερασματικοί σύνδεσμοι:

ώστε, λοιπόν, άρα

π.χ: *«Είχαν απεργία τα ταξί, αναγκαστικά λοιπόν πήρα λεωφορείο»*

2. Υπόταξη:

Εδώ έχουμε 11 κατηγορίες. Ανάλογα με τους υποτακτικούς συνδέσμους με τους οποίους εισάγονται, οι δευτερεύουσες προτάσεις διακρίνονται σε:

1. Ειδικές

π.χ : *«Παραδέχομαι ότι είμαι ένοχος»*

2. Αιτιολογικές

π.χ : *«Φοβάμαι, γιατί είμαι μόνος στο σπίτι»*

3. Υποθετικές

π.χ : *«Αν πας Σαντορίνη, θα τα πούμε»*

4. Εναντιωματικές

π.χ : *«Αν δεν έρθεις, δεν θα τα πούμε»*

5. Τελικές

π.χ : *«Θα πάω στην Αθήνα, για να δω ποδόσφαιρο»*

6. Βουλητικές

π.χ : «*Η λύπη του τον εμπόδισε να φάει*»

7. Συμπερασματικές

π.χ : «*Το κρύο ήταν τσουχτερό, ώστε κανείς δε βγήκε*»

8. Χρονικές

π.χ : «*Καθώς γυρνούσα σπίτι, είδα τον Νίκο*»

9. Αναφορικές

π.χ : «*Κάνε για την υπόθεσή μου ότι εσύ νομίζεις*»

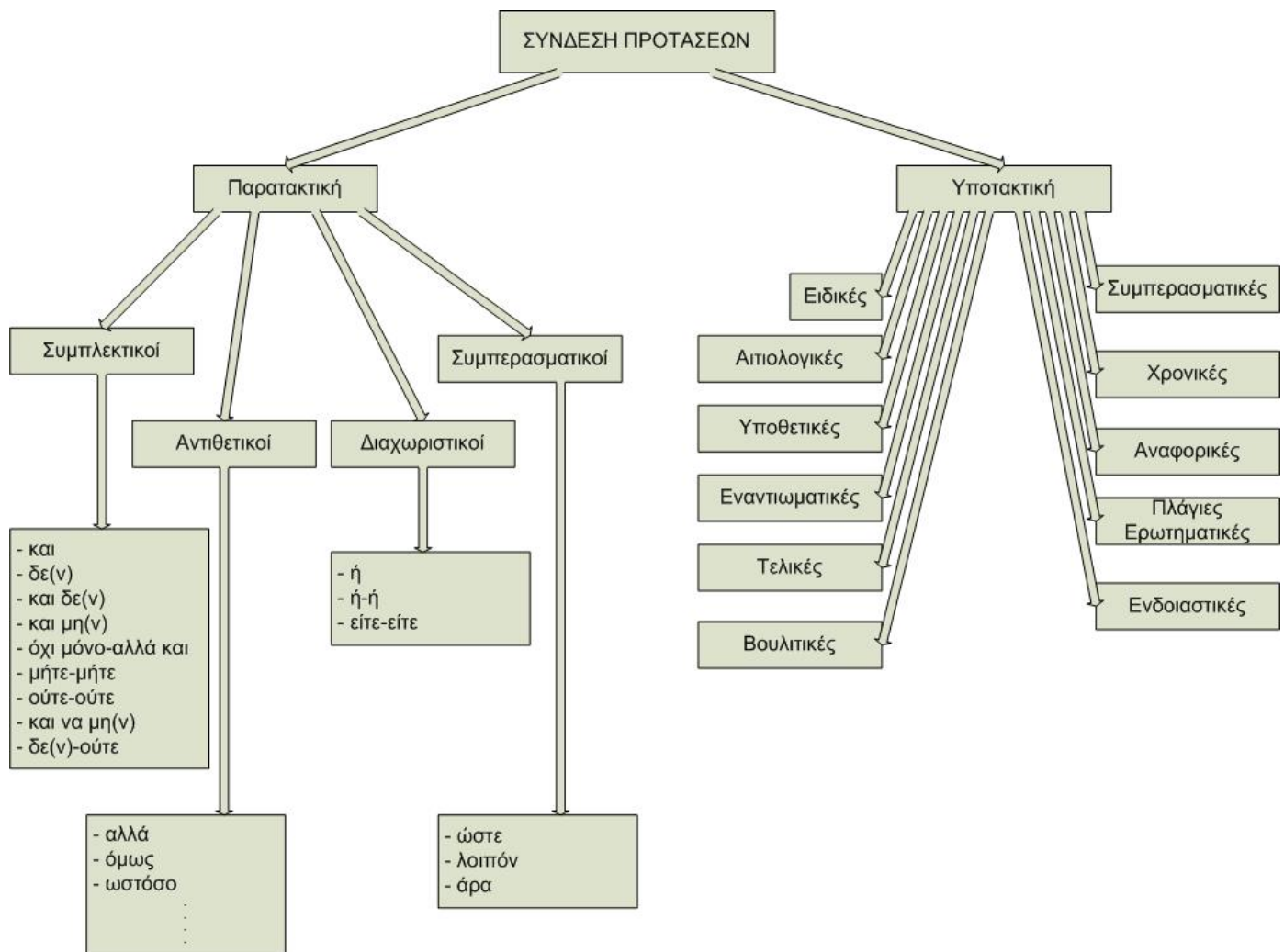
10. Πλάγιες ερωτηματικές

π.χ : «*Δε μου είπε πότε θα έρθει*»

11. Ενδοιαστικές

π.χ : «*Ένα πράγμα φοβάμαι, μήπως δεν έρθεις*»

Σχηματικά η σύνδεση των προτάσεων είναι η ακόλουθη:



Εικόνα 2: Η σύνδεση των προτάσεων

2.2.5 Σχήματα λόγου

Η κανονική θέση των όρων στην πρόταση είναι η ακόλουθη:

A. Υποκείμενο + Ρήμα

π. χ : «Ο Γιώργος έρχεται»

«Ο Νίκος διαβάζει»

B. Υποκείμενο + Ρήμα + Κατηγορούμενο

π. χ: «Ο κόσμος είναι όμορφος»

«Ο Γιάννης σπουδάζει χημικός»

Γ. Υποκείμενο + Ρήμα + Αντικείμενο

π. χ: «Ο Πέτρος αγαπάει τα ζώα»

Δ. Υποκείμενο + Ρήμα + Αντικείμενο(άμεσο) + Αντικείμενο(έμμεσο)

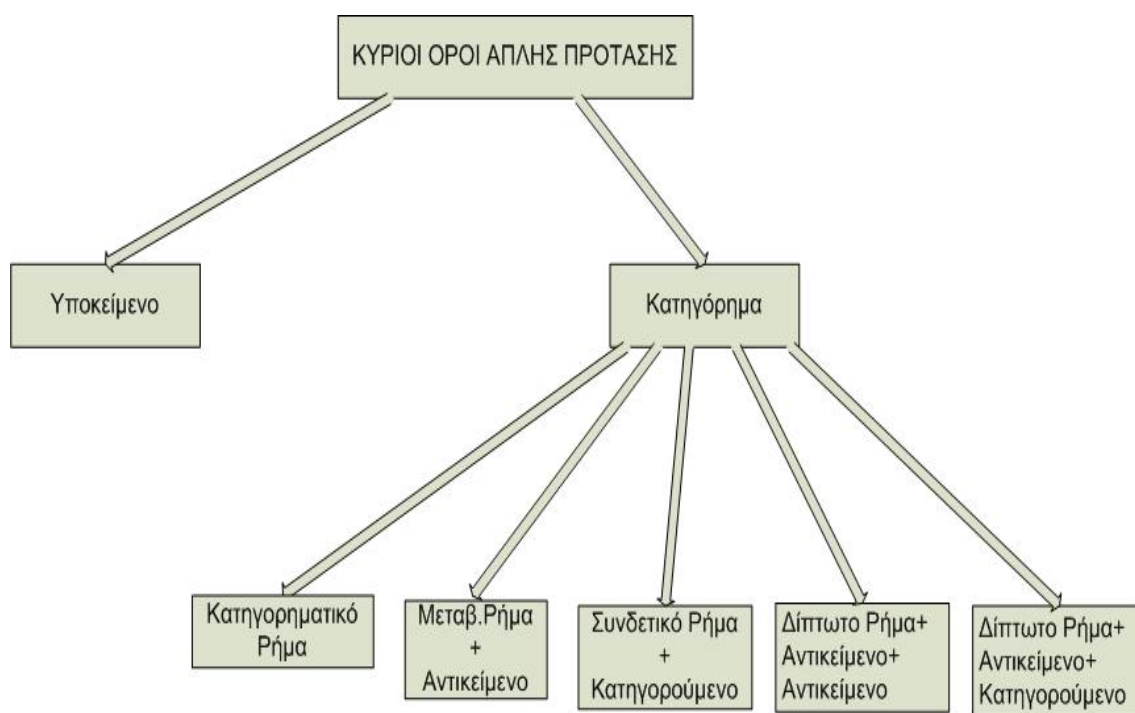
π. χ: «Η γιαγιά χάρισε της Άννας(έμμεσο) ένα βιβλίο(άμεσο) »

«Ο δάσκαλος εξετάζει τον Κώστα(άμεσο) μαθηματικά(έμμεσο)»

Ε. Υποκείμενο + Ρήμα + Αντικείμενο + Κατηγορούμενο

π. χ: «Ο επισκέπτης βρήκε την πόρτα(αντικείμενο) κλειστή(κατηγορούμενο) »

«Θεωρεί τον Κώστα(αντικείμενο) καιροσκόπο(κατηγορούμενο) »



Εικόνα 3: Οι κύριοι όροι απλής πρότασης

Όπως παρατηρούμε παραπάνω, σε μια πρόταση προηγείται το Υποκείμενο, ακολουθεί το Ρήμα και ύστερα από αυτό μπαίνει το Κατηγορούμενο ή το Αντικείμενο.

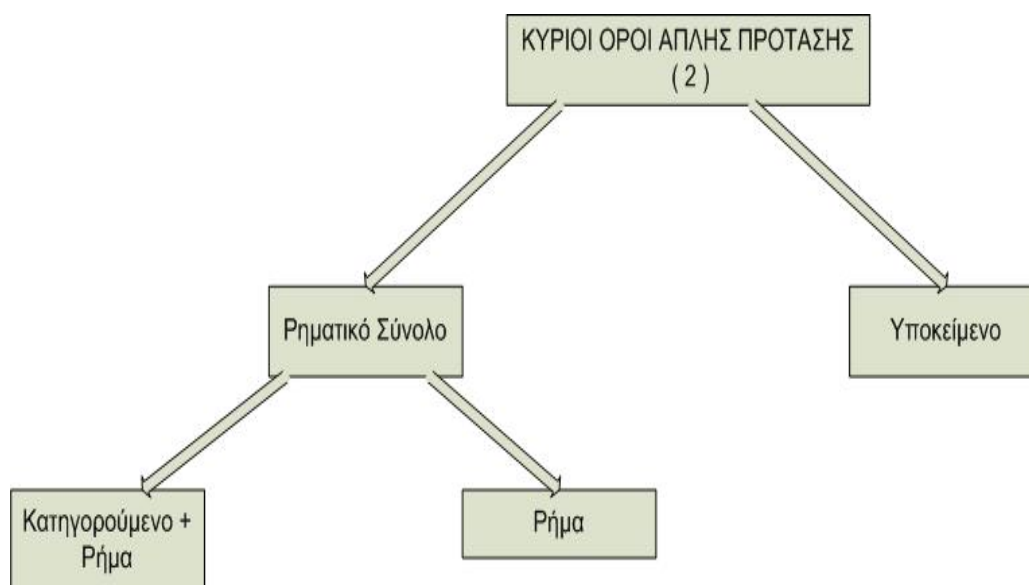
Στις κύριες προτάσεις επιθυμίας, στις επιφωνηματικές και στις ερωτηματικές προτάσεις η σειρά διαταράσσεται:

A. Ρήμα + Υποκείμενο

- a. Επιθυμίας π. χ : «*Ας έρθει ο Γιάννης*»
- b. Επιφωνηματικές π.χ : «*Έφυγε ο καθηγητής!*»
- c. Ερωτηματικές π.χ : «*Τρώει ο Κώστας ;*»

B. Κατηγορούμενο + Ρήμα + Υποκείμενο

- a. Επιθυμίας π. χ : «*Με το καλό να γυρίσετε*»
- b. Επιφωνηματικές π.χ : «*Τι φρικτό που ήταν!*»
- c. Ερωτηματικές π.χ : «*Επιμελητής είναι ο αδερφός σου ;*»



Εικόνα 4: Οι κύριοι όροι απλής πρότασης(2)

2.2.6 Σύνολα λέξεων

Το νόημα ορισμένων λέξεων μέσα στην ίδια πρόταση είναι πιο στενά δεμένο από ότι με το νόημα των υπόλοιπων λέξεων της πρότασης. Οι λέξεις αυτές ,οι οποίες μπορούμε να πούμε, όσον αφορά το νόημά τους, ότι αποτελούν ομάδες μέσα σε μια πρόταση, καλούνται **σύνολα λέξεων**. Για παράδειγμα:

Ο φοιτητής διαβάζει

Η μητέρα εργαζόταν χωρίς σταματημό

Ανάλογα με την συντακτική αντιστοιχία ενός λεκτικού συνόλου, δηλαδή με τι θα μπορούσαμε να το αντικαταστήσουμε, έτσι ώστε να ισχύουν οι συντακτικοί κανόνες μέσα στην πρόταση, μπορούμε να πούμε ότι είναι:

I. **Ονοματικό** σύνολο, όταν η συντακτική του αντιστοιχία είναι **όνομα**:

π.χ.: Ο Νίκος είναι ανήσυχος

II. **Ρηματικό** σύνολο, όταν η συντακτική αντιστοιχία είναι **ρήμα**:

π.χ.: Ο Νίκος είναι ανήσυχος (= ανησυχεί)

III. **Επιρρηματικό** σύνολο, όταν η συντακτική αντιστοιχία είναι **επιρρηματικός προσδιορισμός**:

π.χ.: Η μητέρα εργαζόταν χωρίς σταματημό (= ασταμάτητα)

2.2.6.1 Ονοματικό σύνολο

Το **ονοματικό σύνολο** αποτελείται από ένα όνομα μπροστά στο οποίο βρίσκεται άρθρο ή επίθετο ή αντωνυμία.

Παραδείγματα	Αποτελούνται από
<u>Η μαθήτρια</u> μελετά	άρθρο (οριστικό) + ουσιαστικό
Φυσούσε <u>έναν</u> <u>αέρα</u> δυνατός	άρθρο (αόριστο) + ουσιαστικό
Ήταν <u>εξαιρετικός</u> <u>γιατρός</u>	επίθετο + ουσιαστικό
Εκεί έχει <u>στεκούμενα</u> <u>νερά</u>	μετοχή + ουσιαστικό
Θέλω <u>εκείνο</u> <u>το βιβλίο</u>	αντωνυμία + σύνναρθρο ουσιαστικό
<u>Της νύχτας</u> <u>τα καμώματα</u>	σύνναρθρο ουσιαστικό + σύνναρθρο ουσιαστικό

Πίνακας 1: Οι περιπτώσεις του ονοματικού συνόλου

Κάποτε είναι δυνατόν δύο ονοματικά σύνολα να αποτελούν ένα ευρύτερο ονοματικό σύνολο.

Στη συνέχεια θα δούμε πώς δένεται ένα ουσιαστικό(το απλούστερο ονοματικό σύνολο) σε λεκτικό σύνολο μαζί με:

i. Άρθρο

ii. Ουσιαστικό

iii. Επίθετο

i. *Ουσιαστικό με άρθρο:*

Το άρθρο συμφωνεί με το ουσιαστικό στο γένος , στον αριθμό και στην πτώση. Μπορεί να είναι οριστικό (παρ. β) ή αόριστο (παρ. α)

α. Της πρόσφερε **μια ανθοδέσμη**

β. Δώσε μου **το πράσινο βιβλίο**

ii. *Ουσιαστικό με ουσιαστικό:*

Σε αυτή την περίπτωση μπορούμε να έχουμε είτε **ομοιόπτωτο** είτε **ετερόπτωτο** προσδιορισμό

α. Ομοιόπτωτος προσδιορισμός:

Σε αυτού του τύπου τον προσδιορισμό, τα δύο ουσιαστικά βρίσκονται στην ίδια πτώση και μπορούμε να έχουμε

Παράθεση: Κάποια ουσιαστικά συνοδεύουν κάποια άλλα, προσδίδοντάς τους μια επιπλέον χαρακτηριστική ιδιότητα (η παράθεση αναλύεται σε αναφορική πρόταση), π.χ.:

Ο Κώστας , ο αδερφός μου, έφυγε (=που είναι αδερφός μου)

Επεξήγηση: Χρησιμοποιείται για να επεξηγήσει περισσότερο το ουσιαστικό που προσδιορίζει. Ακολουθεί ή ακολουθείται από τη λέξη δηλαδή, η οποία υπάρχει μέσα στην πρόταση ή μπορεί και παραλείπεται:

Συνάντησα ένα φίλο, τον Αντώνη (= δηλαδή τον Αντώνη)

Στην παράθεση από μια λέξη με έννοια μερικότερη πηγαίνουμε σε μια άλλη με έννοια γενικότερη, που χαρακτηρίζει την πρώτη:

Η Αφροδίτη, η θεά της ομορφιάς, βγήκε από τη θάλασσα

Στην επεξήγηση, από μια λέξη με έννοια γενικότερη πηγαίνουμε σε μια άλλη με έννοια μερικότερη, που επεξηγεί την πρώτη:

Από τη θάλασσα βγήκε η θεά της ομορφιάς, η Αφροδίτη

β. Ετερόπλωτος προσδιορισμός:

Οι ετερόπλωτοι προσδιορισμοί μπορούν να βρίσκονται σε πτώση **γενική ή αιτιατική**

Θα μάθεις δύο κεφάλαια **του βιβλίου (γενική)**

Φοράει παπούτσια με τακούνι **οκτώ πόντους (αιτιατική)**

iii. *Ουσιαστικό με επίθετο*

Σε αυτή τη σύνδεση διακρίνουμε δύο περιπτώσεις:

α. Το επίθετο να είναι ομοιόπλωτος προσδιορισμός ουσιαστικού, αναλυτικότερα:

Επιθετικός προσδιορισμός: Τα επίθετα προσδιορίζουν ουσιαστικά με τα οποία **συμφωνούν στην πτώση, στο γένος και στον αριθμό** και φανερώνουν μια μόνιμη ιδιότητά τους. Μεταξύ των επιθετικών προσδιορισμών **βάζουμε κόμμα** ενώ οι επιθετικοί προσδιορισμοί **συνδέονται μεταξύ τους με συνδέσμους**.

Μια όμορφη , ψηλή και ξανθιά κοπέλα πέρασε μπροστά μας

Κατηγορηματικός προσδιορισμός: Τα επίθετα (και επιθετικές μετοχές) προσδιορίζουν ουσιαστικά με τα οποία **συμφωνούν στην πτώση, στο γένος και στον αριθμό** και δίνουν σ' αυτά μια ιδιότητα παροδική. Οι κατηγορηματικοί προσδιορισμοί **δε συνοδεύονται ποτέ από άρθρο** ενώ **τα ουσιαστικά που προσδιορίζουν είναι πάντα έναρθρα**.

Συνέχισε να τρέχει με τα χέρια δεμένα

β. Το ουσιαστικό να είναι ετερόπτωτος προσδιορισμός επιθέτου, δηλαδή:

Σε πτώση γενική: Ο προσδιορισμός βρίσκεται σε πτώση γενική:

Στάθηκες αντάξιος της εμπιστοσύνης μου

Σε πτώση αιτιατική: Ο προσδιορισμός βρίσκεται σε πτώση αιτιατική:

Ο Νίκος είναι δέκα πόντους ψηλότερους από εμένα

2.2.6.2 Ρηματικό σύνολο

Το σύνολο που αποτελείται από έναν περιφραστικό τύπο του ρήματος ή από ένα ρήμα , μονολεκτικό ή περιφραστικό, με το συμπλήρωμά του (όνομα , επίρρημα κ.ά.) ονομάζεται **ρηματικό σύνολο**.

Παραδείγματα	Αποτελούνται από
Ο ήλιος <u>φώτισε τη γη</u>	ρήμα + ουσιαστικό
Ο ουρανός <u>έγινε σκοτεινός</u>	ρήμα + επίθετο
<u>Τον αγαπούσαν</u> όλοι	αντωνυμία + ρήμα
Ο κόσμος <u>έχει αλλάξει</u>	ρήμα + απαρέμφατο
Ο γέρος <u>κοίταζε ήρεμα</u>	ρήμα + επίρρημα
<u>Έφυγε για το χωριό</u>	ρήμα + προθετικό σύνολο
<u>Όταν έρθεις</u> , θα μιλήσουμε	σύνδεσμος + ρήμα

Όταν έρθεις, <u>θα μιλήσουμε</u>	μόριο + ρήμα
----------------------------------	--------------

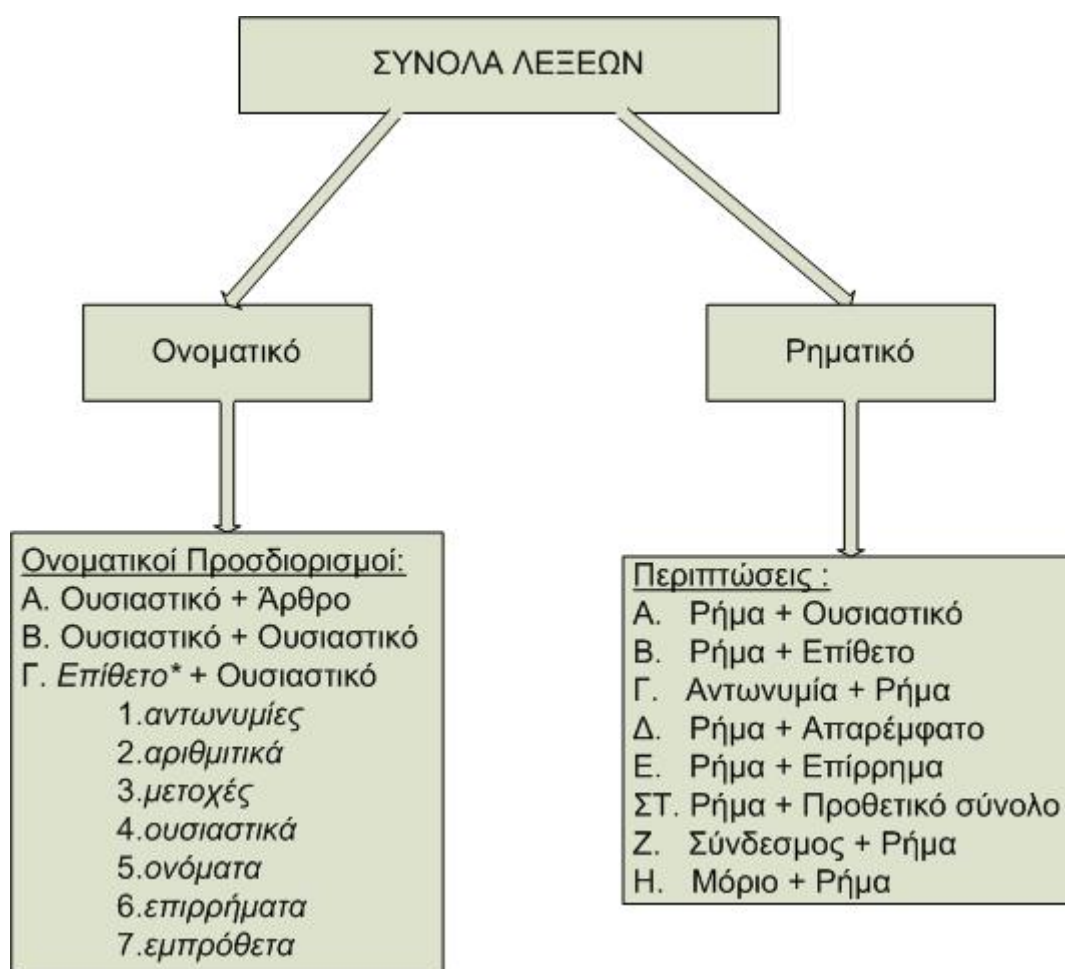
Πίνακας 2: Οι περιπτώσεις του ρηματικού συνόλου

Οι λέξεις που αποτελούν αυτά τα ρηματικά σύνολα δέρονται νοηματικά με το ρήμα στενότερα από ότι με άλλες λέξεις. Κάποτε μάλιστα δέρονται τόσο στενά, ώστε το νόημά τους να μπορεί να αποδοθεί με ένα μόνο ρήμα:

Ο ουρανός έγινε σκοτεινός (= σκοτείνιασε)

Γι' αυτό λέμε ότι το ρηματικό σύνολο αντιστοιχεί **συντακτικά** με το ρήμα

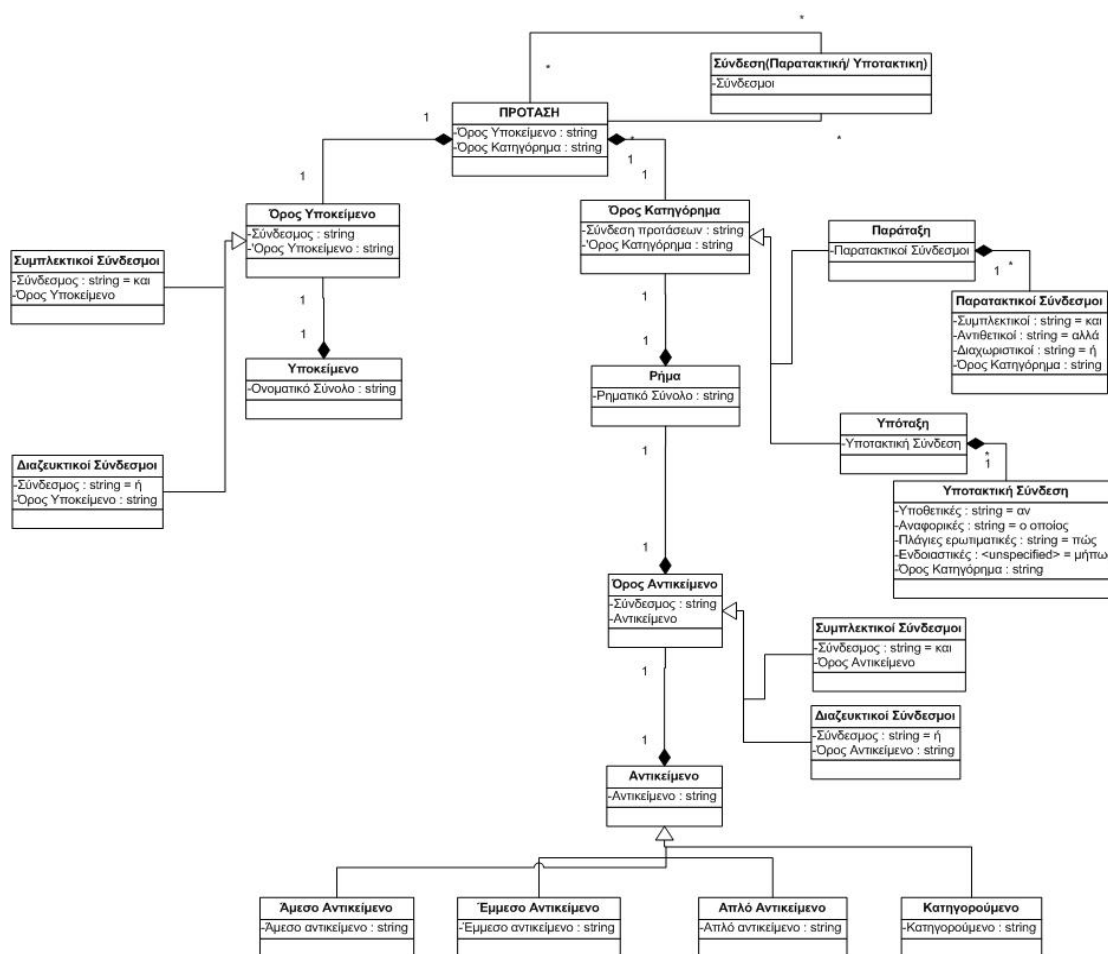
Σχηματικά όλα τα παραπάνω:



Εικόνα 5: Τα σύνολα των λέξεων

2.2.7 Γλωσσικό μοντέλο

Τώρα που παρουσιάσαμε το συντακτικό της ελληνικής γλώσσας θεωρητικά και πριν περάσουμε στην επόμενη παράγραφο, θα δώσουμε το γλωσσικό μοντέλο που προκύπτει για την ελληνική γλώσσα. Το μοντέλο είναι το ακόλουθο:



Εικόνα 6: Το γλωσσικό μοντέλο που προκύπτει από τη συντακτική ανάλυση

Το μοντέλο αυτό είναι σύμφωνο με τη θεωρία που αναπτύχθηκε παραπάνω και αναπαριστά τους βασικούς συντακτικούς όρους της ελληνικής γλώσσας, πώς αυτοί ενώνονται μεταξύ τους αλλά και πώς συνθέτουν μία έκφραση. Οι συντακτικοί αυτοί όροι ξεκινούν από γενικοί (**πρόταση**) και, όπως φαίνεται στο μοντέλο, αναλύονται σε ειδικότερους (**υποκείμενο** / **ρήμα** / **αντικείμενο**) περνώντας από διάφορα στάδια συντακτικής ανάλυσης με τους αντίστοιχους συντακτικούς όρους (π.χ. η **πρόταση** αναλύεται σε **όρο υποκείμενο** και **όρο κατηγορημα**).

Επίσης, το μοντέλο της Εικόνας 6 απευθύνεται σε ένα σύνολο ακολουθιών-προτάσεων και όχι σε μία πρόταση μόνο, μέσω της μονάδας **Σύνδεση (Παρατάξη / Υπόταξη)**. Υποστηρίζει δηλαδή τη *σύνδεση προτάσεων* της ελληνικής γλώσσας (παρ. 2.2.4). Η εμφάνιση στο μοντέλο των μονάδων **Όρος Υποκείμενο** και **Όρος**

Κατηγορήμα φανερώνει πως αυτό τηρεί τα *σχήματα λόγου* (παρ.2.2.5) . Επίσης, με τις μονάδες **Παράταξη** και **Υπόταξη** λαμβάνει υπόψη του τα διάφορα *είδη των προτάσεων* (παρ. 2.2.3). Αυτά κάνουν το παραπάνω μοντέλο πλήρες και αξιόπιστο.

2.3 Κανονικές εκφράσεις

Οι κανονικές εκφράσεις (regular expressions, regexp ή regex) χρησιμοποιούνται για την περιγραφή γλωσσών (ο όρος γλώσσα εδώ χρησιμοποιείται όπως χρησιμοποιείται στα μαθηματικά, τη λογική και την επιστήμη υπολογιστών) με απλά σύμβολα, το $\emptyset$ και συνδυασμούς που προκύπτουν με εφαρμογή ένωσης ($\cup$), του αστεριού Κλήνυ (Kleene Star) ($*$) ή και παρενθέσεων. Οι κανονικές εκφράσεις αποτελούν δομικό στοιχείο του κώδικα της εφαρμογής μας και στις επόμενες παραγράφους θα προσπαθήσουμε να παρουσιάσουμε εκείνα τα χαρακτηριστικά τους που χρησιμοποιήθηκαν στην εφαρμογή και είναι απαραίτητα για την καλύτερη κατανόηση του τρόπου λειτουργίας της εφαρμογής.

2.3.1 Ορισμός κανονικών εκφράσεων

Κανονικές εκφράσεις επί του Σ^* ορίζονται ως όλες οι συμβολοσειρές (strings) επί του $\Sigma \cup \{ (,), *, \emptyset \}$ που σχηματίζονται ακολούθως:

Το κενό και κάθε στοιχείο του Σ είναι κανονική έκφραση.

Αν a και b είναι κανονικές εκφράσεις τότε και η συναλύσωσή τους (concatenation), ab , είναι κανονική έκφραση.

Αν a και b είναι κανονικές εκφράσεις τότε και η ένωσή τους (union), $a \cup b$, είναι κανονική έκφραση.

Αν a είναι κανονική έκφραση τότε και η a^* είναι κανονική έκφραση.

Καμία άλλη στοιχειοσειρά δεν είναι κανονική έκφραση εκτός αν ικανοποιεί τους κανόνες 1 έως 4.

όπου

Σ το αλφάβητο,

Σ^* το σύνολο των συμβολοσειρών επί του αλφαβήτου Σ .

$\emptyset$ το κενό σύνολο,

* το αστέρι Κλήνυ (Kleene Star),

Υη πράξη της ένωσης.

2.3.2 Η γλώσσα Tcl

Η Tcl (συντομογραφία του “Tool Command Language”) είναι μια γλώσσα προγραμματισμού που βοηθάει στην επέκταση των εφαρμογών. Τα τελευταία χρόνια, ανάμεσα στις διάφορες εφαρμογές που χρησιμοποιείται είναι και εφαρμογές φυσικής γλώσσας. Αυτό γιατί είναι μια απλή γλώσσα για κατασκευή και επέκταση εφαρμογών ενώ παράλληλα περιέχει όλες τις βασικές προγραμματιστικές λειτουργίες όπως μεταβλητές , πίνακες , δομές ελέγχου , μαθηματικές εκφράσεις κ.α.

Αυτό όμως που έκανε την Tcl σύντομα γνωστή και έκανε και εμάς να την χρησιμοποιήσουμε στην εφαρμογή μας , είναι τα πλεονεκτήματά της:

α) Με την χρήση της TCL μπορούμε να κατασκευάσουμε πολύ πιο εύκολα , γρήγορα και με λιγότερες γραμμές κώδικα εφαρμογές γραφικού περιβάλλοντος σε σχέση με άλλες γλώσσες προγραμματισμού, όπως με την C

β) Ο προγραμματιστής μπορεί να επεκτείνει εύκολα τις εφαρμογές του χρησιμοποιώντας ελάχιστες εντολές.

γ) Το TCL μπορεί πολύ εύκολα να ενσωματωθεί σε υπάρχουσες εφαρμογές

δ) Η TCL χρησιμοποιεί διερμηνέα και όχι μεταγλωττιστή, έτσι οι εφαρμογές μπορούν να τρέξουν σε διάφορα λειτουργικά συστήματα και αρχιτεκτονικές.

ε) Η TCL μπορεί εύκολα να ενσωματώσει και άλλες βιβλιοθήκες όπως το TK⁹.

Η TCL παρέχει τις εντολές και δομές ώστε να μπορέσουμε να γράψουμε διάφορα είδη εφαρμογών. Μας παρέχει : Μεταβλητές , Τελεστές , Πίνακες , Λίστες , Δομές ελέγχου , Αρχεία , Συναρτήσεις , Εκφράσεις , Διαδικασίες και χειρισμό λαθών.

⁹ Το TK είναι ένα «κουτί» εργαλείων που χρησιμοποιείται μαζί με την TCL για την κατασκευή user interfaces. Το TK είναι η πιο γνωστή και χρήσιμη επέκτασή του TCL

Όλα τα παραπάνω έκαναν την Tcl ιδανική γλώσσα για να μπορέσουμε να εκφράσουμε τις κανονικές εκφράσεις που προέκυψαν από την εξαγωγή των συντακτικών κανόνων. Έτσι, οι συντακτικοί και γραμματικοί κανόνες είναι εκφρασμένοι, με την μορφή κανονικών εκφράσεων, στην γλώσσα Tcl, κερδίζοντας όλα τα παραπάνω πλεονεκτήματα μα κυρίως τα δύο τελευταία, την ανεξαρτησία τους από το σύνολο της εφαρμογής(δ) καθώς και την δυνατότητα επέκτασής τους.

2.4 Μορφοσυντακτική ανάλυσή

Στις ακόλουθες παραγράφους θα περιγράψουμε την μορφοσυντακτική ανάλυση που χρησιμοποιείται στην εφαρμογή μας. Αυτή η ανάλυση γίνεται με τη βοήθεια ενός μορφοσυντακτικού σχολιαστή ο οποίος αναπτύχθηκε το 2000 στο Πανεπιστήμιο Πατρών από τον κ.Ορφανό Γιώργο στα πλαίσια της διδακτορικής του διατριβής με θέμα «Υπολογιστική Μορφοσυντακτική Ανάλυση της Νέας Ελληνικής».

2.4.1 Γενικά

Η παρούσα ενότητα πραγματεύεται την υπολογιστική *μορφοσυντακτική ανάλυση* της Νέας Ελληνικής. Η μορφοσυντακτική ανάλυση είναι μια γνωστική διεργασία που κινείται μεταξύ μορφολογίας και σύνταξης, με στόχο την απόδοση μονοσήμαντης *μορφοσυντακτικής πληροφορίας* στις λέξεις ενός κειμένου. Με τον όρο *μορφοσυντακτική πληροφορία* για μια λέξη εννοούμε τη μορφολογική της προέλευση και τις μορφοσυντακτικές της ιδιότητες (π.χ. η λέξη *ανθρώπου* είναι γενική ενικού του αρσενικού ουσιαστικού [*άνθρωπος*]). Η μορφοσυντακτική ανάλυση έρχεται να καλύψει το κενό μεταξύ της μορφολογικής και της συντακτικής ανάλυσης, και πιο συγκεκριμένα να επιλύσει τη *μορφοσυντακτική ασάφεια* που εισάγεται από τη μορφολογική ανάλυση (π.χ. η λέξη *απαντήσεις* προέρχεται είτε από το ρήμα [*απαντώ*] είτε από το ουσιαστικό [*απάντηση*]), απαλλάσσοντας τη συντακτική ανάλυση από την αντιμετώπιση αυτής της ασάφειας.

Μετά από επισκόπηση των μοντέλων μορφοσυντακτικής αποσαφήνισης που παρουσιάζονται στη διεθνή βιβλιογραφία, χρησιμοποιείται ένα μοντέλο

προσαρμοσμένο στην έντονη μορφολογία της Νέας Ελληνικής. Το μοντέλο αυτό αποτελείται από δύο επίπεδα. Το πρώτο επίπεδο ακολουθεί την προσέγγιση της *μηχανικής μάθησης* και επιλύει ένα σημαντικό (το δυσκολότερο) μέρος της ασάφειας με τη βοήθεια *δέντρων απόφασης*, τα οποία επάγονται αυτόματα από παραδείγματα αποσαφήνισης. Το δεύτερο επίπεδο ακολουθεί τη *γλωσσολογική* προσέγγιση και επιλύει την υπόλοιπη ασάφεια με τη βοήθεια χειρωνακτικά κωδικοποιημένων *συντακτικών κανόνων*.

Για την επαγωγή των δέντρων απόφασης χρησιμοποιήθηκαν τρεις αλγόριθμοι, οι οποίοι επεκτείνουν το στάνταρτ αλγόριθμο ID3 κατά τρόπο ώστε να ανταποκρίνεται στη γλωσσολογική φύση των δεδομένων εκπαίδευσης. Για την περιγραφή των συντακτικών κανόνων χρησιμοποιήθηκε ο φορμαλισμός των *definite-clause γραμματικών*. Η επιτυχία του μοντέλου επιβεβαιώνεται από τα πειραματικά αποτελέσματα της εφαρμογής του πάνω σε σώμα νεοελληνικών κειμένων.

Το αποτέλεσμα αυτής της έρευνας βρήκε εφαρμογή στην κατασκευή ενός *μορφοσυντακτικού σχολιαστή*, βασική προδιαγραφή του οποίου ήταν η *ευρωστία*: η ικανότητα δηλ. να αναλύει οποιοδήποτε νεοελληνικό κείμενο, ανεξαρτήτως περιεχομένου. Η σπουδαιότητα του μορφοσυντακτικού σχολιαστή έγκειται στο ότι, εκτός από αυτόνομο εργαλείο γλωσσικής ανάλυσης, συνιστά το πρώτο επίπεδο επεξεργασίας σε ευρύτερο πεδίο εφαρμογών, όπως είναι η συγκρότηση μορφοσυντακτικά σχολιασμένων σωμάτων κειμένων, η συντακτική ανάλυση, ο εντοπισμός και η διόρθωση γλωσσικών λαθών, η σημασιολογική αποσαφήνιση, η ανάκτηση πληροφορίας, η εξόρυξη πληροφορίας, η εξαγωγή περιλήψεων, η θεματική κατηγοριοποίηση κειμένων κτλ.

2.4.2 Μορφοσυντακτική ανάλυση

Ένα ηλεκτρονικό κείμενο, ως ακολουθία ASCII χαρακτήρων, περιέχει μόνο τους (ορθο)γραφικούς τύπους των λέξεων, δηλ. τις ακολουθίες γραμμάτων από τις οποίες αποτελούνται οι λέξεις. Μορφοσυντακτική ανάλυση είναι η αντιστοίχιση των γραφικών τύπων των λέξεων σε μορφοσυντακτικές λέξεις. Μορφοσυντακτική λέξη (*morphosyntactic word*, βλ. [Lyons 1977, σελ.377][9]) είναι ένα ζεύγος της μορφής [λέξη_{μα}4 : μορφοσυντακτικά χαρακτηριστικά₅], από το οποίο προσδιορίζεται πλήρως μια λέξη σε μορφοσυντακτικό επίπεδο, π.χ. η μορφοσυντακτική λέξη για τον

τύπο αγαπά είναι [αγαπώ: ACT+A_P+IND+PRES+SING+SUBJ+V]¹⁰. Για παράδειγμα στην πρόταση:

Ο Βάκης αισθάνθηκε το αίμα να χτυπάει με ορμή στις φλέβες του, μυρμηγκιάζοντας στις κλειδώσεις.

η μορφοσυντακτική της ανάλυση θα μας δώσει:

Ο/[ο: **ACC+ART+MASC+SING**]
Βάκης/[Βακης: **ACC+N+ MASC +NOM+SING+VOC**]
αισθάνθηκε/[αισθάνομαι: **AOR+C_P+IND+PASS+SING+V**]
το/[ο: **ACC+ART+NEUT+NOM+SING**]
αίμα/[αίμα: **ACC+N+NEUT+NOM+SING+VOC**]
να/[να: **PARTICLE**]
χτυπάει/[χτυπώ: **ACT+C_P+IND+INFORMAL+PRES+SING+SUBJ+V**]
με/[με: **PREP**]
ορμή/[ορμή: **ACC+FEM+N+NOM+SING+VOC**]
στις/[στον: **ACC+ART+FEM+PLUR**]
φλέβες/[φλέβα: **ACC+FEM+N+NOM+PLUR+VOC**]
του/[μου: **ART+GEN+MASC+NEUT+SING**]
./Κόμμα
μυρμηγκιάζοντας/[μυρμηγκιάζω: **ACT+PART+PRES+V**]
στις/[στον: **ACC+ART+FEM+PLUR**]
κλειδώσεις/[κλείδωση: **ACC+FEM+N+NOM+PLUR+VOC**]
./Τελεία

Στην παραπάνω διαδικασία, η μορφοσυντακτική ανάλυση λαμβάνει ως είσοδο μια ακολουθία ASCII χαρακτήρων, αναγνωρίζει το όρια των προτάσεων και των λέξεων και αντιστοιχίζει τις λέξεις σε μορφοσυντακτικές λέξεις. Στόχος της μορφοσυντακτικής ανάλυσης είναι να αντιστοιχίσει κάθε λέξη σε μία ακριβώς μορφοσυντακτική λέξη. Όμως, κατά κανόνα, δεν υπάρχει 1-1 αντιστοιχία μεταξύ λέξεων και μορφοσυντακτικών λέξεων· αρκετές λέξεις αντιστοιχούν σε περισσότερα

¹⁰ Η ανάλυση των συντομογραφιών των όρων βρίσκεται στο παράρτημα 1

του ενός λεξήματα ή/και είναι ασαφείς ως προς κάποιο(α) από τα μορφοσυντακτικά τους χαρακτηριστικά, ένα φαινόμενο που περιγράφεται με τον όρο *μορφοσυντακτική ασάφεια*. Για παράδειγμα, η λέξη *κλειδώσεις* στο παραπάνω παράδειγμα αντιστοιχεί στις μορφοσυντακτικές λέξεις:

[κλείδωση:Ουσ(ΘηλΠληΟνο)],
[κλείδωση:Ουσ(ΘηλΠληΑιτ)],
[κλείδωση:Ουσ(ΘηλΠληΚλτ)] και
[κλειδώνω:Ρήμ(ΕνεΑόρΥποΕνιΒ')].

Η επιλογή της κατάλληλης μορφοσυντακτικής λέξης για μια λέξη που παρουσιάζει μορφοσυντακτική ασάφεια είναι απόρροια εξέτασης του φραστικού περιβάλλοντος στο οποίο εμφανίζεται η λέξη. Για το λόγο αυτό, η μορφοσυντακτική ανάλυση είναι γνωστή και ως *ρηχή συντακτική ανάλυση*.

2.4.3 Μορφοσυντακτική αποσαφήνιση στη Νέα Ελληνική

Το πρώτο στάδιο για την αντιμετώπιση κάθε προβλήματος —άρα και της μορφοσυντακτικής ασάφειας - είναι η καταγραφή της έκτασης και της πολυπλοκότητάς του. Η καταγραφή της μορφοσυντακτικής ασάφειας της Νέας Ελληνικής [Γάκης κ.ά. 1999][10] έγινε με τη βοήθεια του μορφολογικού λεξικού που αναπτύχθηκε από την Ερευνητική Μονάδα 2 του Ινστιτούτου Τεχνολογίας Υπολογιστών [Tsalidis and Orphanos 1995, Orphanos and Tsalidis 1999][11]. Η πληρότητα του λεξικού (Πίνακας 3) εγγυάται την κάλυψη του φαινομένου της μορφοσυντακτικής ασάφειας σε όλο του το εύρος. Οι ασαφείς τύποι προήλθαν μετά από ανάκληση όλων των λεξικών τύπων με κοινή ορθογραφική αναπαράσταση που επαναλαμβάνονται στο λεξικό. Οι κατηγορίες μορφοσυντακτικής ασάφειας οργανώθηκαν μετά από μελέτη των μορφοσυντακτικών χαρακτηριστικών που διαφοροποιούν τους ασαφείς αυτούς τύπους, με κύριο χαρακτηριστικό κατηγοριοποίησης το ΜτΛ.

Μέρος του Λόγου	πλήθος Τύπων
Ουσιαστικά	153.929
Επίθετα	178.794

Ρήματα	414.044
Μετοχές	127.089
Επιρρήματα	5.949
Υπόλοιπα	669
Σύνολο Τύπων (σε όλα τα Μέρη του	880.474
Σύνολο μοναδικών Τύπων στο λεξικό	873.701
Ασαφείς Τύποι από διαφορετικά	8.543
Ασαφείς Τύποι στο ίδιο λέξημα	58.050
Σύνολο Ασαφών Τύπων	66.593

Πίνακας 3: Στοιχεία για το μορφολογικό λεξικό της EM2 του ΙΤΥ

Ο Πίνακας 3 παρουσιάζει το πλήθος των λεξικών τύπων για το κάθε μέρος του λόγου. Στη μέτρηση συμμετέχουν οι ορθογραφικά διαφορετικοί τύποι για το κάθε μέρος του λόγου, π.χ. το λεξικό έχει 153.929 διαφορετικούς ορθογραφικούς τύπους ουσιαστικών, πράγμα που σημαίνει ότι ένας τύπος που είναι και ουσιαστικό και ρήμα (π.χ. *πράξεις* < [*πράξη*], [*πράττω*]) έχει μετρηθεί και ως ουσιαστικό και ως ρήμα. Στο πλήθος των ασαφών τύπων από διαφορετικά λεξήματα (8.543) έχουν συμπεριληφθεί οι τύποι που ανήκουν ταυτόχρονα σε δύο ή τρία λεξήματα του λεξικού, χωρίς απαραίτητα να διαφοροποιούνται ως προς το ΜτΛ (π.χ. *φακών* < [*φακός*], [*φακή*], [*φάκα*]). Αξίζει επίσης να σημειώσουμε ότι υπάρχει ένα μεγάλο ποσοστό επανάληψης τύπων μέσα στην κλίση του ιδίου λεξήματος (58.050 τύποι συνολικά), οι οποίοι διαφέρουν ως προς τις μορφοσυντακτικές τους ιδιότητες. Για παράδειγμα, ο τύπος *διεθνή* στο λέξημα [*διεθνής*] υπάρχει δύο φορές:

διεθνή – γενική, αιτιατική, κλητική, αρσενικό, θηλυκό, ενικός

διεθνή – ονομαστική, αιτιατική, κλητική, ουδέτερο, πληθυντικός

και όχι μια φορά, γεγονός που θα οδηγούσε στο λανθασμένο χαρακτηρισμό:

διεθνή – ονομαστική, γενική, αιτιατική, κλητική, αρσενικό, θηλυκό, ουδέτερο, ενικός, πληθυντικός

Από τις μετρήσεις που δείχνει ο Πίνακας 3 φαίνεται ότι οι ασαφείς τύποι ανέρχονται σε ποσοστό 7,62% επί των γραφικά μοναδικών τύπων του λεξικού. Το πραγματικό ποσοστό ασαφών τύπων που χρησιμοποιούνται στο λόγο προκύπτει από τη μέτρησή τους σε σώμα κειμένων. Ενδεικτικά αναφέρουμε ότι σε πλήθος 119.420 λέξεων που συλλέχθηκαν από σώμα κειμένων οι 24.565 ήταν ασαφείς ως προς το

ΜτΛ, δηλαδή σε ποσοστό 20,57%. Η σημαντική διαφορά του ποσοστού μορφοσυντακτικής ασάφειας σε σώμα κειμένων από αυτό που μετρήθηκε στο λεξικό οφείλεται στη συχνή εμφάνιση λειτουργικών λέξεων που είναι μορφοσυντακτικά ασαφείς, όπως οι λέξεις *το* (2,1%), *τον* (1,85%), *της* (1,59%), *την* (1,18%), *με* (1,18%) κτλ.

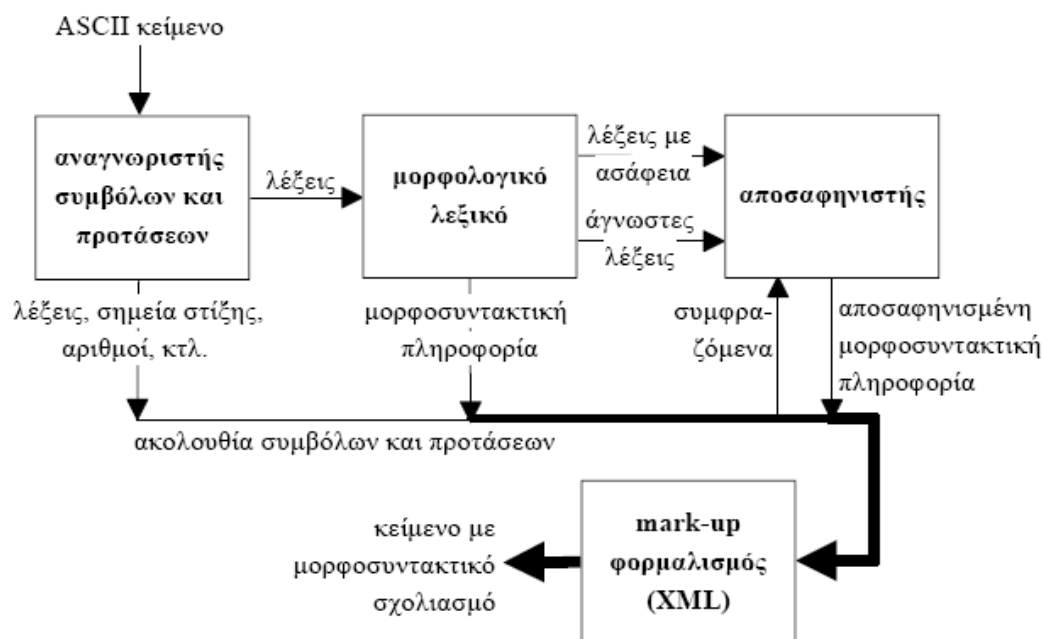
Η μορφοσυντακτική ασάφεια που παρατηρείται στη Νέα Ελληνική είναι δυνατόν να διακριθεί σε:

Προβλέψιμη. Σε αυτή τη μορφή ασάφειας ταυτίζονται ορθογραφικά τύποι με κοινή ετυμολογική προέλευση. Περιλαμβάνει δύο κατηγορίες:

- Στην πρώτη κατηγορία ανήκει η ασάφεια που παρατηρείται στους κλιτικούς τύπους του ιδίου λεξήματος. Ο τύπος *κρίνω*, για παράδειγμα, μπορεί να είναι οριστική ενεστώτα, υποτακτική ενεστώτα, οριστική εξακολουθητικού μέλλοντα, οριστική συνοπτικού μέλλοντα, καθώς και υποτακτική αορίστου. Σε αυτό το επίπεδο ασάφειας ανήκει και η ασάφεια που παρατηρείται μεταξύ της γενικής ενικού του αρσενικού και της γενικής ενικού του ουδετέρου, καθώς και της γενικής πληθυντικού αρσενικού, θηλυκού και ουδετέρου των αντωνυμιών, επιθέτων και μετοχών.
- Στη δεύτερη κατηγορία ανήκει η ασάφεια που παρατηρείται μεταξύ λεξικών τύπων διαφορετικών λεξημάτων με το ίδιο ή διαφορετικό μέρος του λόγου, π.χ. *γλυπτών* < [*γλύπτης*], [*γλυπτός*], πρόσθετε < [*πρόσθετος*], [*προσθέτω*].

Συμπτωματική. Σε αυτή τη μορφή ασάφειας ταυτίζονται ορθογραφικά λέξεις διαφορετικής ετυμολογικής προέλευσης. Για παράδειγμα, ο τύπος *βάλτε* μπορεί να είναι κλητική ενικού του ουσιαστικού [*βάλτος*] ή προστακτική αορίστου του ρήματος [*βάζω*]. Επίσης, υπάρχουν τύποι που είναι αδύνατον να αποδοθούν υπολογιστικά με τη σωστή συντακτική ιδιότητα λόγω απουσίας φωνητικής πληροφορίας, όπως ο τύπος *ήπια*, που μπορεί να είναι α' ενικό οριστικής αορίστου του ρήματος [*πίνω*] ή ονομαστική, αιτιατική και κλητική πληθυντικού ουδετέρου του επιθέτου [*ήπιος*] ή το επίρρημα [*ήπια*] (αν και η ασάφεια μεταξύ επιθέτου-επιρρήματος είναι προβλέψιμη).

2.4.4 Μορφοσυντακτικός σχολιαστής



Εικόνα 7: Αρχιτεκτονική του μορφοσυντακτικού σχολιαστή

Η Εικόνα 6 παρουσιάζει την αρχιτεκτονική του μορφοσυντακτικού σχολιαστή που κατασκευάστηκε από τον κ.Ορφανό Γιώργο στα πλαίσια της διδακτορικής του διατριβής με θέμα «Υπολογιστική Μορφοσυντακτική Ανάλυση της Νέας Ελληνικής».

Ο *αναγνωριστής συμβόλων και προτάσεων* σε πρώτη φάση αναγνωρίζει τα σύμβολα που εμφανίζονται στο πηγαίο κείμενο (λέξεις, σημεία στίξης, αριθμοί, συντμήσεις κτλ.) και σε δεύτερη φάση αναγνωρίζει τα όρια των προτάσεων, μετασχηματίζοντας το πηγαίο κείμενο σε μια ακολουθία συμβόλων μέσα στην οποία έχουν μαρκαριστεί τα όρια των προτάσεων. Όσα σύμβολα αποτελούνται αποκλειστικά από αλφαβητικούς χαρακτήρες (δηλ. οι λέξεις) αναζητούνται στο *μορφολογικό λεξικό*, με τρία πιθανά ενδεχόμενα: α) να τους αποδοθεί από το λεξικό σαφής μορφοσυντακτική πληροφορία, β) να τους αποδοθεί από το λεξικό ασαφής μορφοσυντακτική πληροφορία και γ) να μην τους αποδοθεί καθόλου μορφοσυντακτική πληροφορία, επειδή δεν υπάρχουν στο λεξικό. Ο *αποσαφηνιστής* αναλαμβάνει να αντιμετωπίσει τις περιπτώσεις (β) και (γ). Στην περίπτωση (β) αφαιρεί μορφοσυντακτική πληροφορία που δε συμφωνεί με τα συμφραζόμενα. Στην περίπτωση (γ) παράγει μορφοσυντακτική πληροφορία που συμφωνεί με τα

συμφραζόμενα. Τα αποτελέσματα όλης αυτής της επεξεργασίας ενσωματώνονται στο πηγείο κείμενο υπό τη μορφή σχολίων· τα σχόλια υπακούουν σε συγκεκριμένο *markup* *φορμαλισμό* και δηλώνουν ρητά μέσα στο κείμενο τα όρια των προτάσεων, τα όρια των συμβόλων από τα οποία αποτελούνται οι προτάσεις και τη μορφοσυντακτική πληροφορία που αντιστοιχεί στα σύμβολα.

Όλα τα λειτουργικά μέρη του μορφοσυντακτικού σχολιαστή αναπτύχθηκαν με τη γλώσσα προγραμματισμού C.

Κεφάλαιο 3

Συντακτικός αναλυτής

Ο συντακτικός αναλυτής υλοποιεί την συντακτική ανάλυση μιας συγκεκριμένης γλώσσας, στην παρούσα εργασία της Νέας Ελληνικής. Δέχεται ως είσοδο μια ακολουθία λεκτικών μονάδων, των οποίων έχει προηγηθεί η γραμματική ανάθεση σε όρους (ουσιαστικό / επίθετο /αντωνυμία κ.ά.) και στη συνέχεια ελέγχει εάν η ακολουθία αυτή είναι σύμφωνη με το σύνολο των συντακτικών κανόνων της γλώσσας. Έπειτα αντιστοιχίζει τους όρους αυτούς σε συντακτικούς όρους και έχουμε τη συντακτική ανάλυση της αρχικής ακολουθίας.

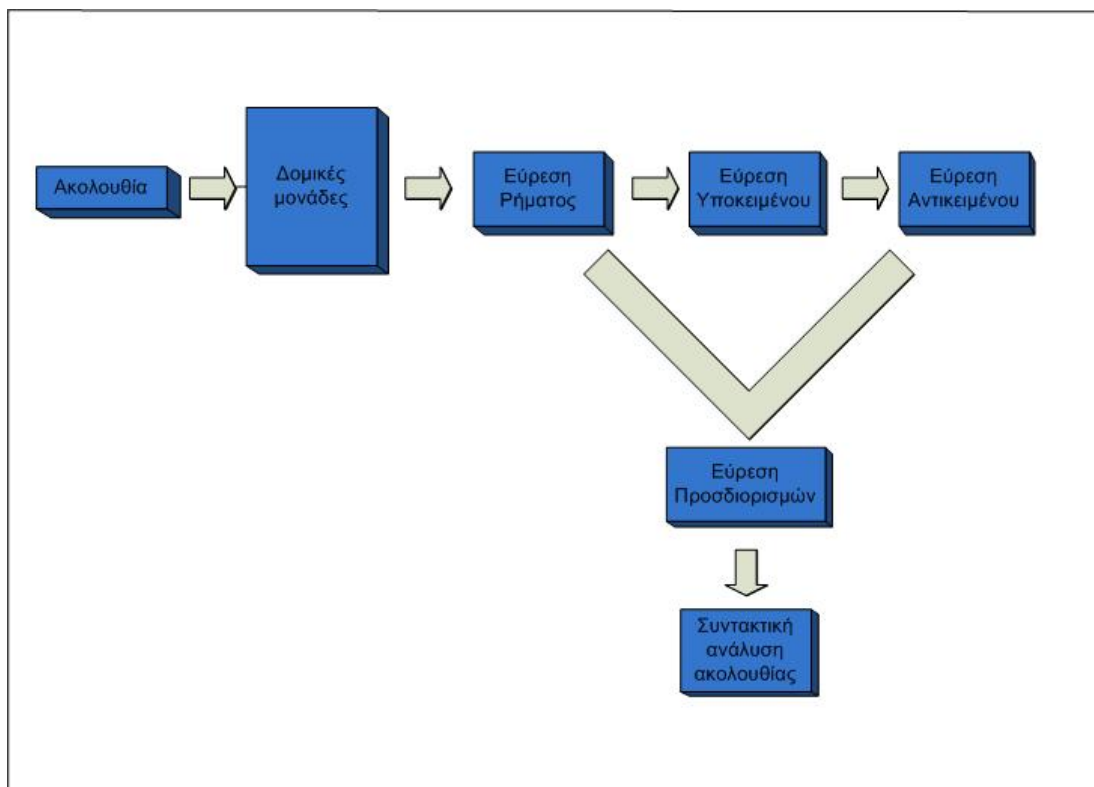
Η αντιστοίχιση των γραμματικών όρων (ουσιαστικό / επίθετο /αντωνυμία κ.ά.) σε συντακτικούς (υποκείμενο / αντικείμενο / κατηγορούμενο κ.ά.) είναι το βασικότερο κομμάτι του συντακτικού αναλυτή. Για να πραγματοποιηθεί αυτή η ανάλυση με επιτυχία, αναπτύξαμε ένα μοντέλο για το συντακτικό αναλυτή.

Έτσι λοιπόν ο αναλυτής παρέχει τις παρακάτω λειτουργίες:

- Διαχωρίζει το εισερχόμενο κείμενο σε λεκτικές μονάδες (*tokens*) και το μεταφέρει με τον τρόπο αυτό στο συντακτικό αναλυτή.
- Αναγνωρίζει λεκτικά λάθη στην είσοδο (π.χ. σύμβολα που δεν επιτρέπονται στη γλώσσα).
- Αφαιρεί τα σχόλια και λεκτικές μονάδες που δεν περιέχουν χρήσιμη συντακτική πληροφορία
- Συσχετίζει αριθμούς γραμμών με στοιχεία της εισόδου του.

Στις επόμενες παραγράφους θα προσπαθήσουμε να παρουσιάσουμε τον τρόπο με τον οποίο γίνεται αυτή η αντιστοίχιση-ανάλυση στο μοντέλο του συντακτικού αναλυτή της Νέας Ελληνικής που κατασκευάσαμε.

3.1 Το μοντέλο του συντακτικού αναλυτή



Εικόνα 8: Το μοντέλο του συντακτικού αναλυτή

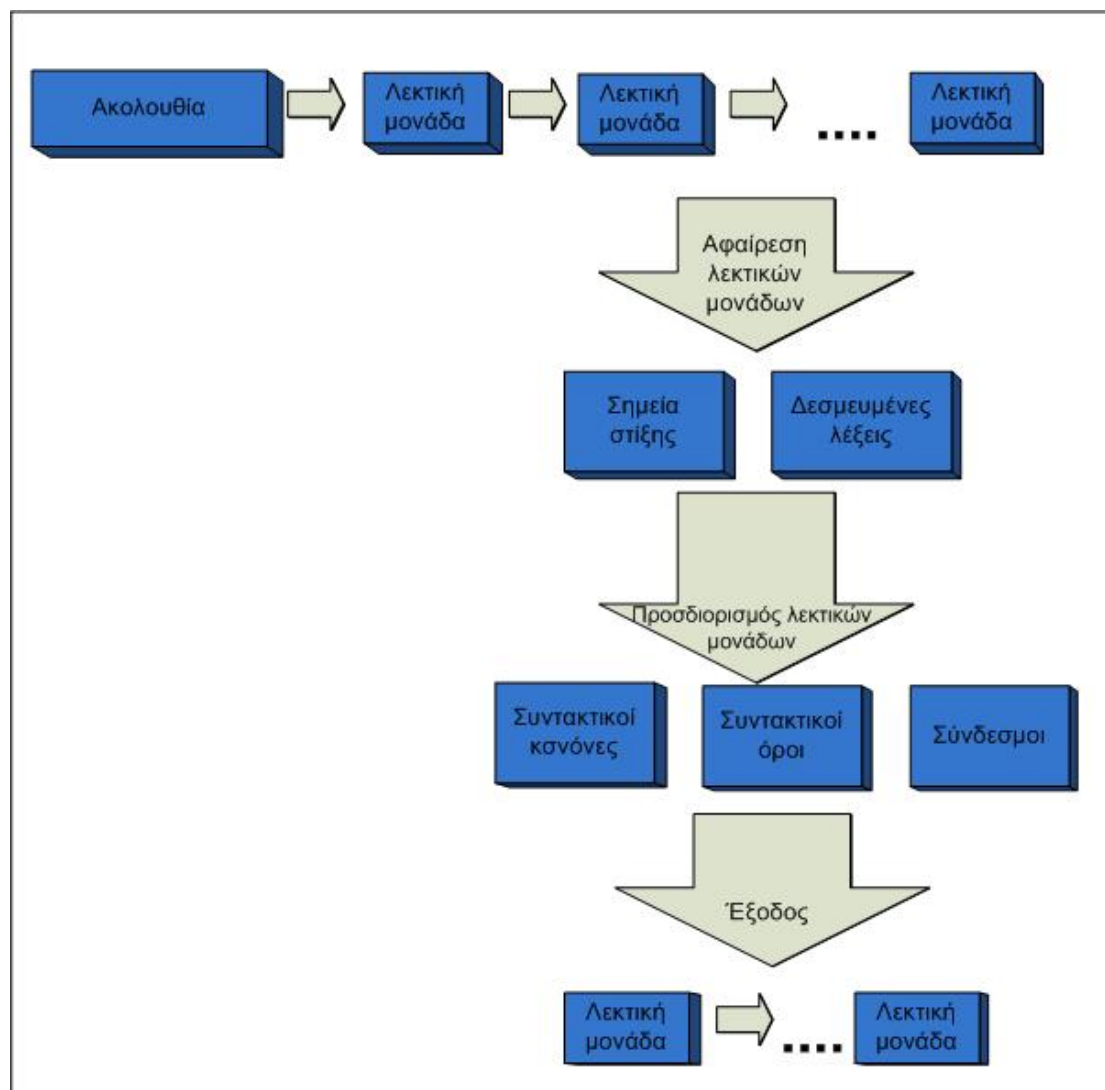
Στην Εικόνα 8, βλέπουμε το μοντέλο του συντακτικού αναλυτή που αναπτύξαμε. Στο μοντέλο αυτό μπορούμε να δούμε τα διάφορα στάδια από τα οποία πρέπει να περάσει μια προς ανάλυση ακολουθία, προκειμένου να γίνει η συντακτική της ανάλυση. Έτσι αρχικά η ακολουθία εισέρχεται σε ένα σύνολο δομικών μονάδων οι οποίες την αναλύουν κατάλληλα. Στη συνέχεια γίνεται η εύρεση του ρήματος, έπειτα του υποκειμένου και τέλος του αντικειμένου. Αφού εντοπίσουμε αυτούς τους βασικούς συντακτικούς όρους, προχωράμε στην εύρεση των προσδιορισμών κάθε όρου και ολοκληρώνεται η συντακτική ανάλυση της ακολουθίας.

Είναι λογικό και αναμενόμενο, για να μπορέσουμε να έχουμε έναν ολοκληρωμένο συντακτικό αναλυτή με ορθά αποτελέσματα, τα στάδια της ανάλυσής του να καλύπτουν όλους τους συντακτικούς όρους που μπορούν να εμφανιστούν στη γλώσσα και οι οποίοι παρουσιάστηκαν παραπάνω στην παράγραφο 2.2. Στο συγκεκριμένο μοντέλο αυτή η βασική προϋπόθεση τηρείται.

Η ανάλυση του μοντέλου θα έχει την ακόλουθη σειρά: Αρχικά θα παρουσιαστούν οι δομικές του μονάδες. Οι μονάδες εκείνες δηλαδή που ουσιαστικά

το συνθέτουν και είναι αναπόσπαστο κομμάτι του. Στη συνέχεια θα παρουσιαστούν το ρήμα, ο όρος υποκείμενο, ο όρος αντικείμενο και οι προσδιορισμοί. Η σειρά είναι έτσι γιατί στη μετάβαση από τη γραμματική ανάλυση στη συντακτική, ο μόνος όρος που μένει αμετάβλητος, γιατί είναι ταυτόχρονα και γραμματικός αλλά και συντακτικός όρος, είναι το ρήμα. Έχοντας αυτό ως βάση και με τη βοήθεια ενός συνόλου κανόνων μπορούμε να προσδιορίσουμε το υποκείμενο το οποίο υπάρχει **απαραίτητα** μέσα σε κάθε πρόταση. Εάν μετά τους δύο προσδιορισμούς υπάρχουν ακόμη λεκτικές μονάδες προς ανάλυση, τότε ελέγχουμε την ύπαρξη αντικειμένου, το οποίο είναι **προαιρετικό**, και στο τέλος και αυτού του ελέγχου, ψάχνουμε για πιθανούς προσδιορισμούς οι οποίοι είναι επίσης **προαιρετικοί**. Έτσι η σειρά της ανάλυσης και ανίχνευσης συντακτικών όρων δεν είναι τυχαία.

3.2 Δομικές μονάδες



Εικόνα 9: Οι δομικές μονάδες του συντακτικού μας αναλυτή

Θεωρήσαμε απαραίτητο, ο συντακτικός μας αναλυτής να αποτελείται από ένα σύνολο δομικών μονάδων. Με τον τρόπο αυτό διαχωρίζονται οι εργασίες της λεκτικής και της καθαρά συντακτικής ανάλυσης και κάθε μια υλοποιείται με τον πιο αποδοτικό τρόπο.

Οι δομικές αυτές μονάδες περιέχουν τις απαραίτητες πληροφορίες για να λειτουργεί κανονικά ο αναλυτής. Λογικό είναι, το σύνολο των δομικών μονάδων να περιέχουν είτε συντακτική πληροφορία (τι μπορεί να είναι υποκείμενο) είτε συντακτικούς όρους (υποκείμενα - αντικείμενα). Στις επόμενες ενότητες θα παρουσιάσουμε αναλυτικά τις μονάδες αυτές καθώς και τι περιέχει η κάθε μία. Για την ομαλότερη και πληρέστερη κατανόησή μας, θα δουλέψουμε πάνω στην ίδια πρόταση για να δούμε πώς αυτή θα περάσει από τις διάφορες μονάδες του αναλυτή:

«Πώς καθαρίσατε τα ρούχα ο Γιάννης και εσύ ; »(1)

3.2.1 Λεκτικές μονάδες

Οι λεκτικές μονάδες είναι το πρώτο κομμάτι του αναλυτή. Τα στοιχεία της γλώσσας χωρίζονται σε λεκτικές μονάδες. Μέσα σε αυτές τις μονάδες βρίσκονται οι λέξεις ή τα σύμβολα που είναι προς ανάλυση. Απαραίτητα μαζί με την κάθε λέξη βρίσκεται η γραμματική της ανάλυση (ουσιαστικό / επίθετο κ.ά.) όπως αυτή έχει προκύψει από το μορφοσυντακτικό σχολιαστή.

Στις λεκτικές μονάδες που δημιουργούμε στο μοντέλο μας, συναντάμε και επιπλέον πληροφορία που αφορά τη λέξη, όπως τον αριθμό, την πτώση και το γένος της για ουσιαστικά ή επίθετα, τον χρόνο, την έγκλιση, τη φωνή, τον αριθμό και το πρόσωπο για ρήματα. Αυτή η πληροφορία θα χρησιμοποιηθεί κατάλληλα στη συνέχεια από τους συντακτικούς κανόνες. Για παράδειγμα, γνωρίζοντας την πτώση μίας λεκτικής μονάδας μπορούμε να αποφανθούμε εάν αυτή μπορεί να αποτελέσει ή όχι το υποκείμενο μιας προς ανάλυση ακολουθίας.

Επίσης, ο γραμματικός σχολιαστής μας δίνει και τη ρίζα της λέξης, κάτι πολύ χρήσιμο σε περιπτώσεις όπου αναζητάμε λέξεις και βασιζόμαστε στη ρίζα τους για να αποφύγουμε την πολυπλοκότητα που ίσως εισάγουν οι παραπάνω πληροφορίες. Για παράδειγμα το ρήμα «ήταν» είναι ευκολότερο να βρεθεί σε μία αναζήτηση του ρήματος «είναι» όταν γνωρίζεις ότι και τα δύο ρήματα έχουν την ίδια ρίζα, το ρήμα «είμαι».

Οι λεκτικές μονάδες είναι απαραίτητες στο μοντέλο μας. Ουσιαστικά αποτελούν τον κερματισμό της ακολουθίας εισόδου και ένα πρώτο (γραμματικό) σχολιασμό του αποτελέσματος. Έτσι μετά το πέρασμα της ακολουθίας από αυτή τη δομική μονάδα, δε χρειάζεται να δουλεύουμε με το σύνολο της ακολουθίας εισόδου αλλά με λεκτικές μονάδες.

Στην πρότασή μας (1), συναντάμε επτά λεκτικές μονάδες:

Πώς – επίρρημα

καθαρίσατε – ρήμα / χρόνος:αόριστος, έγκλιση:οριστική, φωνή:ενεργητική, αριθμός:πληθυντικός, πρόσωπο: τρίτο, ρίζα «καθαρίζω»

τα – άρθρο / αριθμός: πληθυντικός, γένος: ουδέτερο, πτώση:αιτιατική

ρούχα – ουσιαστικό / αριθμός: πληθυντικός, γένος: ουδέτερο, πτώση:αιτιατική

ο – άρθρο / αριθμός: ενικός, γένος: αρσενικό, πτώση:ονομαστική

Γιάννης – ουσιαστικό / αριθμός: ενικός, γένος: αρσενικό, πτώση:ονομαστική

και – σύνδεσμος

εσύ – αντωνυμία / αριθμός: ενικός, πρόσωπο: δεύτερο

; - σημείο στίξης

Αυτές οι λεκτικές μονάδες αποτελούν τα στοιχεία που θα αναλυθούν και θα προωθηθούν στις υπόλοιπες μονάδες για να συνεχιστεί η ανάλυση.

3.2.2 Σημεία στίξης

Όταν μιλάμε ,σταματάμε κάθε τόσο, άλλοτε λιγότερα και άλλοτε περισσότερο, γιατί το θέλει το νόημα ή για να πάρουμε αναπνοή. Άλλοτε πάλι ανεβοκατεβάζουμε τη φωνή μας για να ρωτήσουμε , να δείξουμε την απορία μας , τη χαρά μας ή να εκφράσουμε ότι άλλο αισθανόμαστε.

Είναι βέβαια αδύνατο με το γράψιμο να παραστήσουμε αυτά που δείχνει η φωνή μας. Μεταχειριζόμαστε όμως μερικά σημάδια, που μας δείχνουν που πρέπει να σταματήσουμε και πόσο κάθε φορά, και πώς να χρωματίσουμε τη φωνή μας. Αν έλειπαν αυτά ,θα ήταν δύσκολο να καταλάβουμε το νόημα διαβάζοντας. Τα σημάδια αυτά τα ονομάζουμε **σημεία στίξης**.

Τα συχνότερα σημεία στίξης είναι η **τελεία (.)**, η **άνω τελεία (⋅)**, το **κόμμα (,)** το **ερωτηματικό (;)** και το **θαυμαστικό (!)** ενώ υπάρχουν και άλλα όπως οι **δύο τελείες (:)**, η **παύλα (–)**, οι **αγκύλες []**, τα **ομοιωματικά (»)** κ.ά.

Τα σημεία στίξης παίζουν σημαντικό ρόλο σε κάθε συντακτικό αναλυτή. Πρέπει ο αναλυτής να είναι σε θέση να τα αναγνωρίζει ως σημεία στίξης και να τα αφαιρεί από τη λίστα των λεκτικών μονάδων. Επιπλέον μπορεί να αντλήσει σημαντικές πληροφορίες από αυτά.

Το μοντέλο του συντακτικού αναλυτή που κατασκευάσαμε χρησιμοποιεί τα σημεία στίξης και για να αφαιρέσει λεκτικές μονάδες χωρίς χρήσιμη συντακτική πληροφορία αλλά και για εντοπισμό συνδέσεων όμοιων συντακτικών όρων. Για παράδειγμα, εντοπίζει την ολοκλήρωση μιας πρότασης μετά από μία τελεία, το είδος μίας πρότασης μετά από ένα ερωτηματικό (ερωτηματική πρόταση) ή ακόμη και την έμμεση σύνδεση δύο συντακτικών όρων, χωρίς τη μεσολάβηση συνδέσμου αλλά με τη χρήση κόμματός (πρόταση_1 , πρόταση_2 / υποκείμενο_1 , υποκείμενο_2). Από την άλλη πλευρά, αφαιρεί σημεία στίξης που δεν προσδίδουν συντακτική πληροφορία, όπως το “@” ή το “:”.

Ένας συντακτικός αναλυτής, ανάλογα με την εφαρμογή στην οποία χρησιμοποιείται ενδέχεται να αγνοεί ορισμένα σημεία στίξης ενώ για κάποια άλλα να κάνει συγκεκριμένες ενέργειες. Έτσι για παράδειγμα, ο συντακτικός μας αναλυτής όταν εφαρμόζεται σε κείμενο, αναγνωρίζει πολλά σημεία στίξης όπως η τελεία , το θαυμαστικό και το ερωτηματικό ως παρόμοιας σημασίας, δηλαδή το τέλος μίας πρότασης ή περιόδου. Από την άλλη μεριά, όταν ο αναλυτής εφαρμόζεται σε ένα σύνολο ερωτήσεων μπορεί και χειρίζεται με όμοιο τρόπο την τελεία και το θαυμαστικό αλλά σίγουρα έχει συγκεκριμένες και διαφορετικές ενέργειες για το

ερωτηματικό. Σε κάθε περίπτωση, τα σημεία στίξης αφαιρούνται από τη λίστα με τις λεκτικές μονάδες μετά το πέραςμα από αυτή τη μονάδα.

Όπως λοιπόν γίνεται κατανοητό από τα παραπάνω, ο συντακτικός αναλυτής περιέχει στις δομικές του μονάδες και τα σημεία στίξης και, αναλόγως με την εφαρμογή του, την αντίστοιχη ενέργεια που πρέπει να κάνει, και εάν πρέπει να κάνει, κάθε φορά που τα αναγνωρίζει μέσα στην προς ανάλυση ακολουθία.

Στην πρότασή μας (1) το σημείο στίξης είναι το ερωτηματικό (;). Σίγουρα ο αναλυτής θα αναγνωρίσει ότι πρόκειται για την ολοκλήρωση μιας πρότασης. Τώρα, αναλόγως την εφαρμογή, μπορεί να αναγνωρίσει ότι πρόκειται για ερωτηματική πρόταση και να ενημερώσει την εφαρμογή ή απλά να αφαιρέσει το ερωτηματικό από το σύνολο των προς ανάλυση λεκτικών μονάδων.

3.2.3 Δεσμευμένες λέξεις

Ο συντακτικός μας αναλυτής έχει και μία μονάδα που περιλαμβάνει κάποιες λεκτικές μονάδες τις οποίες πρέπει να χειριστεί με ιδιαίτερο και διαφορετικό τρόπο. Αυτή η μονάδα, και πιο συγκεκριμένα, οι δεσμευμένες λέξεις που περιέχει, είναι άμεσα συνδεδεμένη με την εφαρμογή στην οποία χρησιμοποιείται ο συντακτικός αναλυτής.

Για να γίνει καλύτερα κατανοητή η λειτουργία αυτής της μονάδας θα δώσουμε ένα παράδειγμα εφαρμογής της. Έστω λοιπόν ότι ο συντακτικός αναλυτής μας εφαρμόζεται σε ένα σύνολο ερωτήσεων. Τότε αυτή η μονάδα θα έπρεπε να περιέχει ένα σύνολο λέξεων όπως:

Τι ,
Ποιος, Ποια, Ποιο, (*ερωτηματικές αντωνυμίες*)
Πόσος, Πόση, Πόσο, (*ποσοτικά επιρρήματα*)
Που(*τοπικό*), Πότε (*χρονικό*), Πώς (*τροπικό*), επιρρήματα
Γιατί

Αυτές δεν πρέπει να αναγνωριστούν σαν επιρρήματα ή σαν αντωνυμίες. Για τη συγκεκριμένη εφαρμογή δεν αποτελούν μέρος της προς ανάλυσης ακολουθίας

αλλά τρόπο έκφρασης της ακολουθίας, δηλαδή των ερωτήσεων, και γι' αυτό δε θα πρέπει να συνδεθούν με το συνολικό συντακτικό της ακολουθίας. Επίσης πρέπει , αφού αναγνωριστούν, να αφαιρεθούν από το σύνολο των λεκτικών μονάδων.

Στο μοντέλο του συντακτικού αναλυτή που αναπτύξαμε, αυτή η μονάδα, όπως και αυτή των σημείων στίξης που περιγράφηκε παραπάνω, είναι ευμετάβλητες. Αυτό σημαίνει πως ανάλογα με την εφαρμογή στην οποία εφαρμόζεται ο αναλυτής, αλλάζει και των σύνολο των λεκτικών μονάδων που πρέπει να αφαιρεθούν.

Στην πρότασή μας (1), επειδή ο συντακτικός αναλυτής περιέχει μονάδα με δεσμευμένες λέξεις για ερωτήσεις όπως υποθέσαμε παραπάνω, τότε το αρχικό πώς αναγνωρίζεται και δεν λαμβάνεται υπόψη κατά τη συντακτική ανάλυση της πρότασης.

Επίσης, με κατάλληλες ενέργειες στον αναλυτή μας όπως θα δούμε στη συνέχεια, μπορούμε να χρησιμοποιήσουμε τις δεσμευμένες λέξεις και για κατηγοριοποίηση των προτάσεων. Έτσι στην πρότασή μας (1) , αναγνωρίζοντας το πώς ,μπορούμε να εξάγουμε την πληροφορία ότι πρόκειται για μία ερωτηματική πρόταση που αναζητείται ο τρόπος που έγινε μία ενέργεια.

3.2.4 Συντακτικοί κανόνες

Οι συντακτικοί κανόνες αποτελούν τη βάση του συντακτικού σχολιαστή. Περιέχουν τους κανόνες οι οποίοι θα βοηθήσουν στη συντακτική ανάλυση των λεκτικών μονάδων που έχουν προκύψει. Αυτοί οι κανόνες αναπαριστούν το σύνολο των θεωρητικών κανόνων, όπως αυτοί περιγράφηκαν στην παράγραφο 2.2. και είναι εκφρασμένοι με τη βοήθεια κανονικών εκφράσεων για να γίνουν κατανοητοί από το σχολιαστή και κατ' επέκταση τον υπολογιστή. Εδώ θα παρουσιαστούν με την μορφή απλών **κανονικών εκφράσεων**. Οι κανονικές εκφράσεις που θα χρησιμοποιηθούν έχουν παρουσιαστεί παραπάνω, στην αντίστοιχη ενότητα (2.3 Κανονικές εκφράσεις).

<u>Σύμβολα κανονικών εκφράσεων</u>	
.	Οποιοδήποτε γράμμα
[abc]	Ένα από τα γράμματα a, b, ή c
Έκφραση*	Η έκφραση μηδέν ή περισσότερες φορές
Έκφραση{n}	Η έκφραση n φορές

Έκφραση {n,}	Η έκφραση τουλάχιστον n φορές
Έκφραση {n,m}	Η έκφραση τουλάχιστον n αλλά όχι περισσότερες από m φορές
(Έκφραση)	Το περιεχόμενο στην παρένθεση

Πίνακας 4: Σύμβολα κανονικών εκφράσεων

<u>Σύμβολα που χρησιμοποιούνται</u>	
V	ρήμα
J	επίθετο
N	ουσιαστικό
A	επίρρημα
S	σύνδεσμος
C	κόμμα (,)

Πίνακας 5: Σύμβολα που χρησιμοποιούνται

Είναι σημαντικό να γίνει κατανοητό ότι με τη χρήση των κατάλληλων κανονικών εκφράσεων γίνεται αναπαράσταση πολλών συντακτικών κανόνων σε μια μόνο κανονική έκφραση. Έτσι λοιπόν, για παράδειγμα, οι συντακτικοί κανόνες:

Αντικείμενο + Ρήμα + Αντικείμενο + Υποκείμενο ,
 Αντικείμενο + Ρήμα + Υποκείμενο και
 Ρήμα + Αντικείμενο + Υποκείμενο + Υποκείμενο
 (και άλλες παρόμοιες)

μπορούν να αναπαρασταθούν με την κανονική έκφραση:

*Αντικείμενο * Ρήμα Αντικείμενο * Υποκείμενο **

Άρα, για το σύνολο των συντακτικών κανόνων της ελληνικής γλώσσας και όπως αυτοί πρέπει να υποστηρίζονται με βάση το γλωσσικό μοντέλο που προκύπτει από το συντακτικό αναλυτή (εικόνα 7) έχουμε τις ακόλουθες κανονικές εκφράσεις:

Έκφραση 1:

$[JNA]((([JNA])^* C [JNA])^* V$

Όρος Υποκείμενο – Ρήμα

Η πρόταση αυτή ,ξεκινάει με ένα ή περισσότερα υποκείμενα, τα οποία μπορεί να είναι ουσιαστικό, επίθετο ή επίρρημα, όπως θα δούμε παρακάτω και τα οποία ενώνονται με κόμμα ή με συνδέσμους και ακολουθείται από το ρήμα.

Παράδειγμα έκφρασης:

«Ο Νίκος διαβάζει »

Υποκείμενο – Ρήμα

Έκφραση 2:

$[JNA]((([JNA])^* C [JNA])^* V [JNA]((([JNA])^* C^* [JNA])^*$

Όρος Υποκείμενο – Ρήμα – Όρος Αντικείμενο

Η πρόταση αυτή ,ξεκινάει με ένα ή περισσότερα υποκείμενα, ακολουθείται από το ρήμα και τέλος ένα ή περισσότερα αντικείμενα, τα οποία και αυτά μπορεί να είναι ουσιαστικό, επίθετο ή επίρρημα, όπως θα δούμε παρακάτω και τα οποία ενώνονται με κόμμα ή με συνδέσμους.

Παράδειγμα έκφρασης:

«Ο Κώστας πηγαίνει σχολείο »

Υποκείμενο – Ρήμα – Αντικείμενο

Έκφραση 3:

$[JNA]((([JNA])^* C [JNA])^* V [JNA]((([JNA])^* C^* [JNA])^* [JNA]((([JNA])^* C^* [JNA])^*$

Όρος Υποκείμενο – Ρήμα – Όρος Αντικείμενο – Όρος Αντικείμενο

Η πρόταση αυτή ,ξεκινάει με ένα ή περισσότερα υποκείμενα, ακολουθείται από το ρήμα, έπειτα ένα ή περισσότερα αντικείμενα **προαιρετικά** ακολουθούμενα από ένα ή περισσότερα αντικείμενα **προαιρετικά** αλλά **χωρίς** τη μεσολάβηση συνδέσμου

Παράδειγμα έκφρασης:

«Η γιαγιά χάρισε της Άννας ένα βιβλίο»

Υποκείμενο – Ρήμα – Αντικείμενο(έμμεσο) – Αντικείμενο (άμεσο)

Έκφραση 4:

$V [JNA] ((, [JNA])^* C [JNA])^*$

Ρήμα – Όρος Υποκείμενο

Σε αυτή την έκφραση, η πρόταση ξεκινάει με το ρήμα και ακολουθούν ένα ή περισσότερα υποκείμενα, τα οποία μπορεί να είναι ουσιαστικό, επίθετο ή επίρρημα, όπως θα δούμε παρακάτω και τα οποία ενώνονται με κόμμα ή με συνδέσμους.

Παράδειγμα έκφρασης:

« Σκόραρε ο Κώστας ,ο Γιώργος ή ο Γιάννης ; »

Ρήμα – Υποκείμενο – (,) – Υποκείμενο – Σύνδεσμος – Υποκείμενο ή

Ρήμα – Πολλά Υποκείμενα

Έκφραση 5:

$[JNA]((,[JNA])^*C^*[JNA])^*V((,[JNA])^*C^*[JNA])[JNA]((,[JNA])^*C[JNA])^*$

Όρος Αντικείμενο – Ρήμα – Όρος Αντικείμενο – Όρος Υποκείμενο

Στην έκφραση αυτή, η πρόταση ξεκινάει με ένα ή περισσότερα αντικείμενα **προαιρετικά** , ακολουθεί το ρήμα, ένα ή περισσότερα αντικείμενα **προαιρετικά** και τέλος ένα ή περισσότερα υποκείμενα.

Παράδειγμα έκφρασης:

« Κόκκινη ή κίτρινη κάρτα πήρε ο Γιάννης ; »

Αντικείμενο – Σύνδεσμος – Αντικείμενο – Ρήμα – Υποκείμενο ή

Πολλά Αντικείμενα – Ρήμα – Υποκείμενο

Έκφραση 6:

$V[JNA]((,[JNA])*C*[JNA])*[JNA]((,[JNA])*C*[JNA])*[JNA]((,[JNA])*C*[JNA])^*$;

Ρήμα – Όρος Αντικείμενο – Όρος Αντικείμενο – Όρος Υποκείμενο

Στην έκφραση αυτή, η πρόταση ξεκινάει με το ρήμα, συνεχίζει με ένα ή περισσότερα αντικείμενα **προαιρετικά**, ακολουθεί ένα ή περισσότερα αντικείμενα προαιρετικά αλλά **χωρίς** τη μεσολάβηση συνδέσμου και τέλος ένα ή περισσότερα υποκείμενα.

Παράδειγμα έκφρασης:

«Όρισε επιμελητή τον Κώστα ο διευθυντής ; »

Ρήμα – Αντικείμενο – Αντικείμενο - Υποκείμενο

Είναι προφανές ότι οι παραπάνω εκφράσεις αποτελούν ουσιαστικά μετατροπή των σχημάτων λόγου που παρουσιάστηκαν παραπάνω στην παράγραφο 2.2.5, σε κανονικές εκφράσεις. Η μετατροπή έγινε με μεγάλη προσοχή έτσι ώστε να περιλαμβάνονται όλα τα σχήματα σε όσο το δυνατόν λιγότερες εκφράσεις αλλά και να μη προκύψει σύγκρουση μεταξύ των εκφράσεων, δηλαδή μία πρόταση να ανήκει σε δύο ή περισσότερες εκφράσεις.

Μέχρι στιγμής λοιπόν, μία ακολουθία την έχουμε κερματίσει σε ένα σύνολο λεκτικών μονάδων, αφαιρέσαμε τις μονάδες εκείνες που δε χρειαζόμαστε και μετά το πέρασμα από αυτή τη μονάδα, βλέποντας την σειρά που ακολουθούν οι λεκτικές μονάδες μέσα στην ακολουθία, βρίσκουμε την πιθανή συντακτική έκφραση που αντιστοιχίζεται σε αυτές τις λεκτικές μονάδες.

Όπως γίνεται κατανοητό, η πρότάσή μας (1) ανήκει στην έκφραση 5.

3.2.5 Συντακτικοί όροι

Οι συντακτικοί όροι είναι αναπόσπαστο κομμάτι του συντακτικού μας αναλυτή. Είναι οι μονάδες εκείνες οι οποίες περιέχουν την πληροφορία για την αντιστοίχιση των εναπομείναντων λεκτικών μονάδων (παρ. 2.3.1) σε όλους τους πιθανούς συντακτικούς όρους, αφού λάβουν υπόψη τους και τους συντακτικούς κανόνες (παρ. 2.3.2) αλλά κυρίως λαμβάνοντας υπόψη τις πιθανές γραμματικές αντιστοιχίσεις κάθε όρου.

Κάθε συντακτικός όρος περιέχει όλες τις πιθανές γραμματικές κατηγορίες στις οποίες μπορεί να ανήκει. Πιο συγκεκριμένα, εκφράζει αυτός ο όρος σε τι γραμματική αναπαράσταση μπορεί να εμφανιστεί σε μία πρόταση. Με τον όρο γραμματική αναπαράσταση εννοούμε ουσιαστικό, επίθετο, αντωνυμία κ.ά. Έτσι έχουμε:

Υποκείμενο:

Ουσιαστικό: « **Ο καθηγητής** γράφει »

Επίθετο: « **Ο ψηλός** τελείωσε »

Αντωνυμία: « **Εκείνος** είναι φίλος μου »

Ομάδα λέξεων σε θέση ουσιαστικού: « **Το ταξίδι στο Παρίσι**, ήταν υπέροχο »

Αντικείμενο:

Ουσιαστικό: « Ο Γιώργος έβαψε **τη βιβλιοθήκη** »

Επίθετο: « Ο προπονητής άλλαξε **επιθετικό** »

Αντωνυμία: « Μαρία είδε **εκείνον** στο δρόμο »

Ομάδα λέξεων σε θέση ουσιαστικού: « Εσύ έκανες **όλο τον κόσμο άνω κάτω**; »

Όπως παρατηρούμε από τα δύο παραπάνω, είναι ιδιαίτερα δύσκολη η διάκριση ανάμεσα στους όρους υποκείμενο και αντικείμενο καθώς και οι δύο συναντούνται όχι μόνο σε πολλές και αυθαίρετες περιπτώσεις (κάθε ομάδα λέξεων σε θέση ουσιαστικού) αλλά οι περιπτώσεις αυτές είναι κοινές. Όπως θα δούμε στη συνέχεια, για τη σωστή διάκρισή τους θα στηριχθούμε και σε άλλα στοιχεία τους, όπως η πτώση τους ή το ρήμα που τα συνοδεύει. Αντίθετα το κατηγορούμενο επειδή είναι στενά συνδεδεμένο με το ρήμα, όπως είδαμε και

στην παράγραφο 2.2.2.2 εμφανίζει μικρότερη πολυπλοκότητα και αναγνωρίζεται ευκολότερα.

Κατηγορούμενο:

Ουσιαστικό: « Εγώ είμαι **φοιτητής** »

Επίθετο: « Ο Γιάννης έγινε **στρατηγός** »

Ο τέταρτος βασικός συντακτικός όρος της Νέας Ελληνικής γλώσσας είναι το ρήμα αλλά αυτό από μόνο του είναι παράλληλα και γραμματικός όρος. Έτσι δεν εμφανίζει καμία πολυπλοκότητα, δηλαδή η αντιστοιχία γραμματικού όρου προς συντακτικό είναι ένα προς ένα.

Όταν λοιπόν μία λεκτική μονάδα εισέρχεται σε αυτή τη δομική μονάδα, αποκτά το σύνολο των πιθανών συντακτικών όρων που μπορεί να είναι. Έτσι γίνεται μια πρώτη συντακτική της ανάλυση όπου κάποιες λεκτικές μονάδες αποκλείονται από κάποιους συντακτικούς όρους (μία αντωνυμία δεν μπορεί να είναι κατηγορούμενο) ενώ συσχετίζονται με έναν ή περισσότερους άλλους (μία αντωνυμία μπορεί να είναι υποκείμενο ή αντικείμενο)

3.2.6 Σύνδεσμοι

Οι άκλιτες λέξεις όπως και, ή ,μα ,ώστε ,δηλαδή, γιατί κ.ά. οι οποίες χρησιμοποιούνται μέσα σε προτάσεις για να συνδέουν λέξεις ή προτάσεις λέγονται **σύνδεσμοι**. Υπάρχουν διάφορα είδη συνδέσμων ενώ κάθε σύνδεσμος μπορεί να ανήκει σε περισσότερα από ένα είδη.

<u>Είδη</u>	<u>Σύνδεσμοι</u>
Συμπλεκτικοί	<i>και (κι), ούτε, μήτε, ουδέ, μηδέ, ή ,είτε</i>
Διαχωριστικοί	<i>ή, είτε</i>
Αντιθετικοί	<i>μα, αλλά, παρά, όμως, ωστόσο, ενώ, αν και, μολονότι, μόνο</i>
Συμπερασματικοί	<i>λοιπόν, ώστε, άρα, επομένως ,που</i>
Επεξηγηματικός	<i>δηλαδή</i>
Ειδικοί	<i>πως, που, ότι</i>
Χρονικοί	<i>όταν, σαν, ενώ ,καθώς αφού, αφότου, πριν, μόλις προτού,</i>

	<i>όσπου, ωστόσο, όσο που, οπότε</i>
Αιτιολογικοί	<i>γιατί, επειδή, αφού, τι</i>
Υποθετικοί	<i>αν, σαν, άμα</i>
Τελικοί	<i>να, για να</i>
Αποτελεσματικοί	<i>ώστε, που</i>
Διστακτικοί	<i>μη(ν), μήπως</i>
Συγκριτικός	<i>παρά</i>

Πίνακας 6: Οι σύνδεσμοι

Οι σύνδεσμοι αποτελούν αναπόσπαστο κομμάτι κάθε συντακτικού αναλυτή. Όπως φαίνεται και στο γλωσσικό μοντέλο της εικόνας 7, σε κάθε επίπεδο υπάρχουν και οι αντίστοιχοι σύνδεσμοι που ενώνουν συντακτικούς όρους μεταξύ τους. Έτσι, στο πρώτο επίπεδο υπάρχουν οι σύνδεσμοι που συνδέουν **προτάσεις** μεταξύ τους και έχουν περιγραφεί παραπάνω στην παράγραφο 2.2.4. Σε ένα δεύτερο επίπεδο, ορισμένοι σύνδεσμοι συνδέουν **όρους υποκείμενα** μεταξύ τους. Αυτοί συνήθως είναι οι *συμπλεκτικοί* και οι *διαζευκτικοί* σύνδεσμοι. Οι διάφοροι **όροι κατηγορήματα** συνδέονται με παρατακτική ή υποτακτική σύνδεση, επομένως οι σύνδεσμοι είναι οι αντίστοιχοι *παρατακτικοί* ή *υποτακτικοί* και έχουν παρουσιαστεί στην παράγραφο 2.2.4. Σε ένα τρίτο επίπεδο, οι διάφοροι **όροι αντικείμενα** ενώνονται μεταξύ τους και αυτοί με *συμπλεκτικούς* και *διαζευκτικούς* συνδέσμους. Αυτή είναι μία ακόμη ομοιότητα των όρων υποκείμενα με τους όρους αντικείμενα που κάνει αδύνατη τη διάκριση αυτών των όρων μέσω των συνδέσμων που τους ενώνουν.

Ο συντακτικός αναλυτής μας λαμβάνει υπόψη του τους συνδέσμους που μπορεί να συναντήσει κατά την ανάλυση. Όχι μόνο για να τους αναγνωρίσει αλλά και για να βγάλει χρήσιμα συμπεράσματα από αυτούς. Για παράδειγμα, αναγνωρίζοντας ένα *συμπλεκτικό* σύνδεσμο όπως το *και* ή ένα *διαχωριστικό* όπως το *ή*, μπορεί και αναγνωρίζει ότι υπάρχει ένωση μεταξύ συντακτικών όρων. Έτσι, εάν μόλις έχει ολοκληρώσει την αναγνώριση μιας πρότασης, συναντήσει έναν τέτοιο σύνδεσμο τότε ακολουθεί μία ακόμη πρόταση.

Πιο σημαντική όμως είναι η αναγνώριση που κάνει, όταν πρόκειται για συνδέσμους που ενώνουν όμοιους όρους μέσα σε μία πρόταση. Για παράδειγμα, όταν τρία ουσιαστικά ενώνονται με συνδέσμους και το ένα από τα τρία ή τα δύο από τα τρία έχουν αναγνωριστεί ως αντικείμενα, τότε και το τρίτο ουσιαστικό θα είναι αντικείμενο.

Όλα αυτά μπορούν να διευκολύνουν τη συντακτική ανάλυση ενώ παράλληλα αυξάνουν την πολυπλοκότητα των εκφράσεων που αναγνωρίζονται με τη βοήθεια του αναλυτή. Απλά ο αναλυτής αναγνωρίζει τους συνδέσμους που ενώνουν τις προτάσεις και διαιρεί μια μεγάλη έκφραση σε μικρότερες. Στην περίπτωση σύνδεσης υποκειμένων ή κατηγορημάτων, εάν ο συντακτικός αναλυτής βρίσκεται σε δίλημμα για το τι μπορεί να είναι συντακτικά μία λεκτική μονάδα, μπορεί να διευκολυνθεί εάν αυτή είναι συνδεδεμένη μέσω συνδέσμων με άλλες που έχουν αναγνωριστεί με επιτυχία.

Στην πρότασή μας (1) ενώνονται δύο υποκείμενα με τον σύνδεσμο και, ο οποίος είναι συμπλεκτικός, όπως αναμενόταν από τις παραπάνω παρατηρήσεις στο μοντέλο.

3.3 Το ρήμα

Το ρήμα αποτελεί τη βάση της ορθής λειτουργίας ενός συντακτικού αναλυτή. Το πλεονέκτημα με το ρήμα είναι ότι πέρα από συντακτικός όρος, είναι παράλληλα και γραμματικός όρος. Αυτό σημαίνει ότι όταν μια ακολουθία λεκτικών μονάδων έρχεται στο συντακτικό αναλυτή για αναγνώριση, αυτός δε χρειάζεται να αναγνωρίσει το ρήμα, καθώς αυτό έχει αναγνωριστεί στο προηγούμενο στάδιο, αυτό της γραμματικής ανάλυσης. Έτσι το ρήμα αποτελεί την καλύτερη δυνατή βάση για να αρχίσει η συντακτική ανάλυση.

Με τη χρήση των πληροφοριών που περιέχει το ρήμα από τη γραμματική του ανάλυση, ο συντακτικός αναλυτής μπορεί να προχωρήσει σε δύο βασικές ενέργειες. Την εύρεση των προτάσεων που περιέχει η προς ανάλυση ακολουθία καθώς και ένα πρώτο προσδιορισμό των πιθανών υποκειμένων και αντικειμένων των προτάσεων αυτών. Και οι δύο αυτές ενέργειες θα παρουσιαστούν στη συνέχεια.

3.3.1 Εύρεση προτάσεων

Η εύρεση όλων των προτάσεων που αποτελούν μία ακολουθία είναι απαραίτητη για τη σωστή συντακτική ανάλυση της ακολουθίας. Η πιο σημαντική παρατήρηση στο συντακτικό της ελληνικής γλώσσας είναι ότι **κάθε πρόταση**

αποτελείται από ένα ρήμα. Αυτό σημαίνει ότι ο αριθμός των προτάσεων θα είναι ίσος με τον αριθμό των ρημάτων της προς ανάλυση ακολουθίας.

Είναι πιθανό, από αυτά που αναφέρθηκαν παραπάνω, κάποιος να υποθέσει πως οι προτάσεις μπορούν να διαχωριστούν με τα σημεία στίξης. Αυτό όμως είναι λάθος. Εάν έχουμε σύνδεση προτάσεων με συνδέσμους ή με κόμματα τότε ο συντακτικός αναλυτής θα πρέπει να είναι σε θέση να αναγνωρίσει τις διάφορες προτάσεις και όχι να τις θεωρήσει όλες ως μία ,ενιαία πρόταση μέχρι να βρει το επόμενο σημείο στίξης που σηματοδοτεί το τέλος μίας πρότασης. Για παράδειγμα:

« Ο Κώστας έφυγε , ο Γιάννης διαβάζει **και** εγώ τρώω . »

Η παραπάνω έκφραση, αποτελείται από τρεις προτάσεις, όπου οι δύο πρώτες ενώνονται με κόμμα, ενώ η δεύτερη και η τρίτη ενώνονται με σύνδεσμο. Το σημείο στίξης που δηλώνει το τέλος πρότασης βρίσκεται στο τέλος της τρίτης πρότασης, επομένως ο συντακτικός αναλυτής θα κάνει λάθος εάν στηριχθεί σε αυτό.

Αυτό που κάνει ο συντακτικός αναλυτής είναι να ακολουθεί μία σειρά βημάτων προκειμένου να ενημερώσει κατάλληλα τους δείκτες που σηματοδοτούν την αρχή και το τέλος κάθε πρότασης. Έτσι:

1. Βρίσκει ένα ρήμα μέσα σε μία ακολουθία,

2. Πηγαίνει προς την αρχή της ακολουθίας και

- i. εάν βρει κάποιο σημείο στίξης που σηματοδοτεί τέλος πρότασης, τότε βρήκε την αρχή της τωρινής πρότασης και το τέλος μίας προηγούμενης.
- ii. εάν βρει την αρχή της ακολουθίας, τότε δεν προηγείται τίποτε μπροστά της και η αρχή της προς ανάλυση πρότασης είναι η αρχή της ακολουθίας
- iii. εάν βρει άλλο ρήμα, τότε προηγείται άλλη πρόταση, ψάχνει πώς συνδέονται οι δύο προτάσεις(με κόμμα ή σύνδεσμο) και ενημερώνει κατάλληλα τον δείκτη τέλους της μίας πρότασης και τον δείκτη αρχής της άλλης

3. Αποθηκεύει τους δείκτες της προηγούμενη πρότασης, εάν υπάρχει, και το τέλος της τωρινής και επαναλαμβάνει τα βήματα 1 και 2 μέχρι να τελειώσουν τα ρήματα της ακολουθίας

4. Το τελευταίο ρήμα σηματοδοτεί και τέλος πρότασης κοινό με το τέλος της ακολουθίας.

(ειδικό χειρισμό έχουμε για τις ερωτήσεις οι οποίες μπορεί να ξεκινάνε με ερωτηματικές αντωνυμίες, αλλά σχετική αναφορά γίνεται στην αντίστοιχη παράγραφο στη συνέχεια)

Ουσιαστικά δηλαδή κάθε πρόταση που ακολουθεί τα παραπάνω βήματα, εάν δε βρίσκεται μόνη της στην ακολουθία, ενημερώνει το δείκτη τέλους της πρότασης που προηγείται και τον δείκτη αρχής της ίδιας, ενώ ο δείκτης τέλους της τελευταίας πρότασης είναι ίδιος με το δείκτη τέλους της ακολουθίας. Για παράδειγμα για την έκφραση που δώσαμε ως παράδειγμα παραπάνω, ο συντακτικός αναλυτής θα ακολουθούσε τα ακόλουθα βήματα:

- i. Βρίσκει το ρήμα «έφυγε»
- ii. Πηγαίνει προς την αρχή της ακολουθίας και (ii) βρίσκει την αρχή της ακολουθίας και επομένως η τωρινή πρόταση αρχίζει στην αρχή της ακολουθίας
- iii. Αποθηκεύει τον δείκτη αρχής της πρότασης με το ρήμα «έφυγε»
- iv. Βρίσκει το ρήμα «διαβάζει»
- v. Πηγαίνει προς την αρχή της ακολουθίας και (iii) βρίσκει το ρήμα «έφυγε» που προηγείται, βρίσκει το κόμμα που μεσολαβεί και ενημερώνει το δείκτη τέλους της πρότασης με το ρήμα «έφυγε» και το δείκτη αρχής της πρότασης με το ρήμα «διαβάζει»
- vi. Βρίσκει το ρήμα «τρώω»
- vii. Πηγαίνει προς την αρχή της ακολουθίας και (iii) βρίσκει το ρήμα «διαβάζει» που προηγείται, βρίσκει το σύνδεσμο και που μεσολαβεί και ενημερώνει το δείκτη τέλους της πρότασης με το ρήμα «διαβάζει» και το δείκτη αρχής της πρότασης με το ρήμα «τρώω»
- viii. Το ρήμα «τρώω» είναι και το τελευταίο οπότε ο δείκτης τέλους της πρότασης που το περιέχει είναι το τέλος της ακολουθίας
- ix. Ολοκλήρωση της εύρεσης των προτάσεων(3 ρήματα - 3 προτάσεις) με επιτυχία

Βλέπουμε λοιπόν πόσο σημαντικό είναι το ρήμα για ένα συντακτικό αναλυτή. Στη συνέχεια θα δούμε πώς μπορούμε να χρησιμοποιήσουμε πληροφορίες σχετικές με το ρήμα για ένα πρώτο προσδιορισμό των πιθανών υποκειμένων και αντικειμένων του.

3.3.2 Προσδιορισμός υποκειμένων

Η ιδιότητα που έχει το ρήμα να δείχνει κάθε φορά **τι κάνει το υποκείμενο** ή **τι παθαίνει** ή **σε ποια κατάσταση βρίσκεται** λέγεται **διάθεση** του ρήματος. Ο συντακτικός αναλυτής μπορεί να εκμεταλλευτεί αυτήν την πληροφορία είτε για να προσδιορίσει ακριβώς είτε για να έχει βοηθηθεί στην εύρεση των υποκειμένων της πρότασης.

Όσο αφορά το υποκείμενο, ο συντακτικός αναλυτής πρέπει να γνωρίζει πως όταν το ρήμα βρίσκεται στο πρώτο ή δεύτερο πρόσωπο, ενικό ή πληθυντικό, τότε είναι πολύ πιθανό το υποκείμενό του να παραληφθεί. Για παράδειγμα:

« Πήγαμε με το λεωφορείο » (το υποκείμενο εμείς εννοείτε)

Έτσι σε αυτή την έκφραση, η σωστή αναγνώριση του προσώπου του ρήματος θα βοηθήσει το συντακτικό αναλυτή στη συνέχεια να μην κάνει λάθος στον προσδιορισμό του υποκειμένου. Αυτή η πρόταση έχει υποκείμενο, απλά εννοείται, κάτι που συνηθίζεται ιδιαίτερα στον προφορικό λόγο και γι' αυτό είναι απαραίτητο ένας συντακτικός αναλυτής φυσικής γλώσσας να το λαμβάνει υπόψη του. Σε αυτή την περίπτωση το υποκείμενο **προσδιορίζεται ακριβώς** , μέσω του ρήματος.

Σχετικά με το ρήμα και το υποκείμενο υπάρχει και η ακόλουθη παραδοχή:

Παραδοχή ρήματος υποκειμένου: «...σε μία πρόταση που έχει ένα υποκείμενο το ρήμα συμφωνεί με το υποκείμενό του στο πρόσωπο και στον αριθμό...» Συντακτικό της Νέας Ελληνικής,σελ.23

Η παραπάνω παραδοχή μας βοηθάει να αποκλείσουμε πιθανά υποκείμενα και να επικεντρωθούμε σε πιο συγκεκριμένα, έτσι στην πρόταση:

« Ο Κώστας πηγαίνει στα καταστήματα » (1)

με γραμματική ανάλυση:

Ο – *άρθρο* / αριθμός: ενικός, γένος: αρσενικό, πτώση:ονομαστική

Κώστας – *ουσιαστικό* / αριθμός: ενικός, γένος: αρσενικό, πτώση:ονομαστική

πηγαίνει – *ρήμα* / χρόνος:ενεστώτας , έγκλιση:οριστική, φωνή:ενεργητική, αριθμός:ενικός, πρόσωπο: τρίτο, ρίζα «πηγαίνω»

στα – *άρθρο* / αριθμός: πληθυντικός, γένος:ουδέτερο, πτώση: αιτιατική

καταστήματα – *ουσιαστικό* / αριθμός: πληθυντικός, γένος: ουδέτερο, πτώση:αιτιατική

ο αναλυτής θα μπορούσε να κάνει λάθος στο υποκείμενο καθώς οι κανόνες που αναφέραμε παραπάνω (παρ. 3.2.2) για την σειρά των όρων δεν ακολουθούνται πάντα, ειδικά στον προφορικό λόγο. Έτσι η ίδια πρόταση θα μπορούσε να εκφραστεί ως εξής:

«Στα καταστήματα πηγαίνει ο Κώστας» (2)

Με βάση τον κανόνα της έκφρασης 2 στην παράγραφο 3.2.2 στην πρώτη πρόταση (1) το υποκείμενο είναι ο Κώστας, σωστά, ενώ στη δεύτερη (2) είναι το σχολείο, λανθασμένα.

Χρησιμοποιώντας την παραπάνω παραδοχή, αποκλείουμε το ουσιαστικό *καταστήματα* να είναι υποκείμενο καθώς **δεν συμφωνεί** με το ρήμα στον αριθμό.

Όλα τα παραπάνω μας βοηθούν ιδιαίτερα στον προσδιορισμό του όρου υποκείμενο όπως θα δούμε και στη συνέχεια.

3.3.3 Προσδιορισμός αντικειμένων

Σχετικά με το αντικείμενο, οι πληροφορίες που εξάγουμε από το είδος του ρήματος είναι οι ακόλουθες:

Εάν το ρήμα είναι **μονόπτωτο** ,δηλαδή αυτό έχει μόνο ένα αντικείμενο, τότε για παράδειγμα:

*«Ο Γιάννης έβαψε **τον τοίχο**»*

Η ενέργεια του ρήματος, αφού είναι μονόπρωτο, πηγαίνει σε ένα μόνο αντικείμενο. Αυτή είναι μια πολύ σημαντική πληροφορία για τον αναλυτή.

Από την άλλη, εάν το ρήμα είναι **δίπρωτο**, τότε θα έχουμε σίγουρα δύο αντικείμενα, ένα άμεσο και ένα έμμεσο, για παράδειγμα:

«Ο καθηγητής εξέτασε το Γιάννη μαθηματικά »

Στην παραπάνω πρόταση βλέπουμε ότι η ενέργεια του ρήματος πηγαίνει σε δύο αντικείμενα:

το **άμεσο** (Γιάννη), το οποίο λέγεται έτσι γιατί σε αυτό πηγαίνει ή με αυτό σχετίζεται άμεσα η ενέργεια του υποκειμένου ρήματος και

το **έμμεσο** (*μαθηματικά*), το οποίο λέγεται έτσι γιατί σε αυτό πηγαίνει έμμεσα η ενέργεια του υποκειμένου του ρήματος. Το έμμεσο αντικείμενο μπορεί να αντικατασταθεί από εμπρόθετο προσδιορισμό (*στα μαθηματικά*)

Από τα παραπάνω γίνεται κατανοητό ότι το ρήμα **δίνει πληροφορίες** αλλά **δε μπορεί να προσδιορίσει ακριβώς** το αντικείμενο.

Πριν προχωρήσουμε στην επόμενη παράγραφο να υπενθυμίσουμε ,όπως αναφέρθηκε στην παράγραφο 2.2.2.2 παραπάνω , ότι εάν το ρήμα είναι συνδετικό τότε γνωρίζουμε ότι έχουμε **κατηγορούμενο** και όχι αντικείμενο.

3.4 Ο όρος υποκείμενο

Αφού ολοκληρώσει την ανάλυση του ρήματος και συγκεντρώσει όλη τη δυνατή πληροφορία, ο συντακτικός αναλυτής προχωρά στον προσδιορισμό του υποκειμένου της πρότασης ή των προτάσεων της προς ανάλυση ακολουθίας.

Η εύρεση του υποκειμένου είναι μια ιδιαίτερα δύσκολη ενέργεια εξαιτίας της πολυπλοκότητας της ελληνικής γλώσσας. Σε μία πρόταση μπορεί το υποκείμενο να εννοείται (1) ή να είναι το ίδιο για δύο ή περισσότερες προτάσεις και επομένως να παραλείπεται (2). Για παράδειγμα:

«Πήγες σήμερα σχολείο ;» (το υποκείμενο εσύ **εννοείται**) (1)

«Ο Κώστας ήρθε και έφυγε » (το υποκείμενο της δεύτερης πρότασης είναι ίδιο με αυτό της πρώτης, δηλαδή ο Κώστας και **παραλείπεται**) (2)

Μπορεί επίσης η πρόταση να περιέχει περισσότερα από ένα υποκείμενα (3) ή ακόμη και το ρόλο του υποκειμένου να παίζει ολόκληρη πρόταση (4), για παράδειγμα:

«Ο Κώστας και ο Γιώργος πήγαν σχολείο » (δύο υποκείμενα ο Κώστας και ο Γιώργος) (3)

«Το ταξίδι με το τρένο, ήταν πολύ κουραστικό» (υποκείμενο είναι ολόκληρη η πρόταση το ταξίδι με το τρένο) (4)

Ο συντακτικός αναλυτής μπορεί να αναλύσει με επιτυχία αυτές τις εκφράσεις εάν λάβει υπόψη του κάποιους κανόνες του συντακτικού της ελληνικής γλώσσας, τους χειριστεί κατάλληλα και σε συνδυασμό με όποια πληροφορία μπορεί να έχει από το ρήμα, το οποίο υπενθυμίζουμε ότι έχει αναγνωριστεί **πριν** το υποκείμενο.

Στις επόμενες παραγράφους θα παρουσιάσουμε τους κανόνες εκείνους που μπορούν να διευκολύνουν τον αναλυτή στην εύρεση υποκειμένου καθώς και τον τρόπο με τον οποίο θα πρέπει να χειρίζεται ο αναλυτής τη σύνδεση πολλών υποκειμένων.

3.4.1 Κανόνες εύρεσης

Οι κανόνες εύρεσης του υποκειμένου προέκυψαν έπειτα από την ανάγνωση πολλών και διαφορετικών συντακτικών έτσι ώστε να είναι όσο το δυνατόν πιο πλήρης, ολοκληρωμένοι και κοινώς αποδεκτοί. Οι κανόνες αυτοί αποτελούν επίσημες παραδοχές του συντακτικού της ελληνικής γλώσσας, παρατηρήσεις και κανόνες. Ο συνδυασμός τους με την πληροφορία που μπορεί να έχει ο αναλυτής από το ρήμα μπορούν να περιορίσουν σημαντικά την πιθανότητα λάθους και να αυξήσουν τις σωστές αναλύσεις.

Οι κανόνες που θα περιγραφούν είναι συμπληρωματικοί με τους γραμματικούς κανόνες(παρουσιάζονται παρακάτω) **αλλά** εφαρμόζονται σε συντακτικούς και όχι γραμματικούς όρους, όπως οι γραμματικοί κανόνες. Υπάρχουν

εξαιρέσεις σε αυτούς τους κανόνες αλλά γενικά περιγράφουν τις βασικές αρχές που διέπουν την ελληνική γλώσσα

Κανόνας 1: «...όταν το υποκείμενο είναι πρώτου ή δευτέρου προσώπου, η αντωνυμία (εγώ, εσύ – εμείς, εσείς) συνήθως παραλείπεται, γιατί δηλώνεται από την κατάληξη του ρήματος...» Συντακτικό της Νέας Ελληνικής, σελ.16. Για παράδειγμα:

« Διαβάζεις ; » (το υποκείμενο εσύ εννοείται)

Αυτός ο κανόνας βρίσκεται μεν και αφορά το υποκείμενο, αλλά οι πληροφορίες για την εφαρμογή του βρίσκονται στο ρήμα που έχει αναλυθεί προηγουμένως.

Κανόνας 2: «...το υποκείμενο είτε σύναρθρο είτε άναρθρο τοποθετείται σε πτώση ονομαστική...» Συντακτικό της Νέας Ελληνικής, σελ.16. Για παράδειγμα:

« Οι νέοι διασκεδάζουν » (το σύναρθρο υποκείμενο νέοι βρίσκεται σε ονομαστική)

Ο κανόνας αυτός μπορεί να μας βοηθήσει ιδιαίτερα στην περίπτωση όπου τα πιθανά υποκείμενα είναι πολλά αλλά σε διαφορετικές πτώσεις.

Κανόνας 3: Όταν και τα δύο ουσιαστικά πληρούν και τους δύο προηγούμενους κανόνες τότε το ουσιαστικό που έχει άρθρο είναι πιθανότατα το υποκείμενο. Να πληρούνται από περισσότερα του ενός ουσιαστικά οι δύο παραπάνω κανόνες είναι αδύνατο. Γίνεται όμως εξαιτίας της ασάφειας που μπορεί να εισάγει ο γραμματικός αναλυτής. Έτσι στην πρόταση:

«Ο Κώστας πηγαίνει σχολείο»

το ουσιαστικό *σχολείο* μπορεί να βρίσκεται είτε σε πτώση ονομαστική είτε σε πτώση αιτιατική. Εδώ λοιπόν εφαρμόζεται ο παραπάνω κανόνας που μας υποδεικνύει ως πιθανό υποκείμενο το ουσιαστικό στο οποίο προηγείται το άρθρο.

Τους παραπάνω κανόνες εμείς στο μοντέλου του αναλυτή μας τους έχουμε ενσωματώσει σε ένα σύνολο συναρτήσεων. Κάθε λεκτική μονάδα, εισέρχεται σε

αυτές τις συναρτήσεις και ελέγχεται με ποιους κανόνες είναι σύμφωνη. Έτσι προωθείται η ανάλυση καθώς εξάγουμε περαιτέρω συντακτική πληροφορία.

3.4.2 Σύνδεση υποκειμένων

Όπως είναι λογικό, μία πρόταση μπορεί να έχει περισσότερα από ένα υποκείμενα. Για τον εντοπισμό τους στηρίζομαστε σε εκείνον τον κανόνα ο οποίος αναφέρει πως υποκείμενα περισσότερα του ενός ενώνονται μεταξύ τους **απαραίτητα με ένα τουλάχιστον σύνδεσμο** και με μια σειρά από κόμματα ή συνδέσμους. Έτσι για να συναντήσουμε υποκείμενα περισσότερα του ενός, πρέπει οπωσδήποτε αυτά να ενώνονται με σύνδεσμο ή κόμμα.

Υποστηρίζοντας τη σύνδεση των υποκειμένων, ο αναλυτής αυξάνει την πολυπλοκότητα των εκφράσεων που αναλύει με επιτυχία. Πλέον οι προτάσεις μπορούν να έχουν περισσότερα του ενός υποκείμενα κάτι που πλησιάζει περισσότερο στη φυσική γλώσσα και κατ' επέκταση στον προφορικό λόγο.

Για την εύρεση της σύνδεσης υποκειμένων μεταξύ τους, ο συντακτικός αναλυτής ακολουθεί μια σειρά από βήματα. Έτσι λοιπόν:

1. Βρίσκει το πιθανότερο υποκείμενο με τους κανόνες εύρεσης(παρ.3.4.1)
2. Πηγαίνει προς την αρχή της πρότασης και ελέγχει τι βρίσκεται αμέσως πριν από αυτό και:
 - i. εάν βρει κόμμα και πριν από αυτό πιθανό υποκείμενο τότε ίσως υπάρχει υποκείμενο που προηγείται, ψάχνει όμως απαραίτητα για κάποιο σύνδεσμο που να ενώνει κάποιο πιθανό υποκείμενο με κάποιο επίσης πιθανό υποκείμενο.
 - ii. εάν βρει σύνδεσμο, τότε σίγουρα βρήκε πιθανό υποκείμενο καθώς οι σύνδεσμοι ενώνουν παρόμοια πράγματα **μέσα στην πρόταση** (συνδέουν και προτάσεις μεταξύ τους, αλλά στο σημείο που βρισκόμαστε έχουμε ξεχωρίσει τα όρια των προτάσεων)
 - iii. εάν βρει οτιδήποτε άλλο ολοκληρώνει τον έλεγχο
3. Συνεχίζει τα βήματα 1 και 2 για όλα τα πιθανά υποκείμενα, αποθηκεύοντας κάθε φορά τα πιθανά υποκείμενα που υπακούν στους κανόνες

4. Ολοκληρώνει τον έλεγχο σύνδεσης υποκειμένων

Για να γίνουν περισσότερο κατανοητά τα παραπάνω βήματα θα δώσουμε ένα παράδειγμα. Έστω λοιπόν η πρόταση:

«Στα δωμάτια ο Κώστας, ο Γιώργος και εγώ διαβάζουμε ! »

- i. Βρίσκει το πιθανό υποκείμενο «εγώ»
- ii. Πηγαίνει προς την αρχή της πρότασης και βρίσκει αμέσως πριν από αυτό (i)σύνδεσμο, τότε σίγουρα προηγείται υποκείμενο, ο «Γιώργος».
- iii. Αποθηκεύει το «εγώ» και έχει σαν νέα βάση το υποκείμενο «Γιώργος»
- iv. Πηγαίνει προς την αρχή της πρότασης και βρίσκει αμέσως πριν από αυτό (ii) κόμμα, και επειδή έχει ήδη βρει σύνδεσμο προηγουμένως , προηγείται σίγουρα υποκείμενο, ο «Κώστας»
- v. Αποθηκεύει το «Γιώργος» και έχει σαν νέα βάση το υποκείμενο «Κώστας»
- vi. Πηγαίνει προς την αρχή της πρότασης και βρίσκει αμέσως πριν από αυτό(iii) ένα πιθανό υποκείμενο χωρίς τη μεσολάβηση κόμματος ή συνδέσμου
- vii. Αποθηκεύει το «Κώστας» και ολοκληρώνει τον έλεγχο σύνδεσης υποκειμένων

Βλέπουμε λοιπόν ότι με την παραπάνω σειρά βημάτων ο αναλυτής μπορεί να αναγνωρίσει όλες τις πιθανές συνδέσεις μεταξύ υποκειμένων με επιτυχία. Αυτό βοηθάει το συντακτικό αναλυτή να αυξήσει την πολυπλοκότητα των προτάσεων που μπορεί να αναλύσει με επιτυχία καθώς αυτές μπορούν να περιέχουν σύνθετους όρους. Ο αναλυτής θα σπάει αυτούς τους όρους σε μικρότερους σε μια προσπάθεια να τους αναγνωρίσει σωστά

Στη συνέχεια θα δούμε πώς λειτουργεί ο αναλυτής στην περίπτωση του τρίτου μεγάλου συντακτικού όρου της ελληνικής γλώσσας, του κατηγορήματος.

3.5 Ο όρος αντικείμενο

Ο τρίτος βασικός συντακτικός όρος που πρέπει να βρει ο συντακτικός αναλυτής είναι το αντικείμενο. Όταν φθάσει σε αυτό το σημείο, έχει κάνει τη διάκριση των προτάσεων, έχει βρει το ρήμα κάθε πρότασης και έχει βρει και τους όρους υποκείμενα σε κάθε πρόταση. Οι λεκτικές λοιπόν μονάδες που απομένουν στην ακολουθία που του δόθηκε να αναλύσει, είναι πιθανά αντικείμενα.

Το γεγονός ότι ο όρος αντικείμενο ορίζεται τελευταίος από το συντακτικό αναλυτή, δε σημαίνει πως ότι δεν ανήκει στις άλλες κατηγορίες ανήκει σε αυτό. Και αυτόν τον όρο τον διέπουν κανόνες τους οποίους πρέπει να πληροί μία λεκτική μονάδα προκειμένου να είναι πιθανό αντικείμενο. Αυτοί οι κανόνες υποδεικνύουν τι μπορεί να αποτελέσει ένα αντικείμενο (παράγραφος 3.2.3), που πρέπει να βρίσκεται αυτό μέσα στην πρόταση (παράγραφος 2.2.5) και πώς πρέπει να συνδέεται με άλλα αντικείμενα, όπως παρουσιάζεται στη συνέχεια.

Στις επόμενες παραγράφους θα παρουσιάσουμε τις κατηγορίες αντικειμένων που μπορεί να συναντήσει ο συντακτικός αναλυτής, πώς τις ξεχωρίζει και πώς τις χειρίζεται. Επίσης θα παρουσιαστεί και η σύνδεση των αντικειμένων μεταξύ τους.

3.5.1 Κατηγορίες αντικειμένων

Ο όρος αντικείμενο εμφανίζει τέσσερις κατηγορίες: **απλό**, **άμεσο**, **έμμεσο** και **κατηγορούμενο**. Αυτές διακρίνονται και στο μοντέλο της εικόνας 7. Το βασικό κριτήριο διάκρισης μεταξύ των κατηγοριών είναι το είδος του ρήματος. Εάν το ρήμα είναι **μεταβατικό** τότε αυτό έχει απαραίτητο συμπλήρωμα της έννοιάς του το αντικείμενο. Ένα όνομα δηλαδή, συνήθως ουσιαστικό ή ότι μπορεί να είναι ισοδύναμο με ουσιαστικό, που φανερώνει το πρόσωπο ή το πράγμα στο οποίο μεταβαίνει ή με το οποίο έχει σχέση η ενέργεια του υποκειμένου του μεταβατικού ρήματος.

Η έννοια ενός μεταβατικού ρήματος μπορεί να συμπληρώνεται με ένα μόνο αντικείμενο, ενώ ενός άλλου με δύο αντικείμενα. Στην πρώτη περίπτωση το ρήμα χρειάζεται ως συμπλήρωμα μία μόνο πλάγια πτώση ονόματος και λέγεται **μονόπτωτο**:

*«Η μητέρα σκουπίζει το **σπίτι**»*

Στη δεύτερη χρειάζεται ως συμπλήρωμα δύο πλάγιες πτώσεις ονομάτων και λέγεται **δίπτωτο**:

«*Τα εγγόνια έδωσαν του **παππού** τα **δώρα** »*

Είναι πολύ σημαντικό ο συντακτικός αναλυτής να καταφέρει να κατηγοριοποιήσει το αντικείμενο κατάλληλα. Αυτό γιατί εάν ξέρει τι αντικείμενο έχει, η ανάλυση γίνεται ευκολότερη. Εάν για παράδειγμα έχει άμεσο και έμμεσο αντικείμενο τότε ψάχνει για δύο πιθανά αντικείμενα στην πρόταση, σε συγκεκριμένη πτώση όπως θα δούμε στη συνέχεια. Εάν έχει κατηγορούμενο ή απλό αντικείμενο τότε ψάχνει ένα αντικείμενο, χωρίς να αποκλείεται η σύνδεση μεταξύ αντικειμένων, επίσης στις αντίστοιχες πτώσεις.

3.5.1.1 Απλό αντικείμενο

Απλό αντικείμενο έχουμε στην περίπτωση των μονόπτωτων ρημάτων.

Τα περισσότερα από τα μονόπτωτα ρήματα συντάσσονται με **αιτιατική** (**εξωτερικό αντικείμενο**). Έτσι ο συντακτικός αναλυτής γνωρίζει ότι το αντικείμενο θα βρίσκεται σε πτώση αιτιατική. Για παράδειγμα:

«*Ο δάσκαλος έκλεισε την πόρτα*»

«*Ο γείτονας κόβει το δέντρο »*

Εάν το αντικείμενο είναι από την ίδια ρίζα με το ρήμα ή από διαφορετική που έχει όμως συγγενική σημασία με αυτό, τότε λέγεται **εσωτερικό** ή **σύστοιχο αντικείμενο**:

«*Μερικοί ζουν τη ζωή τους καλά*»

«*Κοιμήθηκε τον ύπνο του δικαίου*»

Το σύστοιχο αντικείμενο συχνά συνοδεύεται από έναν επιθετικό προσδιορισμό:

«*Χορεύει τον καλαματιανό*»

Πολλές φορές, όπως παραπάνω, το ίδιο το σύστοιχο αντικείμενο παραλείπεται και μένει μόνο ο επιθετικός προσδιορισμός ως σύστοιχο αντικείμενο(... τον καλαματιανό χορό)

Υπάρχουν όμως και αντικείμενα τα οποία συντάσσονται με **γενική**. Λίγα είναι τα μονόπρωτα ρήματα που συντάσσονται με γενική ενώ στη θέση αυτής της γενικής μπορεί να υπάρχει εμπρόθετο αντικείμενο (με μια από τις προθέσεις *με, σε, από* + αιτιατική), για παράδειγμα:

«**Μου** ξέφυγε μία καλή ευκαιρία»

«**Από εμένα** ξέφυγε μία καλή ευκαιρία »

Όμως στην περίπτωση που το ρήμα είναι δίπτωτο τότε ο αναλυτής ψάχνει για δύο αντικείμενα στην ίδια ή σε διαφορετική πτώση.

3.5.1.2 Άμεσο και έμμεσο αντικείμενο

Τα δίπτωτα ρήματα έχουν δύο αντικείμενα, όπως αναφέρθηκε παραπάνω.

«*Η* γιαγιά χάρισε της **Μαρίας** ένα **δαχτυλίδι** »

Από τα δύο διαφορετικά αντικείμενα ενός δίπτωτου ρήματος το ένα λέγεται άμεσο αντικείμενο, γιατί σε αυτό πηγαίνει ή με αυτό σχετίζεται άμεσα η ενέργεια του υποκειμένου του ρήματος. Το άλλο λέγεται έμμεσο αντικείμενο, γιατί σε αυτό πηγαίνει έμμεσα η ενέργεια του υποκειμένου του ρήματος. Το έμμεσο αντικείμενο μπορεί να αντικατασταθεί από εμπρόθετο προσδιορισμό (... στη *Μαρία*)

Τα αντικείμενα μπορεί να είναι το άμεσο σε **αιτιατική** και το έμμεσο σε **γενική**:

«**Μου** γνώρισε το **φίλο** του»

«*Κάνεις κακό* του *εαυτού* σου »

Πολλά δίπτωτα ρήματα συντάσσονται με αιτιατική και γενική. Υπάρχει όμως και η περίπτωση όπου και τα δύο αντικείμενα είναι σε **αιτιατική**:

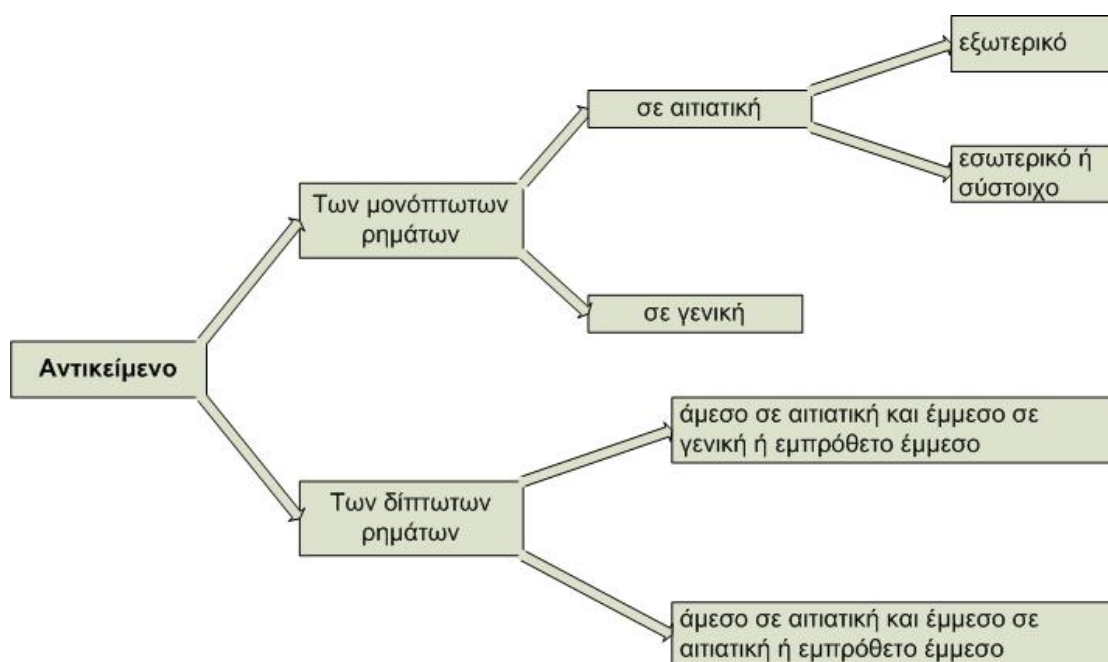
«Σήμερα εξετάζουν τους μαθητές γεωγραφία»

«Ο Γιάννης διδάσκει τον Πέτρο χορό »

Αρκετά είναι τα ρήματα που συντάσσονται με δύο αιτιατικές από τις οποίες συχνά η μία δηλώνει ένα πρόσωπο στο οποίο μεταβαίνει η ενέργεια του υποκειμένου του ρήματος και η άλλη ένα πράγμα. Στις περιπτώσεις αυτές, η αιτιατική του προσώπου ή του προσωποποιημένου πράγματος ή του έμψυχου συνήθως είναι το άμεσο αντικείμενο(...τους μαθητές). Η αιτιατική αυτή είναι έμμεσο αντικείμενο μόνο αν μπορεί να αντικατασταθεί από εμπρόθετο:

«Ο Γιάννης διδάσκει στον Πέτρο χορό»

Έτσι σχηματικά έχουμε:



Εικόνα 10: Οι περιπτώσεις του αντικειμένου

Αυτές είναι οι πιθανές περιπτώσεις που μπορεί και περιμένει να συναντήσει ο συντακτικός αναλυτής στα άμεσα και έμμεσα αντικείμενα. Στη συνέχεια θα δούμε και την τρίτη κατηγορία των αντικειμένων , αυτή του κατηγορουμένου.

3.5.1.3 Κατηγορούμενο

Όταν το ρήμα της πρότασης αποτελείται από έναν τύπο του ρήματος *είμαι* ή άλλου συγγενικού και από ένα όνομα, επίθετο ή ουσιαστικό τότε αυτό, το όνομα, επίθετο ή ουσιαστικό, που με τη μεσολάβησή του ρήματος *είμαι* ή άλλου συγγενικού, φανερώνει ποια ποιότητα ή ιδιότητα αποδίδεται στο υποκείμενο λέγεται **κατηγορούμενο** ενώ το ρήμα που μεσολαβεί μεταξύ του υποκειμένου και του κατηγορουμένου λέγεται **συνδετικό ρήμα**, γιατί συνδέει το κατηγορούμενο με το υποκείμενο.

«Ο Πέτρος είναι **προσεκτικός**»

Για το κατηγορούμενο ισχύουν όσα και για το απλό αντικείμενο που περιγράφηκαν παραπάνω (παρ. 3.5.1.1). Η μόνη διαφορά είναι ότι εδώ έχουμε συνδετικό ρήμα ενώ στο απλό αντικείμενο το ρήμα που συναντάμε είναι μεταβατικό.

Επιπλέον για το κατηγορούμενο υπάρχουν και οι δύο ακόλουθες παραδοχές σε σχέση με το υποκείμενο με το οποίο συνδέεται:

Παραδοχή 1: «...στις απλές προτάσεις το κατηγορούμενο εάν είναι επίθετο, συμφωνεί με το υποκείμενο του ρήματος στο γένος, στον αριθμό και στην πτώση(1) ενώ εάν είναι ουσιαστικό, κανονικά συμφωνεί στην πτώση και μπορεί να συμφωνεί συμπτωματικά στο γένος και τον αριθμό(2)...», Συντακτικό της Νέας Ελληνικής, σελ.24

«Ο Αλέξανδρος είναι **δυνατός**»(1)

«Η Θήβα είναι **πόλη**»(2)

Παραδοχή 2: «...στις σύνθετες προτάσεις όταν υπάρχουν δύο ή περισσότερα υποκείμενα, παρατηρούμε τα εξής:

- i. Το κατηγορούμενο μπαίνει πάντοτε στον πληθυντικό αριθμό (1),(2),
- ii. Αν τα υποκείμενα είναι πρόσωπα του ίδιου γένους, το κατηγορούμενο που είναι επίθετο εκφέρεται στο κοινό γένος των υποκειμένων (1),

- iii. Αν τα υποκείμενα φανερώνουν πράγματα ή αφηρημένες έννοιες, το επιθετικό κατηγορούμενο εκφέρεται στο ουδέτερο γένος, οποιουδήποτε γένους και εάν είναι τα υποκείμενα (2) » , Συντακτικό της Νέας Ελληνικής, σελ.25

«*Η μητέρα και η κόρη της είναι λεπτές*»(1)

«*Ο Γιώργος και η Μαρία είναι καλά παιδιά*»(2)

Οι δύο παραπάνω παραδοχές διευκολύνουν το συντακτικό αναλυτή καθώς όταν ψάχνει για κατηγορούμενα έχει ήδη βρει τα υποκείμενα επομένως μπορεί να τις εφαρμόσει και να εντοπίσει κάποια πιθανά κατηγορούμενα ή να αποκλείσει εκείνα που δεν πληρούν τις παραπάνω παραδοχές.

Μια πολύ σημαντική παρατήρηση έχει να κάνει με όσα παρουσιάστηκαν σε αυτή και την προηγούμενη παράγραφο(παρ. 3.5.1.2). Πολλά είναι τα ρήματα που φαίνονται να συντάσσονται με δύο αιτιατικές ενώ στην πραγματικότητα συντάσσονται με μία. Η άλλη αιτιατική δεν είναι αντικείμενο αλλά χρησιμεύει (μόνη της ή με τα μόρια ως, για, σαν) ως **κατηγορούμενο στο αντικείμενο**. Για παράδειγμα:

«*Θα ορίσουμε το Γιάννη αρχηγό* »

Για να διαπιστώσουμε καλύτερα αν όντως είναι κατηγορούμενο στο αντικείμενο, πρέπει να μετατρέψουμε το ενεργητικό ρήμα σε παθητικό. Τότε, το άμεσο αντικείμενο γίνεται υποκείμενο ενώ η άλλη αιτιατική γίνεται κατηγορούμενο, δηλαδή στην προηγούμενη πρόταση:

«*Ο Γιάννης θα οριστεί αρχηγός* »

Αυτή η περίπτωση είναι πολύ δύσκολο να προβλεφθεί σωστά από το συντακτικό αναλυτή καθώς ο τρόπος επίλυσης του προβλήματος, η μετατροπή δηλαδή του ρήματος σε ενεργητικό και ο μετασχηματισμός της πρότασης, δεν είναι εύκολα υλοποιήσιμος. Απλά σε περίπτωση λάθους, ο συντακτικός αναλυτής θα αναγνωρίσει δύο αντικείμενα αντί για ένα και ένα κατηγορούμενο στο αντικείμενο

3.5.2 Σύνδεση όρων αντικείμενα

Μία πρόταση μπορεί να έχει περισσότερα από ένα αντικείμενα συνδεδεμένα με συνδέσμους μεταξύ τους ή με τη μεσολάβηση κόμματος. Για τον εντοπισμό τους στηρίζομαστε σε εκείνον τον κανόνα ο οποίος αναφέρει πως αντικείμενα περισσότερα του ενός ενώνονται μεταξύ τους **απαραίτητα με ένα τουλάχιστον σύνδεσμο** και με μια σειρά από κόμματα ή συνδέσμους. Έτσι για να συναντήσουμε αντικείμενα περισσότερα του ενός και **να μην έχουμε δίπλωτο ρήμα**, γιατί τότε θα έχουμε ένα άμεσο και ένα έμμεσο αντικείμενο, πρέπει οπωσδήποτε αυτά να ενώνονται με σύνδεσμο ή κόμμα.

Υποστηρίζοντας τη σύνδεση των αντικειμένων, ο αναλυτής αυξάνει την πολυπλοκότητα των εκφράσεων που αναλύει με επιτυχία. Πλέον οι προτάσεις μπορούν να έχουν περισσότερα του ενός αντικείμενα κάτι που πλησιάζει περισσότερο στη φυσική γλώσσα και κατ' επέκταση στον προφορικό λόγο.

Για την εύρεση της σύνδεσης αντικειμένων μεταξύ τους, ο συντακτικός αναλυτής ακολουθεί μια σειρά από βήματα, παρόμοια με αυτή που ακολουθεί για την εύρεση σύνδεσης μεταξύ υποκειμένων. Αυτό είναι λογικό αν αναλογιστούμε ότι τα υποκείμενα και τα αντικείμενα παρουσιάζουν πολλές ομοιότητες. Έτσι λοιπόν:

1. Βρίσκει το πιθανότερο αντικείμενο(ή αντικείμενα αν το ρήμα είναι δίπλωτο)
2. Πηγαίνει προς την αρχή της πρότασης και ελέγχει τι βρίσκεται αμέσως πριν από αυτό και:
 - i. εάν βρει κόμμα και πριν από αυτό πιθανό αντικείμενο τότε ίσως υπάρχει αντικείμενο που προηγείται, ψάχνει όμως απαραίτητα για κάποιο σύνδεσμο που να ενώνει κάποιο πιθανό αντικείμενο με κάποιο επίσης πιθανό αντικείμενο.
 - ii. εάν βρει σύνδεσμο, τότε σίγουρα βρήκε πιθανό αντικείμενο καθώς οι σύνδεσμοι ενώνουν παρόμοια πράγματα **μέσα στην πρόταση** (συνδέουν και προτάσεις μεταξύ τους, αλλά στο σημείο που βρισκόμαστε έχουμε ξεχωρίσει τα όρια των προτάσεων)
 - iii. εάν βρει οτιδήποτε άλλο ολοκληρώνει τον έλεγχο
3. Συνεχίζει τα βήματα 1 και 2 για όλα τα πιθανά αντικείμενα, αποθηκεύοντας κάθε φορά τα πιθανά αντικείμενα που υπακούν στους κανόνες

4. Ολοκληρώνει τον έλεγχο σύνδεσης αντικειμένων

Για να γίνουν περισσότερο κατανοητά τα παραπάνω βήματα θα δώσουμε ένα παράδειγμα. Έστω λοιπόν η πρόταση:

«Ο Κώστας τρέχει στο γήπεδο, στο πάρκο και στο στάδιο »

- i. Βρίσκει το πιθανό αντικείμενο «στάδιο»
- ii. Πηγαίνει προς την αρχή της πρότασης και βρίσκει αμέσως πριν από αυτό (i) σύνδεσμο, τότε σίγουρα προηγείται αντικείμενο, το «πάρκο».
- iii. Αποθηκεύει το «στάδιο» και έχει σαν νέα βάση το αντικείμενο «πάρκο»
- iv. Πηγαίνει προς την αρχή της πρότασης και βρίσκει αμέσως πριν από αυτό (ii) κόμμα, και επειδή έχει ήδη βρει σύνδεσμο προηγουμένως, προηγείται σίγουρα αντικείμενο, το «γήπεδο»
- v. Αποθηκεύει το «πάρκο» και έχει σαν νέα βάση το αντικείμενο «γήπεδο»
- vi. Πηγαίνει προς την αρχή της πρότασης και βρίσκει αμέσως πριν από αυτό (iii) το ρήμα της πρότασης «τρέχει»
- vii. Αποθηκεύει το «γήπεδο» και ολοκληρώνει τον έλεγχο σύνδεσης υποκειμένων

Βλέπουμε λοιπόν ότι με την παραπάνω σειρά βημάτων ο αναλυτής μπορεί να αναγνωρίσει όλες τις πιθανές συνδέσεις μεταξύ αντικειμένων με επιτυχία. Αυτό βοηθάει το συντακτικό αναλυτή να αυξήσει την πολυπλοκότητα των προτάσεων που μπορεί να αναλύσει με επιτυχία καθώς αυτές μπορούν να περιέχουν σύνθετους όρους. Ο αναλυτής θα σπάει αυτούς τους όρους σε μικρότερους σε μια προσπάθεια να τους αναγνωρίσει σωστά

Στη συνέχεια θα δούμε πώς λειτουργεί ο αναλυτής στην περίπτωση των προσδιορισμών που συναντάμε στην ελληνική γλώσσα.

3.6 Ονοματικοί προσδιορισμοί

Όταν ένα ουσιαστικό, επίθετο, μετοχή ή αντωνυμία προσδιορίζουν ως ονόματα άλλα ουσιαστικά με τα οποία δένονται σε σύνολα(τα ονοματικά σύνολα, όπως περιγράφηκαν στην παράγραφο 2.2.6.1), τότε ονομάζονται **ονοματικοί προσδιορισμοί**.

Ένα πλεονέκτημα που κερδίζει ο αναλυτής από τη γνώση των ονοματικών προσδιορισμών είναι η αποφυγή συγχύσεων μεταξύ ονομάτων. Αυτό σημαίνει ότι ένα όνομα μπορεί να είναι, όπως είδαμε και παραπάνω, είτε υποκείμενο είτε αντικείμενο. Οι ονοματικοί προσδιορισμοί είναι επίσης ονόματα. Εάν δεν τους συμπεριλάβουμε στον αναλυτή μας τότε αυτοί οι προσδιορισμοί θα αναγνωρίζονται λανθασμένα από τον αναλυτή ως υποκείμενα ή αντικείμενα.

Η γνώση των πιθανών προσδιορισμών είναι απαραίτητη σε κάθε συντακτικό αναλυτή. Πρέπει ο αναλυτής να μπορεί και να αναγνωρίζει αυτούς τους προσδιορισμούς μέσα στις διάφορες ακολουθίες. Έτσι αυξάνεται και η πολυπλοκότητα των εκφράσεων που μπορεί να αναλύσει. Πλέον, κύριοι συντακτικοί όροι όπως το υποκείμενο και το αντικείμενο θα μπορούν να έχουν επιπλέον ονόματα που λειτουργούν ως προσδιορισμοί.

Επειδή οι ονοματικοί προσδιορισμοί μοιάζουν αρκετά με τους όρους υποκείμενα και αντικείμενα, όσο αφορά τον τρόπο σύνδεσής τους η αντιμετώπιση του αναλυτή είναι ίδια με αυτή της σύνδεσης όρων υποκειμένων και αντικειμένων. Έτσι εδώ δεν θα υπάρξει κάποια ξεχωριστή παράγραφος που να αφορά τη σύνδεση αυτών των όρων.

Για την καλύτερη κατανόηση του τρόπου με τον οποίο χειρίζεται ο αναλυτής τους ονοματικούς προσδιορισμούς ,θα τους διαιρέσουμε σε τρεις κατηγορίες και θα αναλύσουμε την κάθε κατηγορία. Ακολουθεί η ανάλυση των προσδιορισμών ανά κατηγορία.

3.6.1 Ουσιαστικό με άρθρο

Το άρθρο συμφωνεί με το ουσιαστικό στο **γένος** , στον **αριθμό** και στην **πτώση**. Έτσι κατά τη διαδικασία της ανάλυσης, οι λεκτικές μονάδες που χαρακτηρίζονται ως άρθρα συνδυάζονται με τις λέξεις τις οποίες προσδιορίζουν.

Έτσι, ο συντακτικός αναλυτής κάθε φορά που συναντάει άρθρο και έπειτα ουσιαστικό, συνδέει το άρθρο ως προσδιορισμό με το ουσιαστικό. Για παράδειγμα:

«Ο Γιώργος τρέχει »

Το άρθρο συνδέεται και προσδιορίζει το ουσιαστικό. Επομένως αυτό δεν αποτελεί αντικείμενο περαιτέρω συντακτικής ανάλυσης, είναι απλά ένας ονοματικός προσδιορισμός.

Ο συντακτικός αναλυτής μπορεί να επωφεληθεί από αυτής της κατηγορίας τον προσδιορισμό, ιδιαίτερα όταν έπειτα από τη γραμματική ανάλυση μίας λεκτικής μονάδας, υπάρχουν περισσότερες από μία πτώσεις ή/ και αριθμοί ή και γένη . Αυτό μπορεί να συμβεί καθώς ο γραμματικός αναλυτής προσδιορίζει γραμματικά μια-μια τις λεκτικές μονάδες, χωρίς να λαμβάνει υπόψη του άλλες λεκτικές μονάδες γύρω από την προς ανάλυση λεκτική μονάδα. Για παράδειγμα:

«Ο Κώστας πηγαίνει στο σχολείο»

Η γραμματική ανάλυση των δύο τελευταίων λεκτικών μονάδων είναι:

στο – άρθρο / αριθμός: ενικός, γένος: ουδέτερο, πτώση: αιτιατική

σχολείο – ουσιαστικό / αριθμός: ενικός, γένος: ουδέτερο, πτώση: ονομαστική/ αιτιατική/ κλιτική

Εδώ η εφαρμογή της παρατήρησής μας για τη συμφωνία άρθρου και ουσιαστικού παίζει σημαντικό ρόλο. Εάν δεν εφαρμόσουμε την παρατήρηση, τότε τα πιθανά υποκείμενα της πρότασης είναι τα δύο ουσιαστικά που βρίσκονται σε πτώση ονομαστική, δηλαδή το «Κώστας» και το «σχολείο». Αυτό προκαλεί σύγχυση στον αναλυτή και μπορεί να καταστρέψει τη συντακτική ανάλυση όλης της πρότασης. Εφαρμόζοντας όμως την παρατήρηση, το ουσιαστικό «σχολείο» μπορεί να είναι μόνο σε πτώση αιτιατική γιατί προσδιορίζεται από το άρθρο «στο» που προηγείται. Έτσι το ουσιαστικό «σχολείο» είναι σε πτώση αιτιατική και σύμφωνα με αυτά που έχουν ειπωθεί στην παράγραφο 3.4.1 δεν μπορεί να αποτελέσει πιθανό υποκείμενο.

Βλέπουμε λοιπόν πώς οι ονοματικοί προσδιορισμοί μπορούν να διευκολύνουν τον αναλυτή στην πολυπλοκότητα που καμιά φορά εισάγει η γραμματική ανάλυση.

3.6.2 Ουσιαστικό και αντωνυμία

Ένα ουσιαστικό μπορεί να προσδιορίζεται και από αντωνυμία. Σε αυτή την περίπτωση, το ουσιαστικό και η αντωνυμία μοιράζονται την ίδια **πτώση**, τον **αριθμό** και το **γένος**, όπως γίνεται και με το ουσιαστικό με άρθρο. Για παράδειγμα:

«Θέλω εκείνο το **βιβλίο** »

Οι παρατηρήσεις και τα πλεονεκτήματα εδώ είναι τα ίδια με αυτά που αναφέρθηκαν στην προηγούμενη παράγραφο. Πρέπει όμως να τονίσουμε ότι ένα ουσιαστικό μπορεί να έχει περισσότερους από έναν ονοματικούς προσδιορισμούς, όπως στην παραπάνω πρόταση το «εκείνο» και το «το» προσδιορίζουν το βιβλίο. Σε αυτήν την περίπτωση είναι σημαντικό ο συντακτικός αναλυτής να εφαρμόσει σωστά τους κανόνες που διέπουν κάθε κατηγορία προσδιορισμών.

Έτσι, στο προηγούμενο παράδειγμα ο αναλυτής θα πρέπει να βρει πρώτα το άρθρο που είναι άμεσα συνδεδεμένο, δηλαδή δεν μεσολαβεί καμία λεκτική μονάδα ή κόμμα μεταξύ τους, με το ουσιαστικό και σε ένα δεύτερο έλεγχο της πρότασης να βρει την αντωνυμία. Αυτή τώρα πια θα είναι επίσης άμεσα συνδεδεμένη με το ουσιαστικό, καθώς το άρθρο θα έχει παραλειφθεί κατά τον πρώτο έλεγχο για προσδιορισμούς. Αφού λοιπόν είναι άμεσα συνδεδεμένη τότε μπορεί να είναι ονοματικός προσδιορισμός στο ουσιαστικό.

3.6.3 Ουσιαστικό και ουσιαστικό

Το ουσιαστικό μπορεί να σταθεί μόνο τους μέσα σε μία πρόταση, συχνά όμως δεν είναι μόνο του. Συνοδεύεται από ένα άλλο ουσιαστικό, που το προσδιορίζει και συμπληρώνει την έννοιά του.

«*Ηρθε η **Δήμητρα**, η **αδερφή** μου*»(1)

«*Διάβασα το **γράμμα** του **καθηγητή***»(2)

Το δεύτερο αυτό ουσιαστικό(αδερφή, καθηγητή) είναι ονοματικός προσδιορισμός και μπορεί να βρίσκεται στην ίδια πτώση με τη λέξη που προσδιορίζει

(1) και ονομάζεται **ομοιόπτωτος προσδιορισμός** ή να βρίσκεται σε διαφορετική πτώση από τη λέξη που προσδιορίζει (2) και τότε ονομάζεται **ετερόπτωτος προσδιορισμός**. Στη συνέχεια θα αναλυθεί ο τρόπος εντοπισμού από τον αναλυτή αυτών των προσδιορισμών. Η αναλυτική περιγραφή των προσδιορισμών αυτών βρίσκεται στην παράγραφο 2.2.6.1.

3.6.3.1 Ομοιόπτωτος προσδιορισμός

Στον ομοιόπτωτο προσδιορισμό τα δύο ουσιαστικά βρίσκονται στην ίδια πτώση. Οι δύο μορφές με τις οποίες μπορεί να εμφανιστεί ο ομοιόπτωτος προσδιορισμός είναι η *παράθεση* και η *επεξήγηση*. Και οι δύο αυτές μορφές παρουσιάστηκαν στην παράγραφο 2.2.6.1.

Έτσι ο αναλυτής προχωράει στον εντοπισμό αυτής της κατηγορίας προσδιορισμών, βρίσκοντας ουσιαστικά στην ίδια πτώση.

«Ο Νίκος, ο καθηγητής, έφυγε»

Το πρόβλημα που έχει να αντιμετωπίσει με αυτόν τον προσδιορισμό ο αναλυτής είναι να καταλήξει ποιο ουσιαστικό είναι ο προσδιορισμός. Για να το καταφέρει αυτό εκμεταλλεύεται τη σύνταξη των δύο μορφών με τις οποίες ένας ομοιόπτωτος προσδιορισμός μπορεί να εμφανιστεί. Έτσι όταν συναντήσει δύο ουσιαστικά στην ίδια πτώση, και ανάμεσά τους μεσολαβεί κόμμα και **δεν ακολουθεί σύνδεσμος**(*προϋπόθεση για να έχουμε σύνδεση υποκειμένων ή αντικειμένων*), τότε έχουμε ομοιόπτωτο προσδιορισμό. Ο προσδιορισμός είναι το ουσιαστικό που βρίσκεται μετά το κόμμα(όπως αναφέρεται στη σχετική θεωρία) και μπορεί να είναι παράθεση, όπως στην προηγούμενη πρόταση το ουσιαστικό «καθηγητής», ή μπορεί να είναι επεξήγηση όπως στην πρόταση:

«Η θεά της ομορφιάς, η Αφροδίτη, βγήκε από τη θάλασσα »

είναι το ουσιαστικό «Αφροδίτη» ονοματικός προσδιορισμός στο ουσιαστικό «θεά».

Βλέπουμε λοιπόν ότι η πολύ καλή γνώση της συντακτικής μορφής των προσδιορισμών μπορεί να διευκολύνει σημαντικά το συντακτικό μας αναλυτή.

3.6.3.2 Ετερόπρωτος προσδιορισμός

Στην κατηγορία των ετερόπρωτων προσδιορισμών, το ουσιαστικό – προσδιορισμός μπορεί να βρίσκεται σε πτώση γενική(1) ή αιτιατική(2), **απαραίτητα** όμως σε διαφορετική πτώση από το ουσιαστικό που προσδιορίζει.

«*Τα **μάτια** του Κώστα ήταν φοβισμένα* » (1)

«*Είναι ένα **χαλί** τέσσερα μέτρα μάκρος* » (2)

Εδώ ο αναλυτής έχει να αντιμετωπίσει το εξής πρόβλημα: δεν κινδυνεύει να αναγνωρίσει τον προσδιορισμό ως αντικείμενο/ υποκείμενο που ενώνεται με το ουσιαστικό που είναι αντικείμενο/ υποκείμενο καθώς δε μεσολαβεί τίποτα ανάμεσα στα δύο ουσιαστικά, αλλά γι' αυτό ακριβώς, επειδή δε μεσολαβεί τίποτα ανάμεσά τους, κινδυνεύει να αναγνωρίσει το ένα ως υποκείμενο και το άλλο ως αντικείμενο.

Για να λυθεί και αυτό το πρόβλημα, ο αναλυτής εκμεταλλεύεται και πάλι τη σύνταξη. Εάν βρει δύο ουσιαστικά χωρίς να μεσολαβεί κάτι ανάμεσά τους (μάτια/ Κώστα και χαλί/ τέσσερα στις παραπάνω προτάσεις) και το δεύτερο στη σειρά είναι σε πτώση γενική(Κώστα) ή αιτιατική (τέσσερα) τότε αυτό είναι ετερόπρωτος προσδιορισμός στο ουσιαστικό που προηγείται(μάτια και χαλί αντίστοιχα).

3.6.4 Επίθετο και ουσιαστικό

Ένα ουσιαστικό μέσα σε μία πρόταση μπορεί να βοηθηθεί ιδιαίτερα από ένα επίθετο που θα το προσδιορίζει. Όταν ένα ουσιαστικό προσδιορίζεται από ένα επίθετο, τότε αυτό αποκτά ειδικότερη σημασία και καθορίζεται ακριβέστερα. Για παράδειγμα:

«*Μπορείς να μου φέρεις το γαλάζιο βιβλίο ;* »

Στην παραπάνω πρόταση ο προσδιορισμός «γαλάζιο» μας βοηθάει να καταλάβουμε για ποιο βιβλίο μιλάμε, δίνει δηλαδή στο ουσιαστικό «βιβλίο» μία έννοια – χαρακτηριστικό και το κάνει ποιο συγκεκριμένο. Περισσότερες πληροφορίες

υπάρχουν στην αντίστοιχη θεωρία που αναπτύχθηκε στην παράγραφο 2.2.6.1 Στη συνέχεια θα αναλύσουμε πώς αυτό το θεωρητικό υπόβαθρο θα βοηθήσει τον αναλυτή να εντοπίσει αυτούς τους προσδιορισμούς, και ιδιαίτερα την σημαντικότερη κατηγορία τους, τους επιθετικούς προσδιορισμούς, μέσα στις προτάσεις.

3.6.4.1 Επιθετικός προσδιορισμός

Στον επιθετικό προσδιορισμό συναντάμε επίθετα τα οποία προσδιορίζουν ουσιαστικά και συμφωνούν με αυτά στην **πτώση**, το **γένος** και τον **αριθμό**. Αυτά τα επίθετα ονομάζονται επιθετικοί προσδιορισμοί και φανερώνουν μία μόνιμη ιδιότητα του ουσιαστικού το οποίο προσδιορίζουν. Για παράδειγμα:

«Ο ψηλός αρχηγός παραιτήθηκε »

Ένα ουσιαστικό μπορεί να έχει περισσότερους από έναν επιθετικούς προσδιορισμούς. Σε αυτή την περίπτωση είναι δυνατόν:

- i. Ο ένας από τους δύο να προσδιορίζει την κοινή έννοια, την οποία αποτελούν μαζί το ουσιαστικό και το άλλο επίθετο (1) ή
- ii. Και οι δύο (ή και περισσότεροι) επιθετικοί προσδιορισμοί να προσδιορίζουν χωριστά ο καθένας το ουσιαστικό (2)

« Ασχολείσαι με ασήμαντες μικρές υποθέσεις »(1)

«Ήταν σοβαρός και ευγενικός άνθρωπος »(2)

Για να είναι σε θέση να αναγνωρίσει τους επιθετικούς προσδιορισμούς, ο συντακτικός αναλυτής θα πρέπει να γνωρίζει τι μπορεί να αποτελέσει επιθετικό προσδιορισμό. Έτσι λοιπόν, ως επιθετικοί προσδιορισμοί, πέρα από τα επίθετα, χρησιμοποιούνται:

1. *Αντωνυμίες*: **Τόση** χαρά δεν την αντέχω
2. *Αριθμητικά*: **Χίλιοι** οπαδοί ακολούθησαν την ομάδα
3. *Μετοχές*: Το **πληγωμένο** ελάφι ήταν θλιβερή εικόνα
4. *Ουσιαστικά*: **Γέρος** άνθρωπος και δε λέει να αποσυρθεί
5. *Ονόματα*: Φτάσαμε **Κυριακή** μεσημέρι

6. *Επιρρήματα*: Οι **κάτω** χώρες είναι ανάστατες
7. *Εμπρόθετα*: Πήρε άδεια **για μια εβδομάδα**

Έτσι όταν ο συντακτικός αναλυτής εντοπίσει κάτι από τα παραπάνω να ενώνεται άμεσα με ένα ουσιαστικό και συμφωνεί με το ουσιαστικό στην πτώση, το γένος και τον αριθμό, τότε αντιλαμβάνεται ότι αυτό είναι επιθετικός προσδιορισμός.

Σχετικά με τη σύνδεση των επιθετικών προσδιορισμών, όπως και όλων των προσδιορισμών, ο αναλυτής ακολουθεί την ίδια λογική με αυτή της σύνδεσης των υποκειμένων και των αντικειμένων.

ΚΕΦΑΛΑΙΟ 4

Υλοποίηση συντακτικού αναλυτή

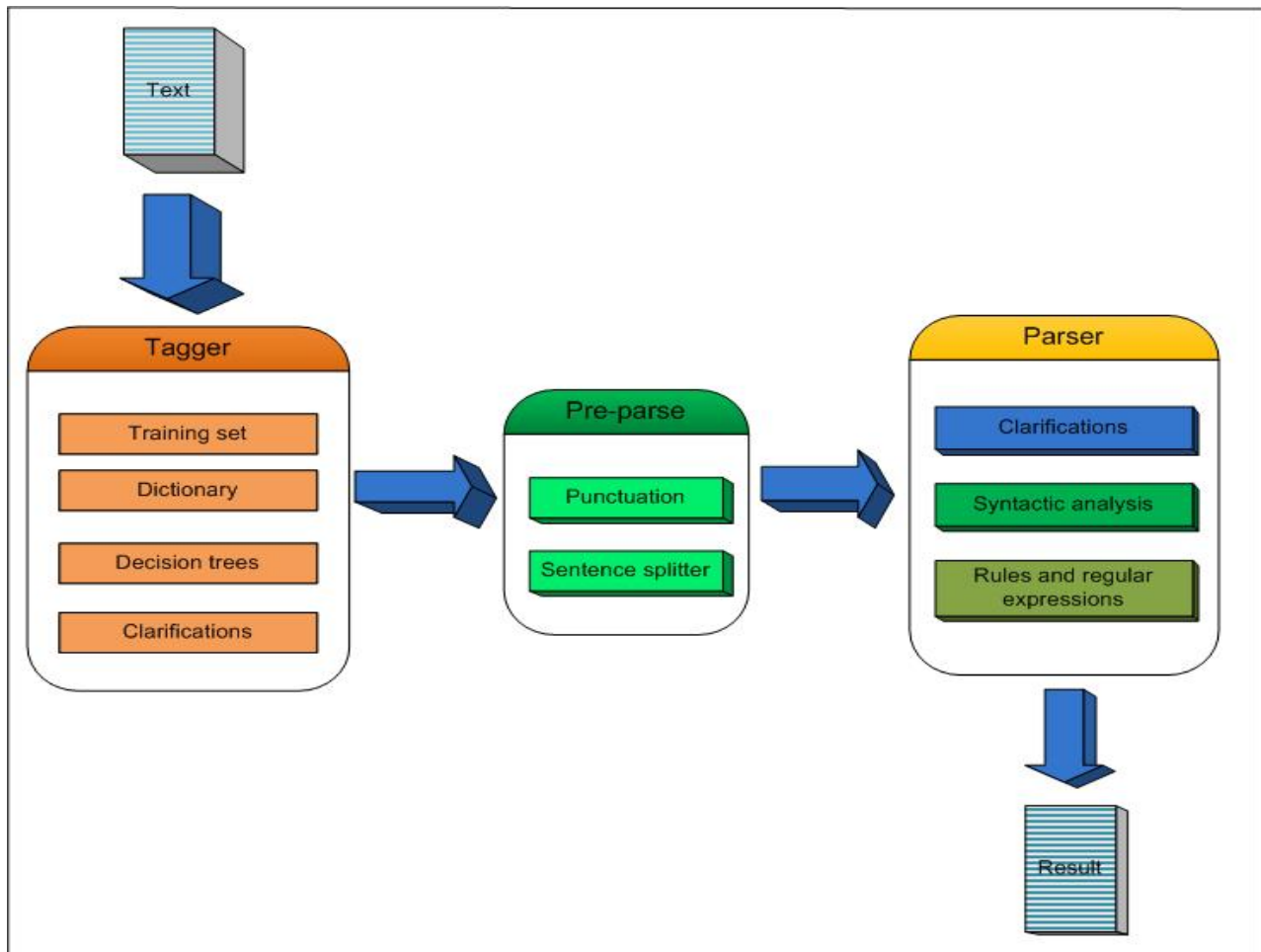
4.1 Εισαγωγή

Η υλοποίηση ενός συντακτικού αναλυτή για την ελληνική γλώσσα απαιτεί γνώση όλων των παραπάνω συντακτικών κανόνων και σωστό συνδυασμό τους. Επίσης πρέπει να δοθεί ιδιαίτερη προσοχή στην αρχιτεκτονική που θα σχεδιαστεί για να υλοποιηθούν οι διάφορες μονάδες.

Σε αυτό το κεφάλαιο θα περιγράψουμε την αρχιτεκτονική που σχεδιάστηκε για την υλοποίηση του συντακτικού αναλυτή. Κάθε μονάδα αυτής της αρχιτεκτονικής θα περιγραφεί χωριστά και αναλυτικά ενώ συχνά θα γίνεται και σύνδεση με το κεφάλαιο 3 το οποίο και περιέχει την βασική θεωρία την οποία υλοποιεί κάθε μονάδα της αρχιτεκτονικής.

4.1 Γενική αρχιτεκτονική

Η γενική αρχιτεκτονική τους συντακτικού αναλυτή που υλοποιήθηκε είναι η ακόλουθη:



Εικόνα 11: Η αρχιτεκτονική του συντακτικού αναλυτή

Όπως παρατηρούμε και στην παραπάνω εικόνα, η αρχιτεκτονική αποτελείται από τρεις βασικές μονάδες: **τη μονάδα του μορφοσυντακτικού σχολιαστή (tagger), τη μονάδα σύνδεσης – απομόνωσης (pre – parse) και τη μονάδα του συντακτικού σχολιαστή (parser)**.

Ένα κείμενο ή μια συλλογή κειμένων ακολουθεί την παρακάτω σειρά βημάτων προκειμένου να αναλυθεί συντακτικά:

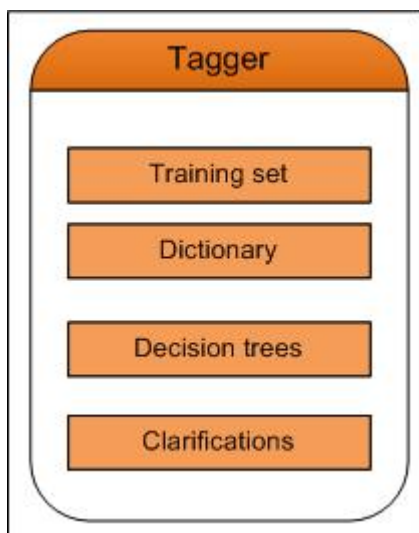
- i. Ένα κείμενο ή μία συλλογή κειμένων (text) εισέρχεται στο συντακτικό αναλυτή Αυτό προωθείται στο μορφοσυντακτικό αναλυτή

- ii. Αυτός κάνει τη γραμματική ανάθεση σε όρους στο κείμενο
- iii. Το μορφοποιημένο κείμενο προωθείται στη μονάδα σύνδεσης – απομόνωσης
- iv. Γίνεται κατακερματισμός του κειμένου σε προτάσεις και αφαιρούνται ορισμένες λεκτικές μονάδες που κρίνονται μη απαραίτητες
- v. Το κείμενο μορφοποιείται κατάλληλα έτσι ώστε να είναι σε θέση στο επόμενο βήμα ο συντακτικός σχολιαστής να κάνει συντακτική ανάλυση.
- vi. Το κείμενο εισάγεται στο συντακτικό σχολιαστή
- vii. Γίνεται η τελική συντακτική ανάλυση του κειμένου, περνώντας το από διάφορες υπομονάδες.
- viii. Στο τέλος το κείμενο εμφανίζεται και πάλι έχοντας τη συντακτική ανάλυση όλων των όρων που περιέχει.

Ακολουθεί η αναλυτική περιγραφή κάθε μονάδας και των υπομονάδων της ενώ στο τέλος παρουσιάζεται και ένα παράδειγμα ροής (system flow)του συστήματος.

4.2 Η μονάδα του μορφοσυντακτικού σχολιαστή

Η μονάδα του μορφοσυντακτικού σχολιαστή υλοποιήθηκε το 2000 στο Πανεπιστήμιο Πατρών από τον **κ.Ορφανό Γιώργο** στα πλαίσια της διδακτορικής του διατριβής με θέμα «Υπολογιστική Μορφοσυντακτική Ανάλυση της Νέας Ελληνικής».Η περιγραφή της μονάδας υπάρχει στην παράγραφο 2.4. Εδώ επειδή χρησιμοποιείται σαν εξωτερική εφαρμογή από τον συντακτικό αναλυτή που υλοποιήσαμε, θα αναφέρουμε περιληπτικά κάποια στοιχεία που αφορούν τη χρήση της μονάδας αυτής από την εφαρμογή μας.



Εικόνα 12: Η μονάδα του μορφοσυντακτικού αναλυτή

Για την υλοποίηση λοιπόν αυτού του συντακτικού αναλυτή, χρησιμοποιήθηκαν τεχνικές μηχανικής μάθησης, νευρωνικά δίκτυα και δέντρα αποφασήσεων. Το σώμα για την εκπαίδευση της εφαρμογής αποτελούταν από 7.624 προτάσεων (137.764 λέξεις, σημεία στίξης, αριθμοί κτλ.), που προέρχονται από απομαγνητοφωνημένο προφορικό λόγο, λογοτεχνικά, δημοσιογραφικά και επιστημονικά κείμενα, ως χαρακτηριστικά δείγματα σύγχρονου νεοελληνικού λόγου. Το λεξικό που χρησιμοποιείται αναπτύχθηκε στο ΙΤΥ. Η ακρίβεια της εφαρμογής αγγίζει το ~95% (ασάφεια επιθέτου – ουσιαστικού – επιρρήματος ~70% ακρίβεια).

Η λειτουργία αυτής της μονάδας είναι η εξής: Παίρνει σαν είσοδο ένα αρχείο κειμένου και με κωδικοποίηση ελληνική. Στη συνέχεια το περνάει από διάφορα στάδια ανάλυσης ενώ συμβουλευεται και ένα λεξικό της ελληνικής γλώσσα. Στο τέλος της διαδικασίας έχει κάνει γραμματική ανάλυση όλων των όρων του κειμένου.

Ο συγκεκριμένος αναλυτής έχει ένα σύνολο πλεονεκτημάτων τα οποία τον έκαναν να υπερτερεί έναντι άλλων μορφοσυντακτικών αναλυτών που υπάρχουν για την ελληνική γλώσσα. Τα πιο σημαντικά από αυτά είναι τα ακόλουθα:

- i. Έχει ένα αρκετά μεγάλο training set σε σχέση με άλλους αναλυτές. Ενδεικτικά αναφέρουμε πως η εφαρμογή *ellogon* που υποστηρίζει και γραμματική ανάθεση σε όρους, είχε training set περίπου 1500 λέξεις
- ii. Το training set είναι βασισμένο σε απομαγνητοφωνημένο προφορικό λόγο, λογοτεχνικά, δημοσιογραφικά και επιστημονικά κείμενα, ως χαρακτηριστικά δείγματα σύγχρονου νεοελληνικού λόγου. Αυτό σημαίνει

οτι θα είναι πολύ κοντά στις απαιτήσεις ενός συντακτικού αναλυτή φυσικής γλώσσας, ενώ αποφεύγεται και ο κίνδυνος η εκπαίδευση του αναλυτή να έχει γίνει σε ένα και μόνο συγκεκριμένο γνωστικό αντικείμενο.

- iii. Τα αποτελέσματα του αναλυτή είναι ακριβέστερα από πολλές άλλες παρόμοιες εφαρμογές
- iv. Το εύρος των γραμματικών αναθέσεων του είναι μεγάλο και περιέχει όρους που δεν περιέχουν άλλοι αναλυτές (αντωνυμίες, επιφωνήματα κ.ά.). Αυτό μας δίνει τη δυνατότητα να χρησιμοποιήσουμε στο μέγιστο δυνατό βαθμό τη συντακτική γνώση που έχουμε και αναλύσαμε στα προηγούμενα κεφάλαια, καθώς μπορούμε να εντοπίσουμε όλους τους πιθανούς γραμματικούς όρους που συνδέονται με συντακτικούς κανόνες.
- v. Μας προσφέρει επιπλέον εξαγωγή χρήσιμων γραμματικών πληροφοριών (πτώση, γένος κ.ά.) οι οποίες είναι ιδιαίτερα χρήσιμες, όπως φάνηκε και στην αντίστοιχη θεωρία, όταν είναι διαθέσιμες.
- vi. Είναι μία stand alone εφαρμογή. Έτσι μπορούμε να τη χρησιμοποιούμε ως εξωτερική εφαρμογή χωρίς να απαιτείται κάποιος ιδιαίτερος χειρισμός για να ενσωματωθεί στην εφαρμογή μας.
- vii. Ο συγκεκριμένος μορφοσυντακτικός αναλυτής είναι ελεύθερα διαθέσιμος για χρήση

Όλα τα παραπάνω πλεονεκτήματα μας έκαναν να επιλέξουμε το συγκεκριμένο μορφοσυντακτικό αναλυτή για την εφαρμογή μας και να επωφεληθούμε από τα πολύ καλά αποτελέσματά του.

Η έξοδος αυτής της μονάδας είναι ένα αρχείο κειμένου, το αρχικό, το οποίο όμως περιέχει δίπλα σε κάθε λεκτική του μονάδα τη γραμματική της ανάλυση. Το αποτέλεσμα μπορεί να έχει τη μορφή αρχείου xml.

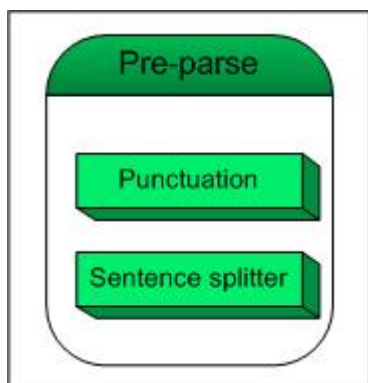
Το αρχείο xml αποτελείται από ένα σύνολο ετικετών. Οι ετικέτες <p>...</p> καθορίζουν τα όρια των παραγράφων. Οι ετικέτες <s>...</s> καθορίζουν τα όρια των προτάσεων. Οι ετικέτες <rh>...</rh> καθορίζουν τα όρια, τις κατηγορίες (μέσω της ιδιότητας cat) και τα μορφοσυντακτικά χαρακτηριστικά (μέσω της ιδιότητας attrs) των φράσεων που αναγνωρίστηκαν κατά το δεύτερο επίπεδο αποσαφήνισης. Οι ετικέτες <gw>...</gw> καθορίζουν τα όρια, το ΜτΛ (μέσω της ιδιότητας cat), τα λοιπά μορφοσυντακτικά χαρακτηριστικά (μέσω της ιδιότητας attrs) και τη

μορφολογική προέλευση (μέσω της ιδιότητας *lexeme*) των ελληνικών λέξεων. Οι ετικέτες `<fw>...</fw>` καθορίζουν τα όρια των ξένων λέξεων. Οι ετικέτες `<punc>...</punc>` καθορίζουν τα όρια των σημείων στίξης. Οι ετικέτες `<num>...</num>` καθορίζουν τα όρια των αριθμητικών συμβόλων. Τέλος, οι ετικέτες `<abbr>...</abbr>` καθορίζουν τα όρια των συντμήσεων.

4.3 Η μονάδα σύνδεσης – απομόνωσης

Η μονάδα σύνδεσης – απομόνωσης είναι η μονάδα που μεσολαβεί ανάμεσα στον μορφοσυντακτικό και στον συντακτικό σχολιαστή, μετά τον *tagger* και πριν τον *parser*. Η λειτουργία της είναι διπλή: συνδέει τους δύο σχολιαστές για να είναι εφικτή η συντακτική ανάλυση όμως ταυτόχρονα μπορεί να απομονώσει το συντακτικό σχολιαστή από το μορφοσυντακτικό σχολιαστή πετυχαίνοντας έτσι την ανεξαρτησία του συντακτικού αναλυτή από συγκεκριμένους μορφοσυντακτικούς σχολιαστές. Με τη μεσολάβηση δηλαδή αυτής της απομονωτικής μονάδας και με ελάχιστες αλλαγές στον τρόπο λειτουργίας της, μπορούμε να συνδέσουμε το συντακτικό μας σχολιαστή με ένα οποιονδήποτε μορφοσυντακτικό σχολιαστή. Φέρνει δηλαδή το κείμενο που βγάζει ο οποιοσδήποτε *tagger* στην μορφή που διαβάζεται από τον *parser*. Η μορφή αυτή είναι ένα αρχείο *xml* με τις ίδιες ετικέτες με αυτές που περιγράφηκαν στην προηγούμενη παράγραφο αλλά περαιτέρω συντακτική ανάλυση. Έτσι τώρα το αρχείο μπορεί να περιέχει περισσότερες προτάσεις και λιγότερα σημεία στίξης.

Φυσικά ο συντακτικός αναλυτής που θα προκύψει δε θα εμφανίζει τα ίδια αποτελέσματα, οι όποιες αλλαγές όμως θα οφείλονται αποκλειστικά και μόνο στην ακρίβεια των αποτελεσμάτων του μορφοσυντακτικού αναλυτή.



Εικόνα 13: Η μονάδα σύνδεσης – απομόνωσης

Πέρα από τις δύο παραπάνω σημαντικές λειτουργίες, η συγκεκριμένη μονάδα διαθέτει και δύο υπομονάδες που προωθούν την συντακτική ανάλυση. Αυτές οι δύο μονάδες είναι η υπομονάδα αφαίρεσης συμβόλων και η υπομονάδα εύρεσης και διαχωρισμού προτάσεων οι οποίες και θα παρουσιαστούν στις επόμενες παραγράφους.

4.3.1 Η υπομονάδα αφαίρεσης σημείων στίξης (punctuation)

Αυτή είναι η πρώτη υπομονάδα από την οποία διέρχεται το κείμενο, αφού έχει περάσει από γραμματική ανάλυση. Εδώ γίνεται λεπτομερής έλεγχος όλων των σημείων στίξης που εμφανίζονται στο κείμενο. Τα σημεία στίξης και η κρισιμότητά τους στην ορθή λειτουργία ενός συντακτικού αναλυτή περιγράφηκαν στην παράγραφο 3.2.4.

Η έννοια του αφαιρετή σημείων στίξης είναι ιδιαίτερα διαδεδομένη και χρησιμοποιείται συχνά σε εφαρμογές που έχουν να κάνουν με γραπτό λόγο. Σε κάθε εφαρμογή όμως ο διαχειριστής οφείλει να προσαρμόσει ποια σημεία στίξης θα αφαιρεθούν, σύμφωνα με τις ανάγκες της εφαρμογής. Ένας κοινός αφαιρετής πέρα από τα κενά και διάφορους χαρακτήρες όπως @ ή « » , που δεν προσδίδουν επιπλέον πληροφορία, θα αφαιρούσε και άλλα σημεία στίξης, απαραίτητα για τη συντακτικής μας ανάλυση, όπως το κόμμα (,), το ερωτηματικό (;) και η τελεία (.). Οι λόγοι που είναι απαραίτητα αυτά τα σημεία στίξης σε έναν συντακτικό αναλυτή αναφέρθηκαν παραπάνω στην παράγραφο 3.2.4.

Στην υλοποίησή μας τα σημεία στίξης που αφαιρούνται είναι τα ακόλουθα:

"(", ") ", " * ", " + ", " - ", " : ", " < ", " = ", " > ", " ? ", " @ ", " « ", " » "

ενώ διατηρούνται κάποια άλλα ιδιαίτερα σημαντικά για τη συνέχεια της συντακτικής ανάλυσης ,όπως το κόμμα (,), το ερωτηματικό (;) ,η τελεία (.) και το θαυμαστικό (!).

4.3.2 Η υπομονάδα εύρεσης και διαχωρισμού προτάσεων (sentence splitter)

Αφού ολοκληρωθεί ο έλεγχος για τα σημεία στίξης , το κείμενο προωθείται στην υπομονάδα εύρεσης και διαχωρισμού προτάσεων. Και αυτή η υπομονάδα έχει διπλή λειτουργία: Βρίσκει τις προτάσεις που υπάρχουν μέσα στο κείμενο και διαχωρίζει τις εκφράσεις εκείνες που αποτελούνται από περισσότερες από μια προτάσεις συνδεδεμένες με κόμμα ή με σύνδεσμο. Η βασική λειτουργία δηλαδή του sentence splitter είναι να βρει τον ακριβή αριθμό των προτάσεων που συνθέτουν το προς ανάλυση κείμενο

Ο πρώτος κερματισμός του κειμένου σε προτάσεις στηρίζεται στον ορισμό της πρότασης (παρ. 2.2.1) και στα σημεία στίξεως που δηλώνουν το τέλος μίας πρότασης (παρ. 3.2.4). Έτσι, αυτή η υπομονάδα βρίσκει και ορίζει ως προτάσεις, τις λεκτικές ακολουθίες του κειμένου, που βρίσκονται μεταξύ δύο σημείων στίξεως που δηλώνουν αρχή ή τέλος πρότασης. Αυτή η τεχνική δεν είναι λάθος αλλά δύο ή περισσότερες προτάσεις που είναι συνδεδεμένες μεταξύ τους με κόμμα ή με συνδέσμους (παρ. 2.2.4) θα αναγνωριστούν ως μία ενιαία και μεγάλη πρόταση. Για να διορθώσουμε αυτό το πρόβλημα αφού ολοκληρωθεί ο πρώτος έλεγχος για προτάσεις, προχωράμε σε δεύτερο έλεγχο, αυτή τη φορά όχι του κειμένου αλλά των προτάσεων που προέκυψαν από τον πρώτο έλεγχο.

Στον δεύτερο έλεγχο προχωράμε σε πιθανό κερματισμό προτάσεων. Ελέγχουμε τον αριθμό των ρημάτων που έχει κάθε πιθανή πρόταση, ο οποίος όπως ξέρουμε από την θεωρία θα υποδηλώνει και τον αριθμό των προτάσεων. Ουσιαστικά ο αριθμός των προτάσεων είναι ίδιος με τον αριθμό των ρημάτων καθώς κάθε ρήμα αντιστοιχεί σε μία και μόνο πρόταση. Εάν σε μία πιθανή πρόταση βρεθούν περισσότερα από ένα ρήματα τότε έχουμε περισσότερες προτάσεις και επομένως σύνδεση προτάσεων. Έπειτα προχωράμε στον εντοπισμό της σύνδεσης ,όπως αυτός

αναφέρθηκε στη θεωρία, και του τρόπου σύνδεσης των προτάσεων(με κόμμα ή με σύνδεσμο).

Μετά το πέρασμα και από αυτή την υπομονάδα πλέον το κείμενο έχει κατακερματιστεί και δουλεύουμε με τις προτάσεις του και όχι με όλο το κείμενο.

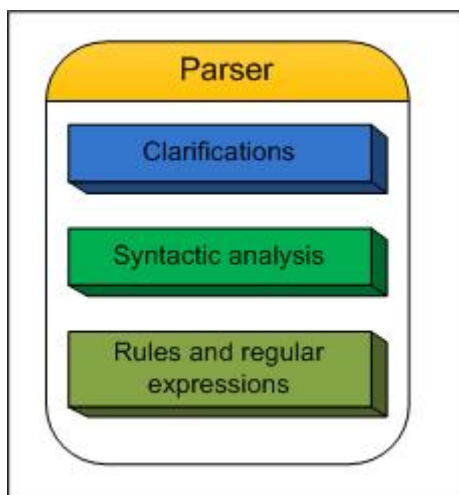
Η έξοδος αυτής της μονάδας είναι ένα αρχείο κειμένου όπου κάθε λεκτική μονάδα έχει και τη γραμματική της ανάλυση, έχουν απομείνει στο κείμενο μόνο οι λεκτικές μονάδες που έχουν νόημα να αναλυθούν συντακτικά (οι λεκτικές μονάδες) ή έχουν να μας δώσουν κάποια συντακτική πληροφορία (τα εναπομείναντα σημεία στίξης). Το επόμενο βήμα είναι η είσοδος αυτού του κειμένου στο συντακτικό σχολιαστή, ο οποίος παρουσιάζεται στη συνέχεια.

4.4 Η μονάδα του συντακτικού σχολιαστή

Η μονάδα του συντακτικού αναλυτή είναι η τελευταία μονάδα από την οποία θα περάσει το προς ανάλυση κείμενο πριν να ολοκληρωθεί η συντακτική του ανάλυση. Αυτή η μονάδα είναι η πιο κρίσιμη της αρχιτεκτονικής καθώς αυτή είναι υπεύθυνη για το τελικό αποτέλεσμα της ανάλυσης.

Πριν όμως φθάσει εδώ το κείμενο έχει περάσει από άλλες δύο μονάδες. Στη μονάδα του μορφοσυντακτικού αναλυτή έγινε γραμματική ανάλυση των όρων του κειμένου. Στην επόμενη μονάδα, αυτή της σύνδεσης – απομόνωσης , το κείμενο κατακερματίστηκε σε προτάσεις ενώ του αφαιρέθηκαν και όποιοι όροι δεν έπρεπε να αναλυθούν συντακτικά ή δεν του προσέδιδαν συντακτική πληροφορία. Επίσης αυτή η μονάδα μορφοποίησε το κείμενο κατάλληλα για να μπορεί αυτό να αναγνωριστεί από τη μονάδα του συντακτικού σχολιαστή.

Ο συντακτικός σχολιαστής για να λειτουργήσει χρειάζεται το κείμενο να βρίσκεται σε μία συγκεκριμένη μορφή. Γι' αυτό έχει φροντίσει η προηγούμενη μονάδα, αυτή της σύνδεσης – αποσύνδεσης. Ο σχολιαστής δουλεύει με προτάσεις στις οποίες δίπλα από κάθε όρο υπάρχει η γραμματική του ανάλυση. Για τις προτάσεις και την τελική μορφή του κειμένου πριν φθάσει εδώ φροντίζει η μονάδα σύνδεσης – απομόνωσης ενώ για τη γραμματική ανάλυση υπεύθυνος είναι ο μορφοσυντακτικός σχολιαστής. Έτσι, λάθη στο τελικό αποτέλεσμα, στο διαχωρισμό προτάσεων ή λάθη που οφείλονται σε λανθασμένη γραμματική αναγνώριση, πρέπει να αναζητηθούν στις αντίστοιχες μονάδες και όχι στο συντακτικό σχολιαστή.



Εικόνα 14: Ο συντακτικός σχολιαστής

Ο συντακτικός σχολιαστής υλοποιήθηκε σε 3 επίπεδα. Αυτά κρίθηκαν αναγκαία μετά από μελέτη του συνόλου των κανόνων και την ανάγκη αποσαφήνισης πολλών πραγμάτων, πριν την τελική εφαρμογή των κανονικών εκφράσεων. Σε αυτά τα επίπεδα έχουμε:

- i. Στο πρώτο επίπεδο(clarifications) γίνεται μία πρώτη ομαδοποίηση των δεδομένων
- ii. Στο δεύτερο επίπεδο(syntactic analysis) γίνεται εύρεση και ανάθεση των βασικών συντακτικών όρων στις ομάδες λέξεων που προέκυψαν από το προηγούμενο επίπεδο
- iii. Στο τρίτο επίπεδο(rules and regular expressions) γίνεται η τελική ανάθεση των όρων με βάση το είδος της πρότασης που θέλουμε να υποστηρίξουμε

Στις επόμενες ενότητες θα παρουσιάσουμε κάθε επίπεδο χωριστά και θα αναλύσουμε τι ακριβώς κάνει και ποιο κομμάτι της θεωρίας του συντακτικού αναλυτή υλοποιεί.

4.4.1 Η υπομονάδα γραμματικής αποσαφήνισης (clarifications)

Αυτή η υπομονάδα βρίσκεται στο πρώτο επίπεδο της σχεδίασης του συντακτικού σχολιαστή. Αυτό που κάνει είναι να προχωράει σε μια πρώτη αποσαφήνιση των λεκτικών μονάδων της προς ανάλυση ακολουθίας.

Η μόνη πληροφορία που είναι διαθέσιμη σε αυτό το επίπεδο για την κάθε λεκτική μονάδα, είναι η γραμματική της ανάλυση. Έτσι η υπομονάδα, για να προχωρήσει στην αποσαφήνιση πρέπει να στηριχθεί σε όση γνώση μπορεί να εξάγει από τη γραμματική ανάλυση. Η σύνδεση γραμματικής ανάλυσης και συντακτικών όρων έχει περιγραφεί σε διάφορα σημεία της παραγράφου 2.2. Το αποτέλεσμα αυτής της σύνδεσης είναι ένα σύνολο κανόνων που εφαρμόζονται στις λεκτικές μονάδες και έτσι γίνεται μια πρώτη αποσαφήνιση.

Οι κανόνες που προκύπτουν αφορούν πάντα κατηγορίες γραμματικών ορισμών, ανεξάρτητα από το ρόλο του ορισμού στην συγκεκριμένη πρόταση. Έχουν συλλεχθεί μέσα από ανάγνωση διαφορετικών συντακτικών της ελληνικής γλώσσας και είναι όσο το δυνατόν πιο περιεκτικοί. Δεν είναι αμφισβητήσιμοι, είναι επίσημες παραδοχές του συντακτικού της ελληνικής γλώσσας.

Ουσιαστικά αυτό που γίνεται σε αυτή την υπομονάδα είναι γραμματικοί όροι όπως ουσιαστικό, επίθετο, αντωνυμία κ.ά, να αντιστοιχίζονται στους πιθανούς συντακτικούς όρους, όπως υποκείμενο, ρήμα, αντικείμενο και κατηγορούμενο, στηριζόμενοι για την αντιστοίχια στους κανόνες που αναφέρθηκαν παραπάνω. Οι κανόνες αντιστοίχισης που αφορούν κάθε συντακτικό όρο έχουν περιγραφή στην αντίστοιχη παράγραφο στο κεφάλαιο 3.

Ακολουθούν μερικά χαρακτηριστικά παραδείγματα κανόνων για την υλοποίηση της αποσαφήνισης που λαμβάνει χώρα σε αυτήν την υπομονάδα:

Παράδειγμα 1: Ουσιαστικό, αποσαφήνιση: πιθανή συντακτική αντιστοίχιση

π.χ. «Ο Νίκος πήγε σχολείο» πιθανά υποκείμενα: Νίκος/ σχολείο
πιθανά αντικείμενα: Νίκος/ σχολείο

Παράδειγμα 2:

Ουσιαστικό με άρθρο, αποσαφήνιση: Προσδιορίζει ή όχι το άρθρο το ουσιαστικό

π.χ «Ο ήλιος δύνει»

Παράδειγμα 3:

Ουσιαστικό και ουσιαστικό, αποσαφήνιση: Ομοιόπτωτος ή ετερόπτωτος προσδιορισμός

π.χ « Ο Κώστας, ο αδερφός μου, έφυγε » -- παράθεση
« Ο αδερφός μου, ο Κώστας, έφυγε » -- επεξήγηση
« Τα μάτια του Γιάννη ήταν δακρυσμένα » -- ετερόπτωτος

Παράδειγμα 4:

Επίθετο και ουσιαστικό, αποσαφήνιση: Επιθετικός ή κατηγορηματικός προσδιορισμός

π.χ « Σκόραρε ο Γιώργος ο ψηλός ; » -- επιθετικός προσδ.
« Προχωρούσε με υψωμένα τα χέρια » -- κατηγορηματικός προσδ.

Τα αποτελέσματα αυτής της υπομονάδας προωθούνται στην επόμενη υπομονάδα, μέσω του ίδιου αρχείου xml, για το δεύτερο επίπεδο συντακτικού σχολιασμού.

4.4.2 Η υπομονάδα συντακτικής αποσαφήνισης (syntactic anaysis)

Αυτή η υπομονάδα υλοποιεί το δεύτερο επίπεδο αποσαφήνισης. Δέχεται ως είσοδο τις λεκτικές μονάδες και την πιθανή συντακτική τους ανάλυση και προχωρά σε περαιτέρω αποσαφήνιση. Αυτή η αποσαφήνιση συντελείται κάτω από ορισμένους κανόνες του πρέπει να πληρούν οι συντακτικοί όροι.

Οι κανόνες αυτοί αφορούν τις βασικές κατηγορίες των συντακτικών όρων, δηλαδή το υποκείμενο, το ρήμα, το αντικείμενο και το κατηγορούμενο που αναζητούμε κάθε φορά. Είναι συμπληρωματικοί με τους προηγούμενους κανόνες όμως έχουν μια πολύ σημαντική διαφορά. Εδώ οι κανόνες εφαρμόζονται και αφορούν συντακτικούς και όχι γραμματικούς όρους.

Υπάρχουν εξαιρέσεις σε αυτούς τους κανόνες αλλά γενικά περιγράφουν τις βασικές αρχές που διέπουν την ελληνική γλώσσα

Ακολουθούν μερικά χαρακτηριστικά παραδείγματα κανόνων για την υλοποίηση της αποσαφήνισης που λαμβάνει χώρα σε αυτήν την υπομονάδα:

Παράδειγμα 1:

Αποσαφήνιση υποκειμένου: «...το υποκείμενο είτε σύναρθρο είτε άναρθρο τοποθετείται σε πτώση ονομαστική...» Συντακτικό της Νέας Ελληνικής, σελ.16

π.χ: « Ο καθηγητής γράφει »

Παράδειγμα 2:

Αποσαφήνιση ρήματος υποκειμένου: «...σε μία πρόταση που έχει ένα υποκείμενο το ρήμα συμφωνεί με το υποκείμενό του στο πρόσωπο και στον αριθμό...» Συντακτικό της Νέας Ελληνικής,σελ.23

π.χ: «Οι γεωργοί οργώνουν το χωράφι»

Παράδειγμα 3:

Αποσαφήνιση υποκειμένου: «...όταν το υποκείμενο είναι πρώτου ή δευτέρου προσώπου, η αντωνυμία (εγώ, εσύ – εμείς, εσείς) συνήθως παραλείπεται, γιατί δηλώνεται από την κατάληξη του ρήματος...» Συντακτικό της Νέας Ελληνικής, σελ.16

π.χ: «Φεύγουμε αύριο το πρωί ; »

Τα αποτελέσματα αυτής της υπομονάδας προωθούνται στην επόμενη, τρίτη και τελευταία υπομονάδα, μέσω του ίδιου και πάλι αρχείου xml, στο τρίτο επίπεδο συντακτικού σχολιασμού.

4.4.3 Η υπομονάδα κανόνων και κανονικών εκφράσεων (rules and regular expressions)

Αυτή η υπομονάδα αποτελεί ο τρίτο και τελευταίο επίπεδο επεξεργασίας του συντακτικού σχολιαστή. Περιέχει ένα σύνολο κανονικών εκφράσεων πάνω στις οποίες εφαρμόζονται τα αποτελέσματα του προηγούμενου επιπέδου. Οι κανόνες και οι αντίστοιχες κανονικές εκφράσεις που υποστηρίζονται έχουν περιγραφεί στην παράγραφο 3.2.2.

Αυτό το επίπεδο μας δίνει τη δυνατότητα να αυξήσουμε την ακρίβεια των αποτελεσμάτων μας, εφόσον περιορίσουμε το εύρος των προτάσεων εισόδου και εισάγουμε σε αυτό το επίπεδο τους απαραίτητους κανόνες. Η επιλογή των κανονικών

εκφράσεων έγινε για ακριβώς αυτό το λόγο. Εάν θέλουμε να αυξήσουμε την ακρίβεια του συντακτικού αναλυτή, δε χρειάζεται να αλλάξουμε όλη την αρχιτεκτονική του. Μπορούμε απλά να εισάγουμε περισσότερους και πληρέστερους κανόνες σε αυτήν την υπομονάδα.

Οι κανόνες μπορεί να αφορούν:

- i. Απλές προτάσεις
- ii. Ερωτηματικές προτάσεις (3 κατηγορίες)
- iii. Ελλειπτικές προτάσεις (να λείπει κάποιος όρος, συνήθως το υποκείμενο)
- iv. Επανυξημένες προτάσεις (Υποκείμενο + Κατηγορήμα + Δευτερεύοντες όροι)
- v. Σύνθετες προτάσεις (Πολλά Υποκείμενα + Πολλά Κατηγορούμενα)

Ακολουθούν μερικά χαρακτηριστικά παραδείγματα κανόνων και κανονικών εκφράσεων:

Παράδειγμα 1:

(κανονική έκφραση)	$V [JNA] ((, [JNA])^* C [JNA])^*$
(συντακτικός κανόνας)	$R - Y - ((, Y)^* - S - Y)^*$

π.χ: «Σκόραρε ο Κώστας ,ο Γιώργος ή ο Γιάννης ?» 1η κατηγορία ερωτήσεων

Παράδειγμα 2:

(κανονική	έκφραση)
$[JNA]((,[JNA])^*C^*[JNA])^*V((,[JNA])^*C^*[JNA])^*[JNA]((,[JNA])^*C[JNA])^*$	
(συντακτικός κανόνας)	

$A - (, A)^* - S - A)^* - R - A - (, A)^* - S - A)^* - Y - ((, Y)^* - S - Y))^*$

π.χ: «Κόκκινη ή κίτρινη κάρτα πήρε ο Γιάννης ή ο Γιώργος ?»

Μετά και από αυτή την υπομονάδα, έχει ολοκληρωθεί η συντακτική ανάλυση του κειμένου.

Η έξοδος αυτής της μονάδας είναι το αρχικό μας κείμενο με τις λεκτικές μονάδες που αναλύθηκαν συντακτικά να έχουν δίπλα τους τη συντακτική τους ανάλυση.

4.5 Οι προτάσεις που υποστηρίζονται

Η συγκεκριμένη υλοποίηση του συντακτικού αναλυτή υποστηρίζει και αναλύει με επιτυχία ένα αρκετά μεγάλο αριθμό προτάσεων. Εάν προσπαθήσουμε να κατηγοριοποιήσουμε τις προτάσεις αυτές, τότε οι κατηγορίες των προτάσεων που υποστηρίζονται και αναλύονται είναι οι ακόλουθες:

I. Ανάλυση απλών προτάσεων(Y-P / Y-P-A / Y-P-K / Y-P-A-A / Y-P-A-K):

π.χ: « Ο Γιάννης είναι σπίτι »

II. Ανάλυση προτάσεων με πολλά υποκείμενα, συνδεδεμένα με τους συνδέσμους «και», «ή» αλλά και με κόμμα , εφόσον ακολουθεί σύνδεσμος :

π.χ: « Σκόραραν ο Κώστας, ο Γιώργος και ο Γιάννης ; »

III. Ανάλυση προτάσεων με πολλά αντικείμενα/ κατηγορούμενα, συνδεδεμένα με τους συνδέσμους «και», «ή» αλλά και με κόμμα:

π.χ : « Κέρδισε κάρτα ή αποβολή ο Κώστας ; »

IV. Ανάλυση προτάσεων με πολλά υποκείμενα και πολλά αντικείμενα/ κατηγορούμενα, συνδεδεμένα με τους συνδέσμους «και», «ή» αλλά και με κόμμα:

π.χ: «Έκανε αλλαγή τον Κώστα και τον Γιάννη ο προπονητής ή ο πρόεδρος; »

- V. Ανάλυση προτάσεων που το υποκείμενο του ρήματος εννοείται (ελλειπτικές προτάσεις):

π.χ: « Πήρατε το λεωφορείο ; »

- VI. Εύρεση ομοιόπτωτων, ετερόπτωτων, επιθετικών και κατηγορηματικών προσδιορισμών:

π.χ: « Σκόραρε ο Γιάννης , ο αδερφός μου ; »

« Είναι αυτό το σπίτι του Κώστα ; »

« Πέρασε ένας ψηλός όμορφος νέος άνθρωπος ; »

« Κατάχλομος ο ασθενής κοίταζε το γιατρό »

- VII. Πρώτη κατηγορία ερωτήσεων ($P - Y / K - P - Y$) αλλά και όλες οι πιθανές αποκλείσεις από την κανονική σειρά:

π.χ: « Σκόραρε ο Κώστας ; »

« Ο Κώστας σκόραρε ; »

π.χ: « Κάρτα πήρε ο Γιάννης ; »

« Πήρε κάρτα ο Γιάννης ; »

« Ο Γιάννης πήρε κάρτα ; »

« Πήρε ο Γιάννης κάρτα ; »

Συμπέρασμα: Υποστηρίζεται η αντιμεταθετικότητα των συντακτικών όρων

- VIII. Σύνδεση προτάσεων, όλων των παραπάνω υποστηριζόμενων κατηγοριών, με τους συνδέσμους «και», «ή» αλλά και με κόμμα:

π.χ: « Σκόραρε ο Κώστας ή αποβλήθηκε ο Γιώργος ; »

Στην επόμενη παράγραφο θα δούμε ένα παράδειγμα λειτουργίας του συντακτικού αναλυτή με ένα κείμενο

4.6 System flow

Στη συγκεκριμένη ενότητα θα δειχθεί η ροή πληροφορίας κατά την επεξεργασία ενός κειμένου. Θα δειχθούν όλα τα στάδια, από τη φάση που ο χρήστης εισάγει το κείμενο στον συντακτικού αναλυτή μέχρι αυτός να μας επιστρέψει το τελικό αποτέλεσμα. Το κείμενο που επιλέχθηκε αποτελείται από λίγες λεκτικές μονάδες όμως περιέχει ένα ικανοποιητικό σύνολο των περιπτώσεων που αναλύονται με επιτυχία από το συντακτικό αναλυτή.

Αρχικά λοιπόν ο χρήστης εισάγει στο συντακτικό αναλυτή το ακόλουθο κείμενο:

Χωρίς τον Γενικό Γραμματέα του ΟΗΕ, θα διεξαχθεί η τελετή έναρξης των Ολυμπιακών Αγώνων του Πεκίνου. Παράλληλα η ολυμπιακή φλόγα βρίσκεται στην Αργεντινή ενώ οι Κινέζοι προετοιμάζονται για το ενδεχόμενο βροχής την ημέρα της τελετής έναρξης των Αγώνων.

Εικόνα 15: Το αρχικό κείμενο πριν την εισαγωγή στον αναλυτή

Αυτό το κείμενο προωθείται στη μονάδα του μορφοσυντακτικού αναλυτή.

Η μορφοσυντακτική ανάλυση ενός ASCII κειμένου έχει ως αποτέλεσμα την παραγωγή πληροφορίας σχετικής με α) τα δομικά στοιχεία του κειμένου, δηλ. τις παραγράφους, τις προτάσεις, τις λέξεις κτλ., και β) τις μορφοσυντακτικές ιδιότητες των λέξεων. Η πληροφορία αυτή ενσωματώνεται στο κείμενο υπό τη μορφή σχολίων κατά τρόπο ώστε να είναι επεξεργάσιμη από τον υπολογιστή και κατανοητή από τον άνθρωπο. Η μορφή και η δομή των σχολίων ακολουθούν το πρότυπο της XML¹¹ (eXtensible Markup Language).

Έτσι λοιπόν μετά το πέραςμα από το μορφοσυντακτικό σχολιαστή και τη γραμματική ανάθεση σε όρους, το αποτέλεσμα είναι το ακόλουθο xml αρχείο:

¹¹ Η XML είναι ένα υποσύνολο της SGML (Standard Generalized Markup Language, ISO standard 8879:1986) και σχεδιάστηκε για να διευκολύνει την ανταλλαγή δομημένης πληροφορίας στο Διαδίκτυο. Στην ουσία, η XML δεν είναι μια προκαθορισμένη markup γλώσσα· είναι μια μεταγλώσσα, δηλ. μια γλώσσα για τον ορισμό markup γλωσσών.

```

<?xml version="1.0" encoding="ISO-8859-7"?>
<d>
<p from="0" to="267">
  <s from="0" to="100">
    <gw from="0" to="5" PoS="Prp" lemma="χωρίς">Χωρίς</gw>
    <gw from="6" to="9" PoS="Art" attrs="MscSngAcc" lemma="ο">τον</gw>
    <gw from="10" to="16" PoS="Adj" attrs="MscSngAcc NtrSngNomAccVoc"
lemma="γενικός">Γενικό</gw>
    <gw from="17" to="26" PoS="N" attrs="MscFemSngGenAccVoc"
lemma="γραμματέας">Γραμματέα</gw>
    <gw from="27" to="30" PoS="Art" attrs="MscNtrSngGen" lemma="ο">του</gw>
    <gw from="31" to="34" PoS="N" lemma="?">ΟΗΕ</gw>
    <punc from="34" to="35">,</punc>
    <gw from="36" to="38" PoS="Pcl" lemma="θα">θα</gw>
    <gw from="39" to="48" PoS="V" attrs="PsvPstSjvSng3rd PsvFutIndSng3rd PsvPstInf"
lemma="διεξάγω">διεξαχθεί</gw>
    <gw from="49" to="50" PoS="Art" attrs="FemSngNom" lemma="ο">η</gw>
    <gw from="51" to="57" PoS="N" attrs="FemSngNomAccVoc" lemma="τελετή">τελετή</gw>
    <gw from="58" to="65" PoS="N" attrs="FemSngGen" lemma="έναρξη">έναρξης</gw>
    <gw from="66" to="69" PoS="Art" attrs="MscFemNtrPlrGen" lemma="ο">των</gw>
    <gw from="70" to="80" PoS="Adj" attrs="MscFemNtrPlrGen"
lemma="ολυμπιακός">Ολυμπιακών</gw>
    <gw from="81" to="87" PoS="N" attrs="MscPlrGen" lemma="αγώνας">Αγώνων</gw>
    <gw from="88" to="91" PoS="Art" attrs="MscNtrSngGen" lemma="ο">του</gw>
    <gw from="92" to="99" PoS="N" lemma="?">Πεκίνου</gw>
    <punc from="99" to="100">.</punc>
  </s>
  <s from="101" to="267">
    <gw from="101" to="110" PoS="Adv" lemma="παράλληλα">Παράλληλα</gw>
    <gw from="111" to="112" PoS="Art" attrs="FemSngNom" lemma="ο">η</gw>
    <gw from="113" to="122" PoS="Adj" attrs="FemSngNomAccVoc"
lemma="ολυμπιακός">ολυμπιακή</gw>
    <gw from="123" to="128" PoS="N" attrs="FemSngNomAccVoc" lemma="φλόγα">φλόγα</gw>
    <gw from="129" to="138" PoS="V" attrs="PsvPrsFucIndSjvSng3rd"
lemma="βρίσκω">βρίσκεται</gw>
    <gw from="139" to="143" PoS="PrpArt" attrs="FemSngAcc" lemma="στο">στην</gw>
    <gw from="144" to="153" PoS="N" attrs="FemSngNomAccVoc"
lemma="Αργεντινή">Αργεντινή</gw>
    <gw from="154" to="157" PoS="Cnj" lemma="ενώ">ενώ</gw>
    <gw from="158" to="160" PoS="Art" attrs="MscFemPlrNom" lemma="ο">οι</gw>
    <gw from="161" to="168" PoS="N" attrs="MscPlrNomVoc" lemma="κινέζος">Κινέζοι</gw>
    <gw from="169" to="184" PoS="V" lemma="?">προετοιμάζονται</gw>
    <gw from="185" to="188" PoS="Prp" lemma="για">για</gw>
  </s>
</p>

```

```

<gw from="189" to="191" PoS="Art" attrs="MscSngAcc NtrSngNomAcc" lemma="ο">το</gw>
<gw from="192" to="202" PoS="Pcp" attrs="PsvPrsMscSngAcc PsvPrsNtrSngNomAccVoc"
lemma="ενδέχεται">ενδεχόμενο</gw>
<gw from="203" to="209" PoS="N" attrs="FemSngGen" lemma="βροχή">βροχής</gw>
<gw from="210" to="213" PoS="Art" attrs="FemSngAcc" lemma="ο">την</gw>
<gw from="214" to="219" PoS="N" attrs="FemSngNomAccVoc" lemma="ημέρα">ημέρα</gw>
<gw from="220" to="223" PoS="Art" attrs="FemSngGen" lemma="ο">της</gw>
<gw from="224" to="231" PoS="N" attrs="FemSngGen" lemma="τελετή">τελετής</gw>
<gw from="232" to="239" PoS="N" attrs="FemSngGen" lemma="έναρξη">έναρξης</gw>
<gw from="240" to="243" PoS="Art" attrs="MscFemNtrPlrGen" lemma="ο">των</gw>
<symb from="244" to="245">&apos;</symb>
<gw from="245" to="258" PoS="Adj" attrs="MscFemNtrPlrGen"
lemma="υπερσύγχρονος">υπερσύγχρονων</gw>
<symb from="258" to="259">&apos;</symb>
<gw from="260" to="266" PoS="N" attrs="MscPlrGen" lemma="αγώνας">Αγώνων</gw>
<punc from="266" to="267">.</punc>
</s>
</p>
</d>

```

Εικόνα 16: Το xml κείμενο όπως προκύπτει από το μορφοσυντακτικό αναλυτή

Η εικόνα 16 παρουσιάζει το μορφοσυντακτικό σχολιασμό ενός κειμένου σε μορφή XML.

Στο επόμενο βήμα, το παραπάνω κείμενο εισέρχεται στη μονάδα σύνδεσης – απομόνωσης και το αποτέλεσμα είναι το ακόλουθο:

```

<?xml version="1.0" encoding="ISO-8859-7"?>
<d>
<p from="0" to="267">
  <s from="0" to="100">
    <gw from="0" to="5" PoS="Prp" lemma="χωρίς">Χωρίς</gw>
    <gw from="6" to="9" PoS="Art" attrs="MscSngAcc" lemma="ο">τον</gw>
    <gw from="10" to="16" PoS="Adj" attrs="MscSngAcc NtrSngNomAccVoc"
lemma="γενικός">Γενικό</gw>
    <gw from="17" to="26" PoS="N" attrs="MscFemSngGenAccVoc"
lemma="γραμματέας">Γραμματέα</gw>
    <gw from="27" to="30" PoS="Art" attrs="MscNtrSngGen" lemma="ο">του</gw>
    <gw from="31" to="34" PoS="N" lemma="?">ΟΗΕ</gw>
    <punc from="34" to="35">,</punc>
    <gw from="36" to="38" PoS="Pcl" lemma="θα">θα</gw>
    <gw from="39" to="48" PoS="V" attrs="PsvPstSjvSng3rd PsvFutIndSng3rd PsvPstInf"
lemma="διεξάγω">διεξαχθεί</gw>
    <gw from="49" to="50" PoS="Art" attrs="FemSngNom" lemma="ο">η</gw>
    <gw from="51" to="57" PoS="N" attrs="FemSngNomAccVoc" lemma="τελετή">τελετή</gw>
    <gw from="58" to="65" PoS="N" attrs="FemSngGen" lemma="έναρξη">έναρξης</gw>
    <gw from="66" to="69" PoS="Art" attrs="MscFemNtrPlrGen" lemma="ο">των</gw>
    <gw from="70" to="80" PoS="Adj" attrs="MscFemNtrPlrGen"
lemma="ολυμπιακός">Ολυμπιακών</gw>
    <gw from="81" to="87" PoS="N" attrs="MscPlrGen" lemma="αγώνας">Αγώνων</gw>
    <gw from="88" to="91" PoS="Art" attrs="MscNtrSngGen" lemma="ο">του</gw>
    <gw from="92" to="99" PoS="N" lemma="?">Πεκίνου</gw>
    <punc from="99" to="100">.</punc>
  </s>
  <s from="101" to="157">
    <gw from="101" to="110" PoS="Adv" lemma="παράλληλα">Παράλληλα</gw>
    <gw from="111" to="112" PoS="Art" attrs="FemSngNom" lemma="ο">η</gw>
    <gw from="113" to="122" PoS="Adj" attrs="FemSngNomAccVoc"
lemma="ολυμπιακός">ολυμπιακή</gw>
    <gw from="123" to="128" PoS="N" attrs="FemSngNomAccVoc" lemma="φλόγα">φλόγα</g
    <gw from="129" to="138" PoS="V" attrs="PsvPrsFucIndSjvSng3rd"
lemma="βρίσκω">βρίσκεται</gw>
    <gw from="139" to="143" PoS="PrpArt" attrs="FemSngAcc" lemma="στο">στην</gw>
    <gw from="144" to="153" PoS="N" attrs="FemSngNomAccVoc"
lemma="Αργεντινή">Αργεντινή</gw>
    <gw from="154" to="157" PoS="Cnj" lemma="ενώ">ενώ</gw>
  </s>
  <s from="158" to="267">
    <gw from="158" to="160" PoS="Art" attrs="MscFemPlrNom" lemma="ο">οι</gw>
    <gw from="161" to="168" PoS="N" attrs="MscPlrNomVoc" lemma="κινέζος">Κινέζοι</gw>
    <gw from="169" to="184" PoS="V" lemma="?">προετοιμάζονται</gw>
    <gw from="185" to="188" PoS="Prp" lemma="για">για</gw>
  </s>
</p>
</d>

```

```

    <gw from="189" to="191" PoS="Art" attrs="MscSngAcc NtrSngNomAcc" lemma="ο">το</gw>
    <gw from="192" to="202" PoS="Pcp" attrs="PsvPrsMscSngAcc PsvPrsNtrSngNomAccVoc"
lemma="ενδέχεται">ενδεχόμενο</gw>
    <gw from="203" to="209" PoS="N" attrs="FemSngGen" lemma="βροχή">βροχής</gw>
    <gw from="210" to="213" PoS="Art" attrs="FemSngAcc" lemma="ο">την</gw>
    <gw from="214" to="219" PoS="N" attrs="FemSngNomAccVoc" lemma="ημέρα">ημέρα</gw>
    <gw from="220" to="223" PoS="Art" attrs="FemSngGen" lemma="ο">της</gw>
    <gw from="224" to="231" PoS="N" attrs="FemSngGen" lemma="τελετή">τελετής</gw>
    <gw from="232" to="239" PoS="N" attrs="FemSngGen" lemma="έναρξη">έναρξης</gw>
    <gw from="240" to="243" PoS="Art" attrs="MscFemNtrPlrGen" lemma="ο">των</gw>
    Deleted < symb from="244" to="245">&apos;</ symb>
    <gw from="245" to="258" PoS="Adj" attrs="MscFemNtrPlrGen"
lemma="υπερσύγχρονος">υπερσύγχρονων</gw>
    Deleted < symb from="258" to="259">&apos;</ symb>
    <gw from="260" to="266" PoS="N" attrs="MscPlrGen" lemma="αγώνας">Αγώνων</gw>
    <punc from="266" to="267">.</punc>
  </s>
</p>
</d>

```

Εικόνα 17: Το xml κείμενο έπειτα από τη μονάδα σύνδεσης – απομόνωσης

Στην εικόνα 17 μπορούμε να παρατηρήσουμε τις αλλαγές που επέφεραν στο κείμενο οι δύο υπομονάδες της μονάδας σύνδεσης – απομόνωσης.

Η πρώτη υπομονάδα, αυτή της αφαίρεσης σημείων στίξης, έχει αφαιρέσει από το κείμενο στις θέσεις 244 και 258 τα ελληνικά εισαγωγικά μέσα στα οποία βρισκόταν η λεκτική μονάδα *υπερσύγχρονων*. Αυτά δεν προσέδιδαν καμία συντακτική πληροφορία στο κείμενο και γι' αυτό, σωστά, αφαιρέθηκαν.

Η δεύτερη υπομονάδα, αυτή της εύρεσης και διαχωρισμού προτάσεων, εντόπισε τη σύνδεση προτάσεων. Αυτή βρίσκεται στη θέση 158 όπου δύο προτάσεις ενώνονται με το σύνδεσμο *ενώ*. Ο διαχωρισμός των προτάσεων είναι σωστός και απλοποιεί το έργο του συντακτικού σχολιαστή ο οποίος αντί για μια σύνθετη πρόταση έχει να αναλύσει δύο απλούστερες.

Η επόμενη μονάδα που θα περάσει το μορφοποιημένο πλέον κείμενο είναι αυτή του συντακτικού σχολιαστή. Έτσι στο τέλος θα πάρουμε τη συντακτική ανάλυση του αρχικού μας κειμένου, που είναι η ακόλουθη:

***** START SENTENCE *****

Expression 0 : Χωρίς τον Γενικό Γραμματέα του ΟΗΕ θα διεξαχθεί η τελετή έναρξης των Ολυμπιακών Αγώνων του Πεκίνου

Verb: διεξαχθεί Voice:Act Tense:/Fut Mood:/Ind Number:/Sng Person:/3rd

Subject: τελετή POS:N Attribute: SngNomAccVoc

----->>>>with Specification: έναρξης POS:N Attribute:FemSngGen

----->>>>with Specification: Ολυμπιακών POS:Adj Attribute: MscFemNtrPlrGen

----->>>>with Specification: Αγώνων POS:N Attribute: MscPlrGen

----->>>>with Specification Πεκίνου POS:N

Specification: Χωρίς POS: Prp

----->>>>with Specification: Γενικό POS: N AttributeMscSngAcc NtrSngNomAcc

----->>>>with Specification: Γραμματέα POS: N AttributeMscFemSngGenAccVoc

----->>>>with Specification: ΟΗΕ POS: N

***** END SENTENCE *****

***** START SENTENCE *****

Expression 1 : Παράλληλα η ολυμπιακή φλόγα βρίσκεται στην Αργεντινή

Verb:βρίσκεται Voice:Act Tense:/Prs Mood:/Ind Number:/Sng Person:/3rd

Subject: φλόγα POS: N Attribute: FemSngNomAccVoc

----->>>>with Specification:ολυμπιακή POS:Adj Attribute: FemSngNomAccVoc

Object Αργεντινή POS: N Attribute: FemSngNomAccVoc

AdverbΠαράλληλα

Conjunction with other sentence: true Cnj: ενώ

***** NEXT SENTENCE *****

Expression 2 : οι Κινέζοι προετοιμάζονται για το ενδεχόμενο βροχής την ημέρα της τελετής έναρξης των υπερσύγχρονων Αγώνων

Verb:προετοιμάζονται Voice:Act Tense:/Prs Mood:/Ind Number:/Plr Person:/3rd

SubjectΚινέζοι POS: N Attribute: MscPlrNomVoc

Object ενδεχόμενο POS: Psp Attribute: PsvPrsMscSngAcc
PsvPrsNtrSngNomAccVoc

Object βροχής	POS: N	AttributeFemSngGen
Object ημέρα	POS: N	AttributeFemSngNomAccVoc
----->>>>>with Specification:τελετής	POS: N	Attribute: SngNomAccVoc
----->>>>>with Specification: έναρξης	POS: N	Attribute:FemSngGen
----->>>>>with Specification:υπερσύγχρονων	POS: Adj	Attribute:
		MscFemNtrPlrGen
----->>>>>with Specification:Αγώνων	POS: N	Attribute:
		MscPlrGen
***** END SENTENCE *****		

Εικόνα 18: Η συντακτική ανάλυση του κειμένου

Παραπάνω φαίνεται η τελική συντακτική ανάλυση του κείμενου, όπως αυτή προκύπτει μετά την ολοκλήρωση του περάσματος του κειμένου από το συντακτικό μας αναλυτή.

Η ανάλυση τελικά έχει καταφέρει τους αρχικούς μας στόχους:

- ✓ Αφαίρεσε τα σημεία στίξης
- ✓ Ανέλυσε με επιτυχία το κείμενο στις προτάσεις του
- ✓ Εντόπισε σύνδεση προτάσεων
- ✓ Ανέλυσε σωστά συντακτικά τις προτάσεις που προέκυψαν

Έτσι σε αυτό το κεφάλαιο είδαμε μια υλοποίηση του συντακτικού μας αναλυτή. Είδαμε επίσης και βήμα προς βήμα την ανάλυση ενός κειμένου με επιτυχία. Στο επόμενο κεφάλαιο θα ασχοληθούμε με μία γραφική διεπαφή του συντακτικού αναλυτή που έχουμε κατασκευάσει.

ΚΕΦΑΛΑΙΟ 5

ΣΥΣΤΗΜΑ ΑΝΑΛΗΨΗΣ ΔΕΔΟΜΕΝΩΝ

5.1 Εισαγωγή

Για την καλύτερη αξιοποίηση του συντακτικού αναλυτή, τον χρησιμοποιήσαμε σε συστήματα ανάληψης πληροφορίας αλλά και σε τυχαία κείμενα του διαδικτύου. Έτσι θα μπορέσουμε να τον αξιολογήσουμε καλύτερα, σε πραγματικά δεδομένα αλλά και να κατανοήσουμε καλύτερα τα πλεονεκτήματα που μας προσφέρει η γνώση της συντακτικής ανάλυσης ενός κειμένου για εφαρμογές ανάληψης πληροφορίας.

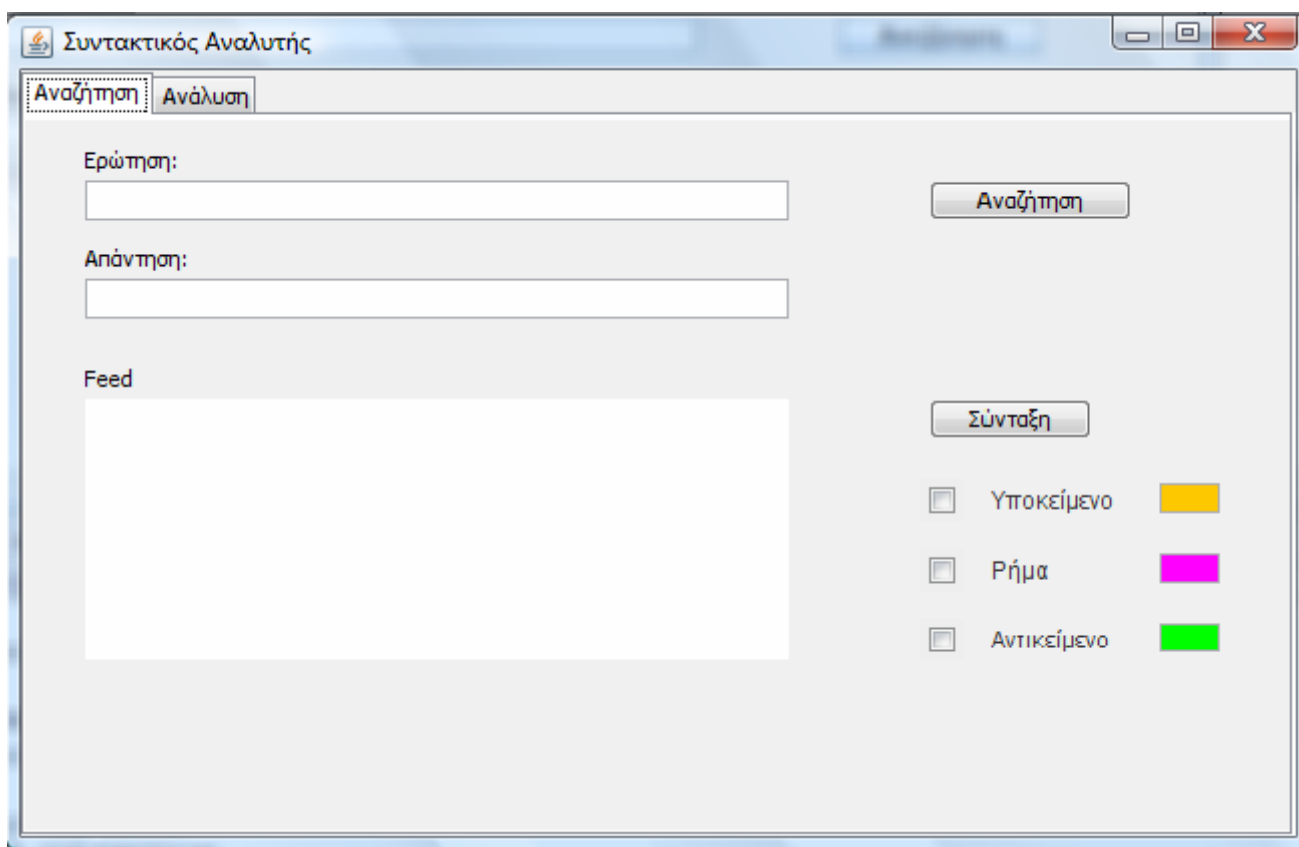
Στις επόμενες παραγράφους θα παρουσιάσουμε μια γραφική διεπαφή του συντακτικού μας αναλυτή. Θα παρουσιάσουμε τα σενάρια χρήσης της, τον τρόπο με τον οποίο αυτή λειτουργεί για να ικανοποιήσει αυτά τα σενάρια και στο τέλος θα κάνουμε μια αξιολόγηση της διεπαφής και της καταλληλότητάς της.

5.2 Σενάρια χρήσης

Η διεπαφή η οποία δημιουργήσαμε αποτελείται από δύο βασικές ενέργειες. Η μία είναι η αναζήτηση της απάντησης σε μία ερώτηση που εισάγει ο χρήστης και απευθύνεται σε ένα σύνολο δεδομένων που δέχεται η διεπαφή από το διαδίκτυο και η δεύτερη είναι η συντακτική ανάλυση ενός κειμένου που εισάγει ο χρήστης. Στη συνέχεια θα περιγράψουμε τα πιθανά σενάρια χρήσης της κάθε ενέργειας.

Σενάριο 1:

Το σενάριο αυτό απευθύνεται στην πρώτη ενέργεια που παρέχει η διεπαφή μας, αυτή της αναζήτησης απαντήσεων σε ερωτήσεις. Το παράθυρο που εμφανίζεται για αυτή την ενέργεια είναι το ακόλουθο:



Εικόνα 19: Screen shot της ενέργειας της αναζήτησης

Ο χρήστης εισάγει μία ερώτηση στο αντίστοιχο πεδίο της διεπαφής, το πεδίο «Ερώτηση», η απάντηση της οποίας μπορεί να στηρίζεται σε συντακτικά κριτήρια, δηλαδή να αναζητά τα υποκείμενα μιας ενέργειας ή τα αντικείμενα στα οποία πηγαίνει η ενέργεια ενός ρήματος. Στη συνέχεια επιλέγει να γίνει αναζήτηση της ερώτησης, από το αντίστοιχο κουμπί «Αναζήτηση».

Εάν υπάρχει απάντηση στην ερώτηση αυτή εμφανίζεται στο σχετικό πεδίο της διεπαφής, στο πεδίο «Απάντηση» και ακριβώς από κάτω, σε ένα δεύτερο πεδίο, αυτό του «Feed», εμφανίζεται το κείμενο από το οποίο προέκυψε η απάντηση αυτή. Επίσης η πρόταση του κειμένου που ευθύνεται για την απάντηση χρωματίζεται.

Εάν η ερώτηση δεν έχει απάντηση μέσα στο σύνολο των κειμένων που εξετάζονται τότε ενημερώνεται ο χρήστης ότι δεν υπάρχει μέσα στα κείμενα πληροφόρηση σχετική με το ερώτημα που έθεσε.

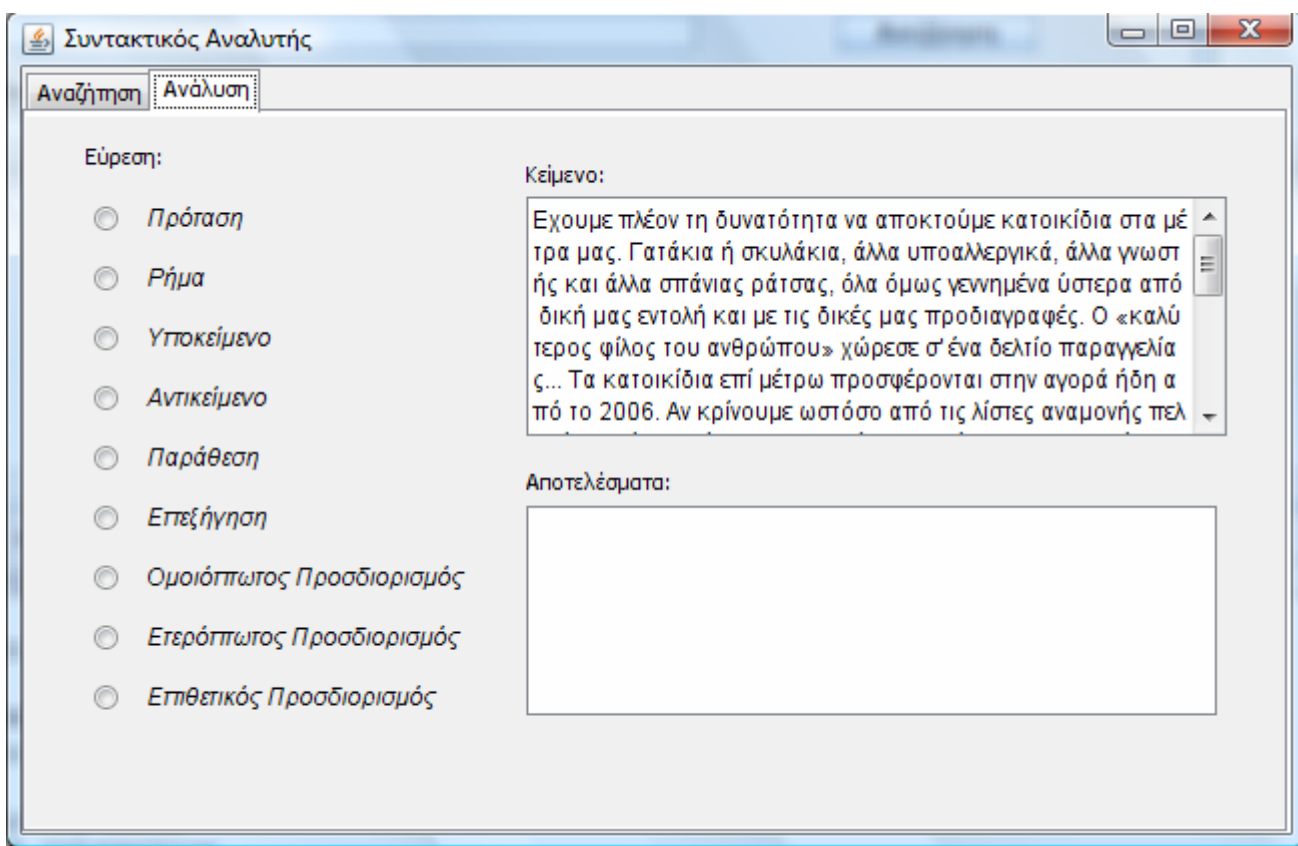
Μετά την ολοκλήρωση της αναζήτησης και εφόσον αυτή επιστρέψει αποτελέσματα ο χρήστης μπορεί να επιλέξει να γίνει συντακτική ανάλυση του κειμένου μέσα στο οποίο βρίσκεται και η απάντηση του ερωτήματός του. Επιλέγοντας λοιπόν τους συντακτικούς όρους που θέλει και πατώντας το κουμπί

«σύνταξη», μπορεί να προχωρήσει σε μία συντακτική ανάλυση ολόκληρου του κειμένου που εμφανίζεται στο πεδίο «Feed» , ενώ οι συντακτικοί όροι θα χρωματιστούν κατάλληλα.

Ο χρήστης μπορεί να επαναλάβει την αναζήτηση όσες φορές επιθυμεί, ενώ οι ερωτήσεις μπορούν να είναι και σύνθετες καθώς υποστηρίζεται η σύνδεση προτάσεων. Η διεπαφή ενημερώνει κάθε φορά κατάλληλα το χρήστη για τα αποτελέσματα της αναζήτησής του.

Σενάριο 2:

Το σενάριο αυτό απευθύνεται στην δεύτερη ενέργεια που παρέχει η διεπαφή μας, αυτή της ανάλυσης κειμένων. Το παράθυρο που εμφανίζεται για αυτή την ενέργεια είναι το ακόλουθο:



Εικόνα 20: Screen shot της ενέργειας της ανάλυσης

Ο χρήστης εδώ επιλέγει το συντακτικό όρο που επιθυμεί να αναζητηθεί μέσα στο κείμενο που εμφανίζεται στο πεδίο «Κείμενο». Επιλέγει λοιπόν έναν από τους συντακτικούς όρους.

Εάν ο συντακτικός αυτός όρος υπάρχει στο κείμενο ,τότε χρωματίζεται στα σημεία εκείνα του κειμένου στα οποία εμφανίζεται. Επίσης στο πεδίο «Αποτελέσματα» , παρουσιάζεται και η γραμματική ανάλυση των όρων που αναζητήθηκαν και βρέθηκαν.

Ο χρήστης μπορεί να επαναλάβει την αναζήτησή του πολλές φορές, αλλάζοντας κάθε φορά το κριτήριο αναζήτησης.

5.3 Λειτουργικότητα

Όπως είδαμε στην προηγούμενη παράγραφο, η γραφική διεπαφή που δημιουργήσαμε προσφέρει στο χρήστη ένα σύνολο επιλογών. Για να είναι διαθέσιμες αυτές οι επιλογές, η διεπαφή έχει μια ορισμένη λειτουργικότητα. Για να είμαστε σε συμφωνία με την προηγούμενη παράγραφο όπου περιγράφηκαν οι επιλογές, θα περιγράψουμε αυτή τη λειτουργικότητα σε σχέση με τα αντίστοιχα σενάρια χρήσης στα οποία εφαρμόζεται.

Στο πρώτο σενάριο που υποστηρίζεται η αναζήτηση ερωτήσεων με βάση συντακτικά κριτήρια, η λειτουργικότητα είναι η ακόλουθη:

Η διεπαφή τροφοδοτείται με ένα αρχείο κειμένου το οποίο περιέχει ένα σύνολο rss feeds, που αφορούν ημερήσιες ειδήσεις γενικά. Αυτό το αρχείο ενημερώνεται συνεχώς με νέες ειδήσεις. Όταν λοιπόν εισάγεται αυτό το αρχείο, γίνεται συντακτική ανάλυση όλων των feeds και αποθηκεύονται σε μία προσωρινή δομή.

Όταν ο χρήστης εισάγει την ερώτησή του, τότε εφαρμόζεται συντακτική ανάλυση και στην ερώτηση του χρήστη. Το αποτέλεσμα αυτής της ανάλυσης είναι η κατηγοριοποίηση της ερώτησης και αποθήκευσή της επίσης σε μία προσωρινή δομή. Επίσης αυτή η προσωρινή δομή που αφορά την ερώτηση περιέχει μέσα και πληροφορίες σχετικά με το τι ακριβώς θα ψάξουμε στα feeds, υποκείμενα, αντικείμενα, συνδυασμούς λέξεων κ.ά.

Στη συνέχεια αναζητούνται αυτά που ζητάει η ερώτηση στο σύνολο των feeds. Εάν βρεθεί απάντηση τότε αυτή επιστρέφεται στο χρήστη μαζί με το feed στο οποίο βρέθηκε. Εάν βρεθούν περισσότερα από ένα feed που περιέχουν την απάντηση, τότε επιστρέφονται όλα. Σε κάθε feed που αποτελεί απάντηση στην ερώτηση, η πρόταση η οποία περιέχει την απάντηση χρωματίζεται ενώ ο συγκεκριμένος

συντακτικός όρος που αναζητήθηκε, το υποκείμενο δηλαδή ή το αντικείμενο επιστρέφουν σε ένα ξεχωριστό πεδίο, όπως περιγράφηκε παραπάνω.

Εάν στη συνέχεια ο χρήστης επιλέξει τη συντακτική ανάλυση του feed τότε αυτή παρέχεται άμεσα καθώς έχει γίνει αρχικά και υπάρχει στην προσωρινή δομή. Κάθε συντακτικός όρος που αναζητείται χρωματίζεται κατάλληλα.

Σε σχέση με το δεύτερο σενάριο χρήσης η λειτουργικότητα είναι η ακόλουθη:

Ο χρήστης εισάγει ένα κείμενο το οποίο θέλει να ελέγξει συντακτικά στη διεπαφή. Με το ξεκίνημα της διεπαφής, γίνεται συντακτική ανάλυση στο κείμενο αυτό. Το αποτέλεσμα της ανάλυσης αποθηκεύεται σε μια προσωρινή δομή. Όταν στη συνέχεια ο χρήστης αναζητήσει ένα συγκεκριμένο συντακτικό όρο, ανατρέχουμε αυτή η δομή και εάν περιέχει αυτούς τους συντακτικούς όρους τότε χρωματίζονται στο αρχικό κείμενο. Ακριβώς από κάτω εμφανίζεται η γραμματική ανάλυση όλων αυτών των όρων.

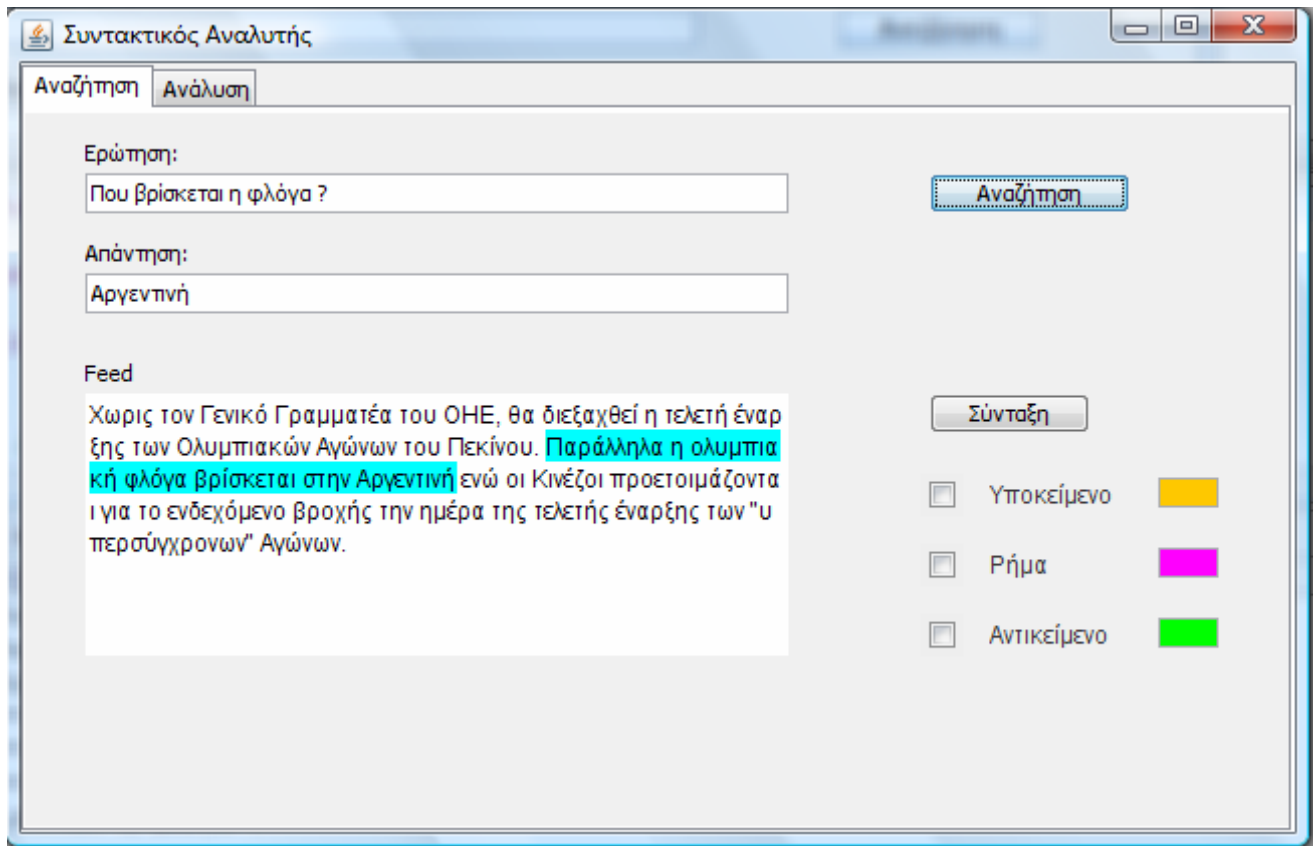
Στη συνέχεια θα παρουσιάσουμε ένα παράδειγμα χρήσης της κάθε επιλογής της διεπαφής.

5.4 Παραδείγματα χρήσης

Για την καλύτερη κατανόηση της λειτουργικότητας που προσφέρεται μέσα από τη γραφική διεπαφή που δημιουργήσαμε, ακολουθούν δύο παραδείγματα:

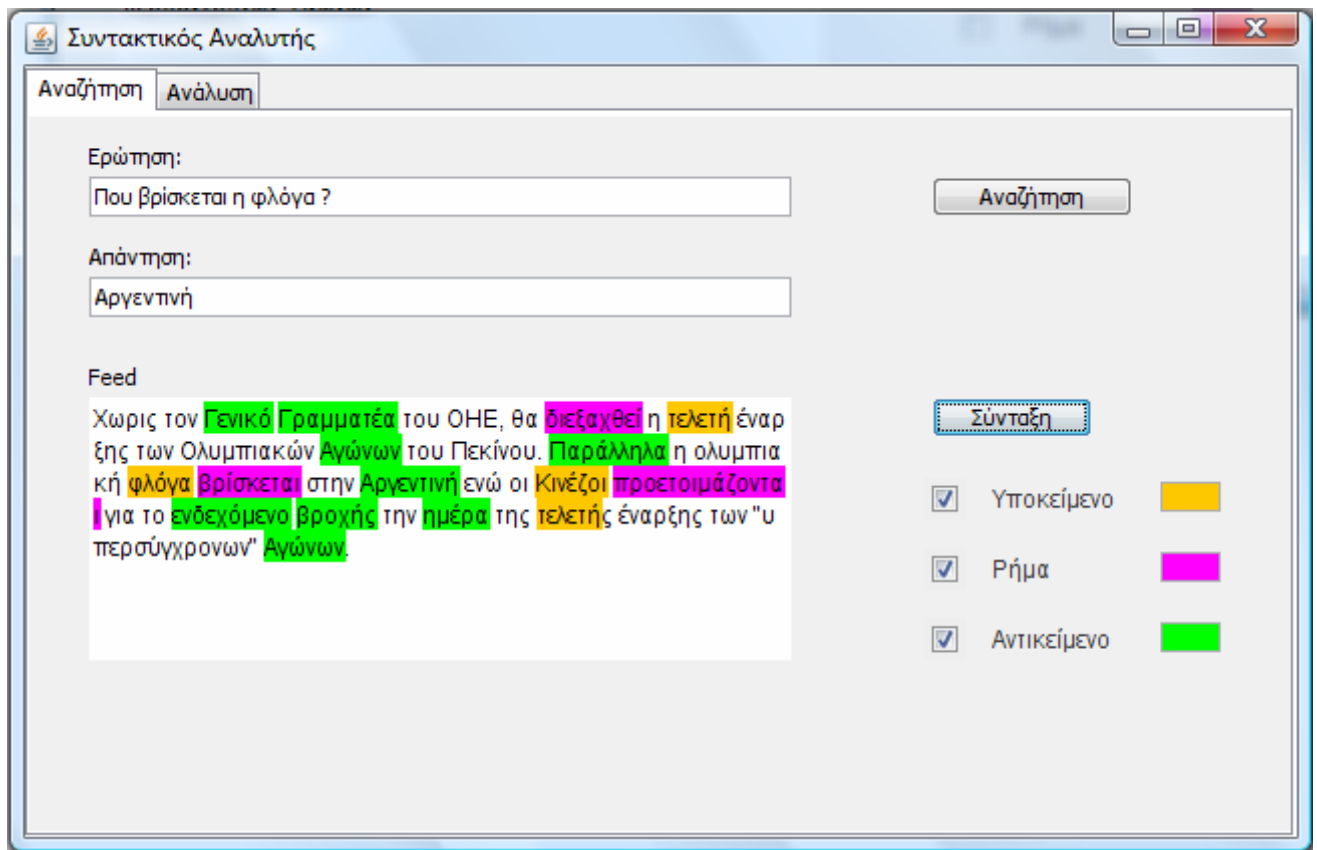
Παράδειγμα 1:

Ο χρήστης εισάγει μία ερώτηση στο σύστημα και υπάρχει απάντηση στην ερώτησή του. Τα αποτελέσματα εμφανίζονται στη ακόλουθη εικόνα:



Εικόνα 21: Παράδειγμα επιτυχημένης ερώτησης(1), ο χρήστης εισάγει την ερώτησή του στο αντίστοιχο πεδίο και πατώντας επάνω στο κουμπί «Αναζήτηση», γίνεται αναζήτηση της ερώτησης. Στη συνέχεια η απάντηση της ερώτησης εμφανίζεται στο αντίστοιχο πεδίο ενώ εάν αυτή είναι θετική, εμφανίζεται και το κείμενο από το οποίο προέκυψε η απάντηση

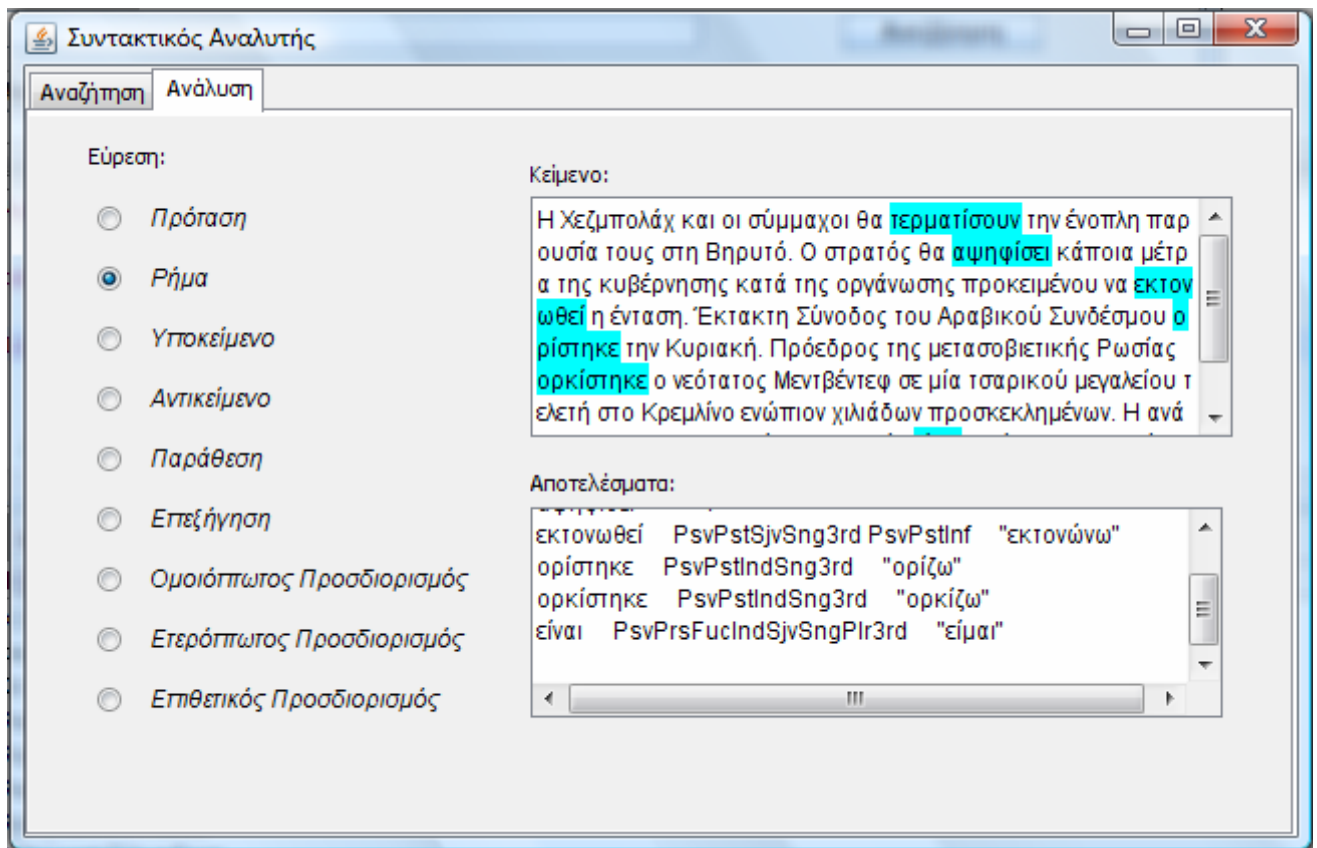
Στη συνέχεια επιλέγει να γίνει και συντακτική ανάλυση του αποτελέσματος.
Τα αποτελέσματα εμφανίζονται στη ακόλουθη εικόνα:



Εικόνα 22: Παράδειγμα επιτυχημένης ερώτησης(2), ο χρήστης πατώντας το κουμπί «Σύνταξη» μπορεί να δει τη σύνταξη του κειμένου απάντησης. Επίσης έχει τη δυνατότητα να επιλέξει να δει συγκεκριμένους συντακτικούς όρους, επιλέγοντας τα αντίστοιχα πεδία

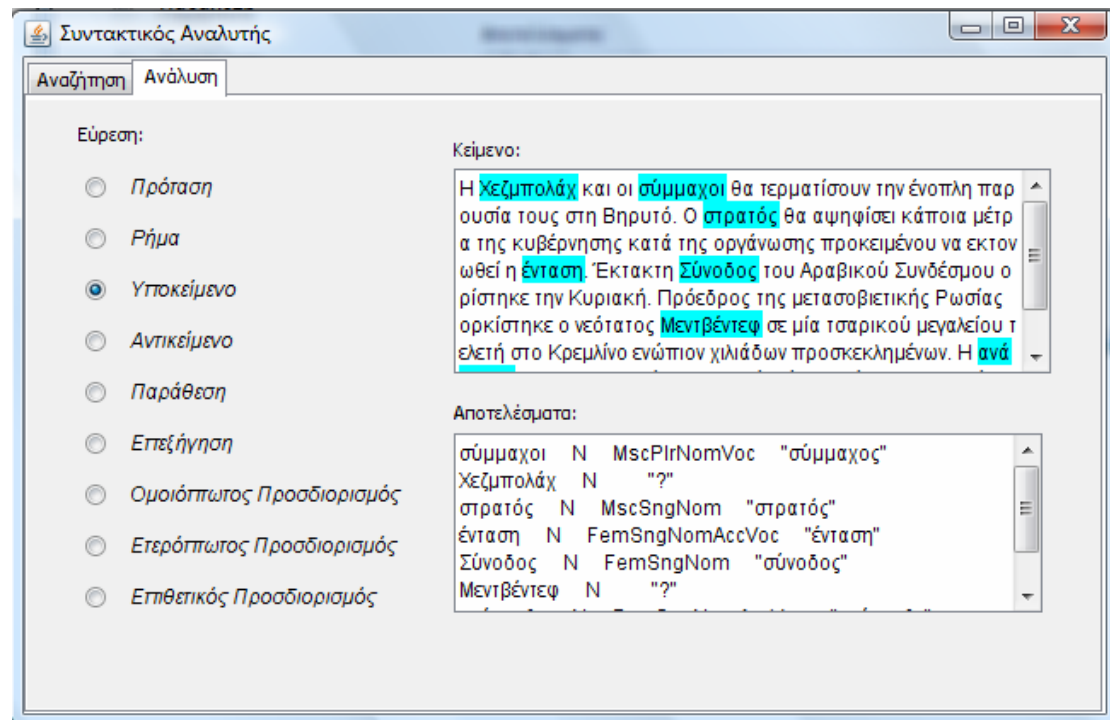
Παράδειγμα 2:

Ο χρήστης εισάγει ένα κείμενο στο σύστημα και αναζητά συγκεκριμένους συντακτικούς όρους. Τα αποτελέσματα εμφανίζονται στη ακόλουθη εικόνα:



Εικόνα 23: Παράδειγμα ανάλυσης κειμένου(1). Ο χρήστης έχει τη δυνατότητα να επιλέξει ένα συγκεκριμένο συντακτικό όρο και αυτός, εάν υπάρχει, θα τονισθεί μέσα στο κείμενο. Κάτω από το κείμενο θα ακολουθήσει η γραμματική ανάλυση του όρου.

Στην παραπάνω ανάλυση αναζητήσαμε ρήματα. Στην ακόλουθη θα αναζητήσουμε υποκείμενα:



Εικόνα 24: Παράδειγμα ανάλυσης κειμένου(2). Παρόμοιο με το παράδειγμα της εικόνας που προηγήθηκε, αλλάζει μόνο ο συντακτικός όρος που αναζητάται.

Με το κεφάλαιο αυτό ολοκληρώνεται η παρουσίαση του συντακτικού αναλυτή. Στο επόμενο κεφάλαιο θα κάνουμε αξιολόγηση των αποτελεσμάτων και θα παρουσιάσουμε πιθανές επεκτάσεις του συντακτικού αναλυτή.

ΚΕΦΑΛΑΙΟ 6

ΑΞΙΟΛΟΓΗΣΗ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1 Αξιολόγηση ορθότητας και καταλληλότητας

Η αξιολόγηση της αρχιτεκτονικής που ακολουθήθηκε και του τελικού αποτελέσματος αποτέλεσε σημαντικό κομμάτι αυτής της διπλωματικής εργασίας. Βλέποντας τους αρχικούς στόχους της εφαρμογής στους οποίους αναφέρεται πως αυτή πρέπει να είναι ανεξάρτητη της κατηγορίας των κειμένων τα οποία δέχεται προς ανάλυση, η αξιολόγησή της έγινε δυσκολότερη, όμως το τελικό αποτέλεσμα ήταν κάτι παραπάνω από ικανοποιητικό.

Για την αξιολόγηση του σωστού διαχωρισμού προτάσεων, ο συντακτικός αναλυτής ανέλυσε ένα σύνολο κειμένων από εφημερίδες. Τα κείμενα αυτά συλλέχθηκαν από το διαδίκτυο. Είναι πολύ σημαντικό να τονίσουμε πως τα κείμενα δεν ανήκαν σε μία συγκεκριμένη κατηγορία αλλά σε όλες τις πιθανές κατηγορίες, πολιτικά, αθλητικά, κοινωνικά κ.ά. που μπορεί να ανήκει ένα κείμενο εφημερίδας. Τα αποτελέσματα αυτής της αξιολόγησης εμφανίζονται στον ακόλουθο πίνακα:

<u>Αριθμός κειμένων: 5(σε παρένθεση ο αριθμός προτάσεων)</u>
Best sentence split 100% (14/14)
Worst sentence split 94% (30/32)
Best true syntactic result 97%(27/28)
Worst true syntactic result 78%(11/14)

Πίνακας 7:Αποτελέσματα evaluation με κείμενα εφημερίδων

Στον παραπάνω πίνακα βλέπουμε τα αποτελέσματα του evaluation σε 5 κείμενα εφημερίδων. Βλέπουμε ότι ο καλύτερος διαχωρισμός προτάσεων(Best sentence split) έγινε σε ένα κείμενο με 14 προτάσεις όπου και οι 14 διαχωρίστηκαν σωστά. Ο χειρότερος διαχωρισμός προτάσεων (Worst sentence split) έγινε σε ένα κείμενο με 32 προτάσεις όπου δύο προτάσεις δεν αναγνωρίστηκαν σωστά. Αυτές δεν

αναγνωρίστηκαν με επιτυχία γιατί δεν ακολουθούσαν τους συντακτικούς κανόνες. Στη μία το ρήμα εννοούταν:

.... Έχουμε πλέον τη δυνατότητα να αποκτούμε κατοικίδια στα μέτρα μας. Γατάκια ή σκυλάκια υποαλλεργικά. Μετά από αυτή την ανακοίνωση...

Στην παραπάνω υπογραμμισμένη πρόταση, ο αναλυτής δε μπορεί να αναγνωρίσει ότι το ρήμα εννοείται. Η πρόταση αυτή εκμεταλλεύεται την ευχέρεια της ελληνικής γλώσσας, χωρίς όμως να ακολουθεί τους συντακτικούς της κανόνες. Στηρίζεται στο γεγονός πως κάποιος που θα διαβάσει και τις υπόλοιπες προτάσεις θα είναι σε θέση να κατανοήσει και αυτή την πρόταση. Ο αναλυτής όμως δεν μπορεί να κάνει μια τέτοια υπόθεση.

Στη δεύτερη πρόταση που δεν αναλύθηκε με επιτυχία, το πρόβλημα προέκυψε από τον τρόπο γραφής της. Αυτή στο τέλος άφηνε τρεις τελείες, δηλαδή:

... τα υπόλοιπα εννοούνται ... Έτσι λοιπόν...

Εδώ το πρόβλημα είναι ότι οι τρεις διαδοχικές τελείες, που για τον αναλυτή συμβολίζουν το τέλος μιας πρότασης, μπερδεύουν τον αναλυτή. Κάθε πρόταση μπορεί να τελειώνει με μία και μόνο τελεία. Οι τρεις τελείες δίνουν μεν έμφαση αλλά παράλληλα προκαλούν και σύγχυση στον συντακτικό αναλυτή ο οποίος δεν προχωρά σε σωστό διαχωρισμό προτάσεων.

Στον ίδιο πίνακα βλέπουμε και τα αποτελέσματα της συντακτικής ανάλυσης αυτών των κειμένων. Τα καλύτερα αποτελέσματα εμφανίζονται σε ένα κείμενο με 28 προτάσεις στο οποίο οι 27 προτάσεις αναλύθηκαν με επιτυχία (Best true syntactic result). Το μοναδικό λάθος εδώ προέκυψε από αστοχία στον προσδιορισμό ενός επιθετικού προσδιορισμού ο οποίος λόγω της ομοιότητάς του με υποκείμενο και επειδή το υποκείμενο εννοούταν, αναλύθηκε σαν αντικείμενο. Η πρόταση ήταν η ακόλουθη:

...Οι πλούσιοι μπορούν να αποκτήσουν ένα αυτοκίνητο...

Στην παραπάνω πρόταση, το υποκείμενο είναι οι άνθρωποι, το οποίο όμως εννοείται. Από την άλλη, ο επιθετικός προσδιορισμός πλούσιοι στο εννοούμενο

υποκείμενο, πληροί όλες τις προϋποθέσεις για να είναι αυτό το υποκείμενο και έτσι οδήγησε λάθος τον συντακτικό αναλυτή.

Τα χειρότερα αποτελέσματα εμφανίζονται στο κείμενο με τις 14 προτάσεις.. Τα λάθη που βρέθηκαν είναι παρόμοια με αυτό που αναλύθηκε παραπάνω. Στο κείμενο αυτό, παρότι ο διαχωρισμός των προτάσεων έγινε με επιτυχία 100% , η συντακτική ανάλυση 3 από αυτές δεν ήταν σωστή

Αυτό είναι λογικό, καθώς οι δύο διαδικασίες είναι πολύ διαφορετικές. Ο διαχωρισμός προτάσεων στηρίζεται σε διαφορετικά κριτήρια από αυτά του συντακτικού σχολιασμού των προτάσεων και αυτό επιτρέπει στο ίδιο κείμενο να εμφανίζει ταυτόχρονα τα καλύτερα αποτελέσματα στη μία ενέργεια και τα χειρότερα στην άλλη.

Επίσης έγινε σύγκριση της ακρίβειας των αποτελεσμάτων σε σχέση μια μία μέθοδο αναζήτησης που κάνει απλό keyword search, επάνω σε RSS(περιλήψεις κειμένων συνήθως με 2 – 3 προτάσεις). Τα αποτελέσματα παρουσιάζονται στον ακόλουθο πίνακα:

Ερωτήσεις	Αριθμός RSS που απαντούν ορθά την ερώτηση	Αποτελέσματα χωρίς συντακτική ανάλυση	Αποτελέσματα με συντακτική ανάλυση
Σκόραρε ο Ανρί;	15	18	15
Θα εκτονωθεί η ένταση;	1	3	1
Διάβασες το βιβλίο;	7	7	7
Ορκίστηκε ο Μεντβέντεφ;	1	2	1
Ποιος ορκίστηκε;	1	4	1
Που βρίσκεται η φλόγα;	2	8	2
Που βρίσκεται η ολυμπιακή φλόγα;	2	5	2
Πότε ενδέχεται να βρέξει;	1	4	1
Υπάρχει γατάκι υποαλεργικό	1	4	4
Δώσε μου τις κάρτες του Γιώργου	3	3	3

Πίνακας 8: Στην πρώτη στήλη του πίνακα εμφανίζονται οι ερωτήσεις. Στη δεύτερη εμφανίζεται ο αναμενόμενος αριθμός rss που απαντούν την ερώτηση χωρίς τη χρήση του συντακτικού αναλυτή, με απλό pattern match των keyword. Στην τρίτη και τελευταία στήλη εμφανίζονται τα αποτελέσματα με χρήση του συντακτικού μας αναλυτή.

Παρατηρούμε ότι με τη χρήση του αναλυτή τα αποτελέσματά μας είναι πάρα πολύ καλά. Αυτό συμβαίνει γιατί ο συντακτικός αναλυτής λαμβάνει υπόψιν του τη σημασιολογική πληροφορία της ερώτησης και του rss. Για παράδειγμα, στην πρώτη ερώτηση, χωρίς τον συντακτικό αναλυτή θα μας επιστραφούν όλα τα κείμενα που περιέχουν και τις δύο λέξεις κλειδιά «Σκόραρε» και «Ανρί». Αντίθετα, με τη χρήση του αναλυτή αναζητούνται τα Rss εκείνα τα οποία έχουν στην ίδια πρόταση το ρήμα «Σκόραρε» και με υποκείμενο το ουσιαστικό «Ανρί».

Πέρα από τα κείμενα των εφημερίδων και επειδή η γραφική διεπαφή που δημιουργήθηκε περιέχει και συντακτικό σχολιασμό ερωτήσεων, έπρεπε η εφαρμογή να αξιολογηθεί και σε ερωτήσεις. Οι ερωτήσεις όμως δεν εμφανίζονται συχνά σε κείμενα εφημερίδων. Έτσι, για την αξιολόγηση των ερωτήσεων, συλλέχθηκαν ερωτήσεις από τις 5 πιο δημοφιλείς ιστοσελίδες της Ελλάδας που βρίσκονται στην κατηγορία «Συχνές ερωτήσεις». Τα αποτελέσματα εδώ ήταν πάρα πολύ καλά καθώς ο συντακτικός αναλυτής έκανε συντακτική ανάλυση όλων των ερωτήσεων, περίπου 200, με επιτυχία κοντά στο 100%. Μερικά λάθη δεν οφείλονταν στο συντακτικό αναλυτή αλλά στον τρόπο που ήταν διατυπωμένη η ερώτηση, καθώς αυτός δεν ακολουθούσε τους συντακτικούς κανόνες και εκμεταλλευόταν την ευελιξία της ελληνικής γλώσσας.

Τέλος, το σύνολο των προτάσεων, οποιασδήποτε κατηγορίας, που βρέθηκαν μέσα στα 4 συντακτικά της νέας ελληνικής, στην προσπάθειά μας να εξάγουμε τους κανόνες που περιγράφηκαν παραπάνω, περνούν με επιτυχία τη συντακτική ανάλυση.

Όλα τα παραπάνω αποδεικνύουν πως η εφαρμογή μας μπορεί να ανταποκριθεί με επιτυχία σε κείμενα οποιασδήποτε κατηγορίας, αρκεί αυτά να ακολουθούν ορθά τους συντακτικούς κανόνες.

6.2 Συμπεράσματα

Η παρούσα διπλωματική εργασία πραγματεύεται την συντακτική ανάλυση της Νέας Ελληνικής. Η ελληνική γλώσσα είναι μια γλώσσα με πολύ πλούσια χαρακτηριστικά που αυτά όμως την κάνουν και ιδιαίτερα σύνθετη. Η συντακτική της ανάλυση θα πρέπει να λαμβάνει υπόψιν της ένα σύνολο κανόνων καθώς και την ευχέρεια της γλώσσας να αγνοεί τους κανόνες αυτούς.

Μετά και από την αξιολόγηση της εφαρμογής, τα συμπεράσματα είναι ιδιαίτερα ενθαρρυντικά. Το μοντέλο που κατασκευάστηκε και η αρχιτεκτονική που υλοποιήθηκε εμφανίζονται να λειτουργούν ορθά.

Γίνεται με πολύ μεγάλη επιτυχία ο εντοπισμός και διαχωρισμός προτάσεων μέσα σε κείμενα. Επίσης παρέχει τη δυνατότητα σύνδεσης προτάσεων αλλά και αναγνώρισης αυτής της σύνδεσης. Αυτή η λειτουργικότητα δεν είναι δυνατόν να προσφερθεί χωρίς τη χρήση τους συντακτικού αναλυτή.

Ακόμη, το μοντέλο και η αρχιτεκτονική που παρουσιάστηκαν, ανταποκρίνονται με επιτυχία στον προσδιορισμό συντακτικών όρων. Απλοί συντακτικοί όροι όπως το υποκείμενο το ρήμα και το αντικείμενο αναγνωρίζονται με επιτυχία. Επιπλέον, σύνθετοι συντακτικοί όροι, όπως ονοματικοί προσδιορισμοί, επιθετικοί προσδιορισμοί αλλά και συντακτικά φαινόμενα όπως η παράθεση και η επεξήγηση, επίσης αναγνωρίζονται.

Πολύ σημαντικά είναι και τα συμπεράσματα που προκύπτουν από την ανάλυση των ερωτήσεων. Κάθε ερώτηση, ανεξάρτητα σε ποια κατηγορία ανήκει αναγνωρίζεται με επιτυχία. Έτσι, είμαστε σε θέση να κατηγοριοποιούμε προτάσεις αλλά και να έχουμε πολύ καλά συντακτικά αποτελέσματα σε ορισμένες από αυτές.

Μαζί με αυτό, έρχεται και η επιτυχία του συντακτικού αναλυτή όταν χρησιμοποιείται ως μηχανή αναζήτησης. Επειδή η ερώτηση που εισάγεται αναγνωρίζεται συντακτικά σωστά σε πάρα πολλές περιπτώσεις, τα αποτελέσματα της μηχανής είναι ιδιαίτερα ενθαρρυντικά. Το μόνο που απομένει για την βελτίωση αυτών των αποτελεσμάτων είναι η βελτίωση της συντακτικής ανάλυσης των πηγών ελέγχου απαντήσεων των ερωτήσεων

Τέλος σημαντική συνεισφορά της παρούσας διπλωματικής εργασίας είναι η εφαρμογή του συντακτικού αναλυτή που προέκυψε. Αυτό είναι ιδιαίτερα σημαντικό καθώς δεν έχει παρουσιαστεί κάποια αντίστοιχη εργασία για την ελληνική γλώσσα. Όπως έχει γίνει και με άλλες γλώσσες, όπως την Αγγλική, η ανάπτυξη εφαρμογών φυσικής γλώσσας θέτει ως προϋπόθεση την ύπαρξη συντακτικού αναλυτή. Έτσι, η ύπαρξη ενός συντακτικού αναλυτή της ελληνικής γλώσσας θα δώσει ώθηση στην περαιτέρω ανάπτυξη των εφαρμογών ελληνικής φυσικής γλώσσας.

ΚΕΦΑΛΑΙΟ 7

ΑΝΑΚΕΦΑΛΑΙΩΣΗ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

7.1 Ανακεφαλαίωση διπλωματικής εργασίας

Σε αυτή την εργασία παρουσιάστηκε ο σχεδιασμός του μοντέλου και η υλοποίηση της αρχιτεκτονικής ενός συντακτικού αναλυτή για τη Νέα ελληνική γλώσσα. Επίσης παρουσιάστηκε μία γραφική διεπαφή η οποία πέρα από συντακτική ανάλυση, έχει τη δυνατότητα να χρησιμοποιηθεί και ως μηχανή αναζήτησης με συντακτικά κριτήρια.

Είναι γνωστό ότι για την ανάπτυξη εφαρμογών φυσικής γλώσσας είναι απαραίτητη η δυνατότητα συντακτικής υποστήριξης της γλώσσας. Επίσης είναι γνωστό πως ένα από τα μεγαλύτερα προβλήματα στη συντακτική ανάλυση μίας γλώσσας είναι η ασάφεια.

Η γνώση του πλαισίου στο οποίο μια ασάφεια εμφανίζεται είναι σημαντικό και κρίσιμο για την επίλυση του. Αυτή η παρατήρηση μας οδηγεί στη προσπάθεια εκμετάλλευσης των συντακτικών κανόνων που διέπουν αυτές τις ασάφειες. Το αποτέλεσμα είναι μία μοντελοποίηση της γλώσσας και των γνώσεων που προκύπτουν από το θεωρητικό υπόβαθρο. Η μεθοδολογία που έχουμε υλοποιήσει είναι επαναχρησιμοποιήσιμη, ανεξάρτητη περιοχής, και προσφέρει τη δυνατότητα επιπλέον επέκτασης. Η μεθοδολογία του διαχωρισμού προτάσεων που επίσης αναπτύχθηκε καταλήγει με επιτυχία στον προσδιορισμό όλων των προτάσεων ενώ εντοπίζει και τη σύνδεση προτάσεων με όλους τους δυνατούς συντακτικούς τρόπους που διαθέτει η ελληνική γλώσσα.

Οι παραπάνω μεθοδολογίες είναι ενσωματωμένες στο μοντέλο του συντακτικού αναλυτή ενώ παρουσιάστηκε και ένα γενικευμένο αρχιτεκτονικό πλαίσιο για τη κατασκευή και χρησιμοποίηση των διεπαφών φυσικής γλώσσας για τις αλληλεπιδράσεις χρηστών με αναλύσεις κειμένων. Η εφαρμογή που αναπτύχθηκε

εξετάζει την ακρίβεια τόσο του μοντέλου όσο και της αρχιτεκτονικής σε κείμενα με πλούσια σημασιολογική βάση.

Τα τμήματα λογισμικού του συντακτικού αναλυτή εξετάζουν ομοιόμορφα μια σειρά προβλημάτων στην ανάλυση της πρότασης τα οποία απαιτούν χωριστούς μηχανισμούς. Μια ενιαία αρχιτεκτονική χειρίζεται τη συντακτική και σημασιολογική ανάλυση, αλλά και τις ασάφειες. Παράλληλα, το πλαίσιο έχει κατασκευαστεί με τέτοιο τρόπο έτσι ώστε να αποφεύγονται οι εξαρτήσεις με συγκεκριμένες κατηγορίες ερωτήσεων ή κείμενα συγκεκριμένων γνωστικών πεδίων με αποτέλεσμα να είναι επαναχρησιμοποιήσιμο σε διαφορετικές εφαρμογές με διαφορετική σημασιολογία περιοχών.

7.2 Μελλοντικές επεκτάσεις

Όσον αφορά το γραφικό περιβάλλον εργασίας και λαμβάνοντας υπόψιν τα διάφορα σχόλια που προέκυψαν από την αξιολόγηση του συστήματος θα μπορούσαν να προταθούν ως μελλοντική επέκταση τα εξής:

- Να δημιουργηθούν και να εφαρμοστούν περισσότεροι κανόνες – κανονικές εκφράσεις. Έτσι το σύστημα θα είναι σε θέση να αυξήσει ακόμη περισσότερο και την πολυπλοκότητα των κειμένων που αναλύει αλλά και την επιτυχία του στις αναλύσεις.
- Εφαρμογή αναζήτησης ερωτήσεων με κριτήρια συγκεκριμένων συντακτικών όρων. Αφού η συντακτική ανάλυση των κειμένων που εισάγονται στη διεπαφή υπάρχει, μπορεί να σχεδιαστεί η αναζήτηση ερωτήσεων στις οποίες το κριτήριο είναι συντακτικό και μόνο. Για παράδειγμα, να αναζητηθούν οι προτάσεις που έχουν ως υποκείμενο μια λεκτική μονάδα ή να βρεθούν τα υποκείμενα όλων των προτάσεων που έχουν ως αντικείμενα μια άλλη λεκτική μονάδα.
- Χρήση ενός λεξικού για μεγαλύτερο εύρος αναζήτησης όρων και πληρέστερα αποτελέσματα
- Εμφάνιση των αποτελεσμάτων με σειρά προτεραιότητας έτσι ώστε να μπορούμε να αξιολογήσουμε τη μηχανή αναζήτησης με περισσότερα κριτήρια

και να συγκρίνουμε τα αποτελέσματά της με αποτελέσματα άλλων μηχανών αναζήτησης

Βιβλιογραφία

- [1] Γραμματική της ελληνικής γλώσσας, 2000, Holton David
- [2] Συντακτικό της νέας ελληνικής, 1987, Οργανισμός Εκδόσεων Διδακτικών Βιβλίων
- [3] Νεοελληνική γραμματική, 1987, Τριανταφυλλίδης Μανόλης
- [4] Αναλυτική γραμματική της νέας ελληνικής, 2002, Αναστάσιος Λώλος, Δημήτρης Κοσμάς, Σαρισχούλης Στράτος
- [5] A.Karanastasi 2007. OntoNL: An Ontology- Based Natural Language Interface Generator for Knowledge Repositories, Master Thesis, Technical University of Crete, Department of Electronic and Computer Engineering
- [6] Karanastasi, S. Christodoulakis: *"Semantic Processing of Natural Language Queries in the OntoNL Framework"*, in the Proceedings of the IEEE International Conference on Semantic Computing (IEEE ICSC), 17-19 September 2007, Irvine, CA
- [7] A. Karanastasi, S. Christodoulakis: *"Ontology-Driven Semantic Ranking for Natural Language Disambiguation in the OntoNL Framework"*, in the Proceedings of the 4th European Semantic Web Conference (ESWC), 3-7 June 2007, Innsbruck, Austria
- [8] A. Karanastasi, S. Christodoulakis: *"The OntoNL Semantic Relatedness Measure for OWL Ontologies"*, in the Proceedings of the Second IEEE International Conference on Digital Information Management (IEEE ICDIM), 28-31 October 2007, Lyon, France
- [9] A. Karanastasi, A. Zotos, S. Christodoulakis: *"The OntoNL Framework for Natural Language Interface Generation and a Domain-Specific Application"*, in the Proceedings of the DELOS Conference on Digital Libraries 2007, 13-14 February 2007, Tirrenia, Pisa, Italy
- [10] A. Karanastasi, A. Zotos, S. Christodoulakis: *"User Interactions with Multimedia Repositories using Natural Language Interfaces - OntoNL: an Architectural Framework and its Implementation"*, in Journal of Digital Information Management - JDIM, Volume 4, Issue 4, December 2006
- [11] A. Karanastasi, A. Zotos, S. Christodoulakis: *"User Interactions with Multimedia Repositories using Natural Language Interfaces:"*

- an Architectural Framework and its Implementation",*
in Proceedings of the Fourth Special Workshop on Multimedia Semantics - WMS06,
June 19-21, 2006
- [12] S. Christodoulakis, A. Karanastasi, J. Koehler, K. Biatov, T. Catarci, S. Kimani:
"Natural Language and Speech Interfaces to Knowledge Repositories",
Poster on the 9th European Conference on Research and Advanced Technology for
Digital Libraries (ECDL 2005), September 2005, Vienna, Austria
- [13] A. Karanastasi, S. Christodoulakis:
*"OntoNL: An Ontology-based Natural Language Interface Generator for Multimedia
Repositories",*
In Proceedings of the AVIVDiLib Seventh International Workshop of the EU
Network of Excellence DELOS on AUDIO-VISUAL CONTENT AND
INFORMATION VISUALIZATION IN DIGITAL LIBRARIES (AVIVDiLib'05),
May 2005
- [14] Orphanos G. and Christodoulakis D. 1999. "Part-of-speech Disambiguation
and Unknown Word Guessing with Decision Trees", *Proceedings of EACL '99*,
Bergen, Norway.
- [15] Orphanos G. and Tsalidis C. 1999. "Combining handcrafted and corpus-
acquired Lexical Knowledge into a Morphosyntactic Tagger", *Proceedings of the 2nd
Research Colloquium for Computational Linguistics in United Kingdom (CLUK)*,
Essex, UK.
- [16] Orphanos G., Kalles D., Papagelis A. and Christodoulakis D. 1999. "Decision
Trees and Natural Language Processing: A case study in Part-of-Speech Tagging",
*Proceedings of ACAI99 Workshop on Machine Learning in Human Language
Technology*, Chania, Greece.
- [17] Ορφανός Γ., Γάκης Π. και Ιορδανίδου Α. 1999. "Μορφοσυντακτική ασάφεια
στη νέα ελληνική: η περίπτωση επιθέτου - ουσιαστικού - επιρρήματος" *Πρακτικά της
20ης Συνάντησης του Τομέα Γλωσσολογίας του Τμήματος Φιλολογίας*, Θεσσαλονίκη,
Ελλάδα.
- [18] Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge,
Massachusetts.
- [19] Pollard, C. and Sag, A. 1987. *Information-Based Syntax and Semantics*. Vol.
1: CSLI Publications.

- [20] Kilgarrieff, A. 1992. *Polysemy*, Ph.D. Thesis, University of Sussex.
- [21] Wilks, Y. and Stevenson, M. 1997. "The Grammar of Sense: Using part-of-speech tags as a first step in semantic disambiguation", *Natural Language Engineering*, 1(1).
- [22] Lyons, J. 1977. *Semantics*. Vol. 1 & 2, Cambridge University Press.
- [23] Γάκης Π., Ορφανός Γ. και Ιορδανίδου Ά. 1999. "Υπολογιστική Επεξεργασία της Νέας Ελληνικής: Καταγραφή της Μορφοσυντακτικής Ασάφειας", *Γλώσσα*, τεύχος 49, Εκδόσεις Νέας Παιδείας.
- [24] Tsalidis C. and Orphanos G. 1995, "Word Description Languages", *Proceedings of the 1st Workshop in Natural Language Processing*, Athens, Greece.
- [25] Jurafsky, D., Martin, J.H. (2000). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New Jersey: Prentice Hall.
- [26] Russel, S., Norvig, P. (1995). *Artificial Intelligence. A Modern Approach*. New Jersey: Prentice Hall.
- [27] Papakitsos, E., Grigoriadou, M. and Ralli, A. 1998. "Lazy Tagging with Functional Decomposition and Matrix Lexica: An Implementation in Modern Greek", *Literary and Linguistic Computing*, 13(4), Oxford University Press.
- [28] Philippaki-Warbuton, I. 1985. "Word-order in Modern Greek", *Transaction of the Philological Society*, 113-143.
- [29] Φιλippάκη-Warbuton, E. 1992. *Εισαγωγή στη θεωρητική γλωσσολογία*. Εκδόσεις Νεφέλη, Αθήνα.
- [30] Pollard, C. and Sag, A. 1987. *Information-Based Syntax and Semantics*. Vol. 1: CSLI Publications.
- [31] Ralli, A. and Galiotou, E. 1987. "A Morphological Processor for Modern Greek", *Proceedings of EACL '87*, Copenhagen, Denmark.
- [32] Tzanidaki, D.I. 1993. *Word-order in Modern Greek*, M.A. dissertation, University College of London.
- [33] Vagelatos A., Triantopoulou T., Tsalidis C. and Christodoulakis D. 1995b. "A Spelling Correction System for Modern Greek", *International Journal of Artificial Intelligence & Tools*.
- [34] Η γλώσσα TCL www.tcl.tk

Παράρτημα Ι. Μορφοσυντακτικά χαρακτηριστικά

Στη συνέχεια παρουσιάζονται τα μορφοσυντακτικά χαρακτηριστικά και οι τιμές τους. Οι χαρακτήρες κάθε λέξης που είναι έντονα μαύροι δημιουργούν μια συντομογραφία για αυτή τη λέξη. Π.χ. η συντομογραφία για το εμπρόθετο άρθρο είναι **ΕμΆρθ**.

Μέρος του Λόγου – Part Of Speech = {αντωνυμία – **pronoun**, προσωπική αντωνυμία – **clitic**, κτητική αντωνυμία (ασθενής τύπος) – **possessive pronoun**, άρθρο – **article**, εμπρόθετο άρθρο – **prepositional article**, επίθετο – **adjective**, επίρρημα – **adverb**, επιφώνημα – **exclamation**, μετοχή – **participle**, μόριο – **particle**, ουσιαστικό – **noun**, πρόθεση – **preposition**, ρήμα – **verb**, σύνδεσμος – **conjunction**}

Γένος – Gender = {αρσενικό – **masculine**, θηλυκό – **feminine**, ουδέτερο – **neuter**}

Αριθμός – Number = {ενικός – **singular**, πληθυντικός – **plural**}

Πτώση – Case = {ονομαστική – **nominative**, γενική – **genitive**, αιτιατική – **accusative**, κλητική – **vocative**}

Βαθμός – Degree = {θετικός – **basic**, συγκριτικός – **comparative**, υπερθετικός – **superlative**}

Πρόσωπο – Person = {Α', Β', Γ' – C'}

Χρόνος – Tense = {ενεστώτας – **present**, αόριστος – **past**, παρατατικός – **past continuous**, παρακείμενος – **present perfect**, υπερσυντέλικος – **past perfect**, συνοπτικός μέλλοντας – **future**, εξακολουθητικός μέλλοντας – **future continuous**, συντελεσμένος μέλλοντας – **future perfect**}

Έγκλιση – Mood = {οριστική – **indicative**, υποτακτική – **subjunctive**, προστακτική – **imperative**}

Φωνή – Voice = {ενεργητική – **active**, παθητική – **passive**}

Κατάληξη – Suffix = { "τατα", "ουσα", "μενα", "ιτσα", "κια", "ικα", "ια", "α", "αι", "οι",
"κι", "ει", "ι", "ουσαν", "μενων", "θεν", "δων", "ουν", "ην", "ων", "ον", "αν",
"ρν", "μενους", "μενος", "μενης", "μενες", "ουσες", "τατες", "τερες", "τερης",
"εντες", "ικης", "ντας", "δες", "εις", "ους", "ας", "ες", "ης", "νς", "ος", "ρς",
"τς", "ως", "τερου", "μενου", "ικου", "ου", "υ", "τερη", "μενη", "τατη", "ικη",
"θη", "η", "τερο", "μενο", "ειτο", "ικο", "ετο", "ο", "ε", "υ", "ζ", "κ", "λ", "μ",
"τ", "ξ", "ρ", "φ" }

Στίξη – Punctuation = { '-', '!', '/', '(', ')', ',', '.', ':', ';', '>', '<' }

Αρχίζει με κεφαλαίο = {ναι, όχι}

Είναι ξένη λέξη = {ναι, όχι}

Κοτόπουλο