



Οι Υποκλάδοι της Γλωσσολογίας και η Σύνδεσή τους με την Ανάκτηση της Πληροφορίας

Λίβερης Αναστάσιος

Τμήμα Αρχειονομεία, Βιβλιοθηκονομίας & Μουσειολογίας

Μεταπτυχιακό Δίπλωμα στην

«Ψηφιακή Επεξεργασία Πολιτιστικής Πληροφορίας»

Περίληψη

Η Ανάκτηση της Πληροφορίας τα τελευταία χρόνια λόγω του διαδικτύου και του μεγάλου αριθμού ιστοσελίδων έχει ως βασικό στόχο την ανάπτυξη αποτελεσματικών εφαρμογών που θα μπορούν να διαχειρίζονται όλο αυτό το ψηφιακό υλικό. Στο σκοπό αυτό συμβάλει η επιστήμη της Γλωσσολογίας και των υποκλάδων της καθώς ο βασικός τύπος των αρχείων που ανακτώνται είναι σε μορφή κειμένου . Στην παρούσα διπλωματική εργασία παρουσιάζονται οι βασικοί υποκλάδοι της γλωσσολογίας και γίνεται προσπάθεια της αναζήτησης της σχέσεώς τους με την ανάκτηση της πληροφορίας μέσω διάφορων εφαρμογών ανάκτησης που έχουν αναπτυχθεί τα τελευταία χρόνια.

Abstract

In recent years Information Retrieval, due to the internet and the large number of web sites, aims at the development of effective applications that can handle all this digital material. For this reason, the science of Linguistics and its subclasses can contribute to this cause because the basic type of retrieved files is in text format. In this thesis the main branches of linguistics are presented and it is attempted to explore the relationship between them and the information retrieval through various retrieval applications that have been developed in recent years.

Περιεχόμενα

Κεφάλαιο 1

1. Εισαγωγή	σελ.14
1.1. Γλωσσολογία	σελ.14
1.1.1. Ιστορική αναδρομή	σελ.16
1.1.1.1. Από την Εποχή του Χαλκού ως τη Ρωμαϊκή Περίοδο	σελ.16
1.1.1.2. Μεσαίωνας	σελ. 18
1.1.1.3. Αναγέννηση	σελ. 19
1.1.1.4. Νεότερη Γλωσσολογία	σελ. 19
1.1.2. Οι υποκλάδοι της Γλωσσολογίας.....	σελ. 20
1.2. Η ανάκτηση της πληροφορίας.....	σελ. 21
1.2.1 Οι μηχανές αναζήτησης.....	σελ.23
1.2.2 Από τα πρώτα συστήματα ανάκτησης ως το σημασιολογικό ιστό.....	σελ.23

Κεφάλαιο 2

2.1. Η σχέση της γλωσσολογίας με την ανάκτηση της πληροφορίας.....	σελ.25
2.2. Η φωνητική.....	σελ.26
2.2.1. Το ανθρώπινο φωνητικό σύστημα	σελ.26
2.2.2. Το Διεθνές Φωνητικό Αλφάβητο	σελ.27
2.2.3. Κατηγορίες Φωνητικής.....	σελ.28
2.2.3.1 Αρθρωτική φωνητική.....	σελ.28
2.2.3.1.1. Αρθρωση συμφώνων.....	σελ.28
2.2.3.1.2. Αρθρωση φωνέντων.....	σελ.29
2.2.3.2. Ακουστική φωνητική.....	σελ.29
2.2.3.3. Ακροατική φωνητική.....	σελ.30
2.2.4. Συστήματα αναγνώρισης φωνής.....	σελ.31
2.3. Η φωνολογία.....	σελ.32
2.3.1. Τα φωνήματα.....	σελ.33

2.3.2. Τα αλλόφωνα.....	σελ.33
2.3.3. Φωνοτακτικοί κανόνες.....	σελ.34
2.3.4. Ο τονισμός.....	σελ.35
2.3.5. Σύνδεση της Φωνολογίας με την Ανάκτηση της Πληροφορίας..	σελ 35
2.3.5.1. Φωνολογικός Βρόγχος του μοντέλου μνήμης.....	σελ 35
2.3.5.2.Συστήματα εισόδου και εξόδου ομιλίας	σελ.38
2.4. Η Μορφολογία	σελ.38
2.4.1. Κατηγορίες μορφημάτων.....	σελ.39
2.4.2. Οι λέξεις.....	σελ.40
2.4.3. Βασικές μορφολογικές διαδικασίες	σελ.41
2.4.4. Περιστολή (stemming)	σελ.41
2.4.5. Υπολογιστική μορφολογία.....	σελ.43
2.5. Η Σύνταξη	σελ.43
2.5.1. Τα συστατικά μιας πρότασης.....	σελ.44
2.5.2. Δενδρική δομή πρότασης.....	σελ.45
2.5.3. Η μορφοσύνταξη	σελ.46
2.5.4. Η γραμματική	σελ.46
2.5.5. Αυτόματη επεξεργασία φυσικής γλώσσας	σελ.48
2.6. Η Σημασιολογία.....	σελ.49
2.6.1. Οι κατηγορίες της σημασιολογίας.....	σελ.49
2.6.2. Η σημασία της λέξης.....	σελ.50
2.6.3. Σημασιολογικές σχέσεις των λέξεων.....	σελ.50
2.6.4. Η σημασιολογική συνάφεια στην ανάκτηση της πληροφορία.....	σελ.51
2.6.5. Οι αλγόριθμοι κατάταξης.....	σελ.52

2.6.6. Οι σημασιολογικές μηχανές αναζήτησης.....σελ.53	σελ.53
2.7. Η Πραγματολογία	σελ.53
2.7.1. Προβλήματα που δυσχεραίνουν την επικοινωνία.....	σελ.54
2.7.2. Γλωσσικές προσεγγίσεις	σελ.56
2.7.3. Ο ρόλος της πραγματολογίας στην ανάκτηση της πληροφορίας...	σελ.57
2.7.3.1. Αυτόματη περίληψη κειμένου.....	σελ.57
2.7.3.2. Αποσαφήνιση νοήματος διφορούμενων λέξεων.....	σελ.58
2.7.3.2.1. Knowledge Based προσέγγιση.....	σελ.59
2.7.3.2.2. Επιβλεπόμενη μάθηση.....	σελ.60
2.7.3.2.3. Μη επιβλεπόμενη μάθηση.....	σελ.61
2.8. Η Κοινωνιολογία	σελ.61
2.8.1. Η γλωσσική ποικιλία.....	σελ.62
2.8.2. Η γλωσσική επικοινωνία.....	σελ.63
2.8.3. Γλωσσικές ανισότητες.....	σελ.64
2.8.4. Social Information Retrieval.....	σελ.64
2.8.4.1. Κοινωνικές μηχανές αναζήτησης	σελ.65
2.9. Η Φιλοσοφία της Γλώσσας.....	σελ.66
2.9.1. Το περιεχόμενο της Φιλοσοφίας της Γλώσσας.....	σελ.67
2.9.2. Ιστορική αναδρομή	σελ.67
2.9.3. Οι λειτουργίες της γλώσσας	σελ.68
2.9.4. Συστήματα ερωτήσεων απαντήσεων	σελ.69
2.10. Η Ψυχολογία της Γλώσσας	σελ.70
2.10.1. Η απόκτηση της γλώσσας	σελ.71
2.10.2. Η σχέση σκέψης – γλώσσας	σελ.71

2.10.3. Η κατανόηση της γλώσσας	σελ.72
2.10.4. Η παραγωγή της γλώσσας	σελ.73
2.10.5. Η μνήμη της γλώσσας.....	σελ.74
2.10.6. Η Πρόβλεψη Λέξεων (Word Prediction)	σελ.74
2.11. Η Νευρογλωσσολογία	σελ.75
2.11.1. Η ανατομία του εγκεφάλου	σελ.75
2.11.2. Η παιδική ηλικία	σελ.76
2.11.3. Οι λειτουργίες του εγκεφάλου	σελ.77
2.11.4. Γλωσσικές διαταραχές και γλωσσολογία	σελ.78
2.11.5. Μηχανισμοί σύνθεσης ομιλίας	σελ.78
2.11.6. Τεχνητά νευρωνικά δίκτυα	σελ.79
2.12. Μαθηματική Γλωσσολογία	σελ.80
2.12.1. Οι γραμματικές	σελ.80
2.12.1.1. Γενετική μετασχηματιστική γραμματική.....	σελ.81
2.12.1.2. Κατηγορική Γραμματική	σελ.81
2.12.2. Σύνδεση της Μαθηματικής Γλωσσολογία με την Ανάκτηση της Πληροφορίας	σελ.82
2.13. Η Λεξικογραφία	σελ.83
2.13.1 Το λεξικό	σελ.83
2.13.2. Είδη λεξικών	σελ.84
2.13.3. Ο ρόλος των λεξικών στην Ανάκτηση της Πληροφορίας	σελ.85
2.13.3.1. Wordnet	σελ.86
2.13.3.2. Η χρήση των λεξικών σε εφαρμογές	σελ.86
2.14. Η Τεκμηρίωση της Γλώσσας	σελ.87

2.14.1. Τα δεδομένα	σελ.88
2.14.2. Η σχέση με την Ανάκτηση της Πληροφορίας	σελ.89
2.14.2.1. Η αρχειοθέτηση	σελ.90
2.15. Η Εκμάθηση της Γλώσσας	σελ. 91
2.15.1. Βασικές κατηγορίες μάθησης της γλώσσας	σελ.91
2.15.1.1. Η Αντίληψη του Λόγου (Speech Perception)	σελ.92
2.15.1.2. Η Παραγωγή του Λόγου (Speech Production)	σελ.92
2.15.2. Επίπεδα Γλωσσικής Ανάπτυξης	σελ.93
2.15.2.1. Η Κατάκτηση της Φωνολογίας.....	σελ.93
2.15.2.2. Η Κατάκτηση της Σύνταξης	σελ.93
2.15.2.3. Η Κατάκτηση της Σημασιολογίας	σελ.94
2.15.3. Θεωρητικές προσεγγίσεις	σελ.94
2.15.3.1. Οι απόψεις του Jean Piaget	σελ.95
2.15.4. Το σύστημα ανταλλαγής δεδομένων παιδικής γλώσσας	σελ.96
2.16. Ιστορική Γλωσσολογία	σελ.97
2.16.1. Μέθοδοι Ιστορικής Γλωσσολογίας	σελ.98
2.16.1.1. Η Ιστορικοσυγκριτική Γλωσσολογία	σελ.98
2.16.1.2. Η μέθοδος της εσωτερικής ανασυγκρότησης	σελ.99
2.16.2. Βασικές αρχές της Ιστορικής Γλωσσολογίας	σελ.99
2.16.3. Χρονολόγηση μεταβολών	σελ.100
2.16.4. Η ετυμολογία των λέξεων	σελ.100
2.16.5. Η σχέση με την Ανάκτηση της Πληροφορίας	σελ.101
2.16.5.1. Το φαινόμενο των greeklish	σελ.102
2.17. Γενική Γλωσσολογία	σελ.102

2.17.1. Η ομιλία	σελ.103
2.17.2. Η γλώσσα	σελ.104
2.17.3. Η δομή της γλώσσας	σελ.105
2.17.4. Οι ομοιότητες μεταξύ διαφόρων γλωσσών	σελ.106
2.17.5. Η γλώσσα και η ομιλία ως μέσω επικοινωνίας στο διαδίκτυο.....	σελ.106
2.18. Δικανική Γλωσσολογία	σελ.107
2.18.1. Η γλώσσα των νομικών κειμένων	σελ.109
2.18.2. Η γλώσσα των νομικών διαδικασιών	σελ.109
2.18.3. Η λογοκλοπή	σελ.110
2.18.4. Η σύνδεση με την ανάκτηση της πληροφορίας	σελ.111
2.19. Η ανάλυση του λόγου	σελ.112
2.19.1. Η κριτική ανάλυση του λόγου	σελ.113
2.19.2. Ανάλυση συναισθημάτων (sentimental analysis)	σελ.115.
2.20. Γλωσσολογία και Λογοτεχνία	σελ.116
2.20.1. Λογοτεχνικό ύφος	σελ.118
2.20.2. Σημασιολογική δομή στην ποίηση	σελ.119
2.20.3. Απόδοση πατρότητας σε κείμενα (Author Recognition).....	σελ.119
2.21. Υπολογιστική Γλωσσολογία	σελ.121
2.21.1. Τα επίπεδα της Υπολογιστικής Γλωσσολογίας	σελ.122
2.21.2. Βασικές μαθηματικές έννοιες	σελ.123
2.21.3. Σύνδεση με την Ανάκτηση της Πληροφορίας	σελ.124
2.22. Η Γνωσιακή Επιστήμη	σελ.125
2.22.1. Δυισμός	σελ.126
2.22.2. Σημασιολογικά πλαίσια	σελ.127

2.22.3. Κατηγοριοποίηση Κειμένου (Text Categorization)	σελ.128
2.23. Η Εφαρμοσμένη Γλωσσολογία	σελ.129
2.23.1. Η εκμάθηση ξένης γλώσσας	σελ.130
2.23.2. Γλωσσικός σχεδιασμός	σελ.131
2.23.3. Σύνδεση με την Ανάκτηση της Πληροφορίας	σελ.131
2.24. Συστήματα Γραφής	σελ.132
2.24.1. Είδη συστημάτων γραφής	σελ.132
2.24.2. Σημαντικά αλφάβητα	σελ.134
2.24.3. Σύστημα γραφής Braille	σελ.135
2.24.4. Οπτική Αναγνώριση Χαρακτήρων	σελ.135
2.25. Τυπολογία	σελ.136
2.25.1. Ομοιότητες βάσει γλωσσικής οικογένειας	σελ.136
2.25.2. Ομοιότητες βάσει μορφολογίας	σελ.137
2.25.3. Ομοιότητες βάσει καθολικών χαρακτηριστικών	σελ.138
2.25.4. Cross – Lingual Retrieval	σελ.139
2.26. Διερμηνεία και Μετάφραση	σελ.140
2.26.1. Η Μετάφραση	σελ.141
2.26.2. Οι μεταφραστικές μονάδες	σελ.142
2.26.3. Η σύνδεση με την Ανάκτηση της Πληροφορίας	σελ.143
2.26.3.1. Η Μηχανική Μετάφραση	σελ.143
2.26.3.2. Η αυτοματοποιημένη διερμηνεία	σελ.144
2.27. Σώματα κειμένων	σελ.145
2.27.1. Κριτήρια δημιουργίας σωμάτων κειμένου	σελ.145
2.27.2. Οι υπογλώσσες	σελ.146

2.27.3. Είδη σωμάτων κειμένου	σελ.147
2.27.4. Η σύνδεση με την Ανάκτηση της Πληροφορίας	σελ.147
2.27.4.1. Ανάλυση σωμάτων κειμένου	σελ.148
2.27.4.2. Ηλεκτρονικά σώματα κειμένου	σελ.148
2.27.4.3. Κανονικές Εκφράσεις	σελ.149
2.28. Η Ανθρωπογλωσσολογία	σελ.149
2.28.1. Σχέση γλώσσας – κουλτούρας	σελ.150
2.28.2. Σχέση γλώσσας –γεωγραφίας	σελ.150
2.28.3. Σχέση γλώσσας – σκέψης	σελ.151
2.28.4. Geographic Information Retrieval	σελ.152

Κεφάλαιο 3

3. Συμπεράσματα	σελ.153
Βιβλιογραφία	σελ.154

Κεφάλαιο 1^ο

1. Εισαγωγή

Πολύ μεγάλος αριθμός αρχείων κειμένου βρίσκεται στο διαδίκτυο, σε βάσεις δεδομένων και σε βιβλιοθήκες. Η αξιοποίηση και η εύκολη πρόσβαση σε όλον αυτό τον μεγάλο όγκο πληροφοριών δεν είναι τόσο απλή υπόθεση. Γι' αυτό το λόγο η ανάπτυξη τεχνικών για την διαχείριση και εκμετάλλευση του προκειμένου να παραχθεί γνώση, απαιτεί συνδυασμό των νέων τεχνολογιών λαμβάνοντας υπόψη όμως και μεθόδους επεξεργασίας της γλώσσας.

Η επιστήμη της γλώσσας κατά τον τελευταίο αιώνα έχει σημειώσει πολύ μεγάλη πρόοδο και μάλιστα παίρνοντας στοιχεία από άλλες επιστήμες έχει διευρυνθεί τόσο, ώστε να είναι απαραίτητη η δημιουργία ξεχωριστών υποκλάδων της. Το ζητούμενο είναι με ποιον τρόπο όλοι αυτοί οι υποκλάδοι συνεισφέρουν, συνεργάζονται και αξιοποιούνται από τις νέες πρακτικές που προσφέρει η τεχνολογία, προκειμένου η ανάκτηση κειμενικού και γλωσσικού υλικού να βρίσκει εφαρμογή στην καθημερινότητα του ανθρώπου αλλά και σε διάφορα άλλα επιστημονικά πεδία.

Στα πλαίσια της παρούσας διπλωματικής εργασίας πραγματοποιήθηκε περιγραφή των βασικότερων υποκλάδων της γλωσσολογίας και για κάθε μια από αυτές αναφέρονται βασικά χαρακτηριστικά και στοιχεία. Επιπλέον δεδομένου ότι το κύριο προϊόν της ανάκτησης της πληροφορίας είναι κειμενικής μορφής, γίνεται προσπάθεια μελέτης της σχέσης της γλώσσας και της γλωσσολογίας με την ανάκτηση της πληροφορίας. Συγκεκριμένα για κάθε υποκλάδο της γλωσσολογίας εκτός από μια σύντομη περιγραφή γίνεται και προσπάθεια εντοπισμού της σχέσης του με κάποια από τις εφαρμογές και λειτουργίες της Ανάκτησης της Πληροφορίας.

1.1. Η γλωσσολογία

Η επιστήμη της γλωσσολογίας μελετάει και αναλύει το λόγο. Η μελέτη του λόγου αποτελείται από δύο μέρη τα οποία είναι η μελέτη της γλώσσας και η μελέτη της ομιλίας. Η ανθρώπινη γλώσσα αποτελείται από ένα σύνολο ήχων και φθόγγων που

έχουν κάποιο νόημα και δίνουν τη δυνατότητα στον άνθρωπο να επικοινωνεί και να εκφράζει τις σκέψεις του. Η ομιλία περιλαμβάνει τη φώνηση¹ δηλαδή τα ακουστικά σήματα που παράγουν τα όργανα του ανθρώπινου φωνολογικού συστήματος. Η γλώσσα περιλαμβάνει όλο το γραμματικό σύστημα το οποίο είναι εσωτερικευμένο στον ομιλητή και το χρησιμοποιεί για να σκέφτεται, να εκφράζεται και να επικοινωνεί. Ο λόγος είναι η ικανότητα του ανθρώπου να επικοινωνεί χρησιμοποιώντας λέξεις και γραμματικές και συντακτικές δομές. Αν και διαφέρουν μεταξύ τους η γλώσσα και ο λόγος, μοιράζονται πολλά κοινά σημεία και υπάρχει αλληλεξάρτηση μεταξύ τους καθώς η ομιλία πραγματοποιείται μέσω της γλώσσας, ενώ απαραίτητη προϋπόθεση για την επικοινωνία είναι η ομιλία .

Οι δύο σημαντικοί γλωσσολόγοι οι οποίοι συνέβαλλαν στην αλλαγή του τρόπου μελέτης της γλώσσας και στην καθιέρωση της ως αυτόνομη επιστήμη ήταν ο Ferdinand de Saussure και ο Noam Chomsky. Πριν την διατύπωση των θεωριών τους η μελέτη της γλώσσας επικεντρώνονταν αποκλειστικά στη γραμματική και τη φιλολογία.

Ο Ferdinand de Saussure θεώρησε ότι η γλωσσολογία θα έπρεπε να καθιερωθεί ως αυτόνομη επιστήμη. Βασικό σημείο της θεωρίας του ήταν ότι η γλώσσα περιλαμβάνει το γλωσσικό σημείο το οποίο αποτελείται από δύο μέρη: το «σημαίνον» δηλαδή ένα αντικείμενο ή μια ιδέα και το «σημαινόμενο» δηλαδή μια ακουστική εικόνα η οποία αποτελεί την αντίστοιχη ονομασία του².

Ο Noam Chomsky ασχολήθηκε με ζητήματα όπως η κατάκτηση της γλώσσας από τους ανθρώπους και τη δομή και λειτουργία του ανθρώπινου νου. Υποστήριξε την ύπαρξη μιας καθολικής γραμματικής στην οποία οφείλεται η ικανότητα του ανθρώπου να μαθαίνει γρήγορα τη μητρική του γλώσσα και να παράγει και να κατανοεί πολύ μεγάλο αριθμό φράσεων.

Η Γλωσσολογία στη σημερινή της μορφή όπως έχει διαμορφωθεί με το πέρασμα των χρόνων μέσα από τα έργα μεγάλων γλωσσολόγων και προσπαθεί να απαντήσει σε διάφορα ερωτήματα όπως: ποια είναι η φύση της γλώσσας και η σχέση της με το νου, αν υπάρχουν ομοιότητες μεταξύ διαφορετικών γλωσσών κ.α. .

¹ De Saussure,F.(1979). *Μαθήματα Γενικής Γλωσσολογίας*. Αθήνα: Εκδόσεις Παπαζήση.

² De Saussure,F.(1979). *Μαθήματα Γενικής Γλωσσολογίας*. Αθήνα: Εκδόσεις Παπαζήση.

1.1.1. Ιστορική Αναδρομή

Η ιστορία της γλωσσολογίας ή αλλιώς η «Ιστορία της Γλωσσικής σκέψης», είναι πολύ μεγάλη και ξεκινάει από την Αρχαιότητα συνεχίζει τη Ρωμαϊκή και Βυζαντινή Περίοδο τον Ευρωπαϊκό Μεσαίωνα και την Αναγέννηση. Σταθμός στη γλωσσολογία θεωρείται το έργο του John Fearn «Grammatical Opus Magnum» το 1824³. Εν συνεχεία σταθμό αποτέλεσε η θεωρία του στρουκτουραλισμού του Ferdinand Saussure και αργότερα ο Νοαμ Τσόμσκι με τη Καθολική Γραμματική.

1.1.1.1 Από την Εποχή του Χαλκού ως τη Ρωμαϊκή Περίοδο

Μια από τις αρχαιότερες γλωσσικές παραδόσεις συναντάμε στην Αρχαία Μεσοποταμία στις περιοχές μεταξύ των ποταμών Τίγρη και Ευφράτη το 3000 π.Χ. και διήρκεσε περίπου 2500 χρόνια. Τα παλαιότερα κείμενα που υπάρχουν ήταν γραμμένα με σφηνοειδή γραφή στη Σουμεριακή Γλώσσα η οποία ήταν μια γλώσσα χωρίς κάποια συγγένεια με σύγχρονες γλώσσες. Στη Σουμεριακή Γλώσσα είναι γραμμένα τα πρώτα σωζόμενα κείμενα και άρα θεωρείται το πρώτο είδος γραφής. Ένα από τα παλαιότερα κείμενα είναι «Το βιβλίο της Σουμερικής Λίστας των Βασιλέων». Η σουμεριακή γλώσσα αντικαταστάθηκε από την Ακκαδική γλώσσα⁴ η οποία άνηκε σε έναν λαό που αναμείχθηκε με τους Σουμερίους. Η Ακκαδική Γλώσσα θεωρείται σήμερα εξαφανισμένη. Στην Ακκαδική Γλώσσα είναι γραμμένο και το σημαντικότερο κείμενο της εποχής «ο Κώδικας του Χαμουραπί⁵». Παρόλη την επικράτηση της Ακκαδικής Γλώσσας, η Σουμεριακή Γλώσσα φαίνεται ότι διατηρήθηκε κυρίως από τους λόγιους για λογοτεχνικούς και θρησκευτικούς σκοπούς.

³ Parret, H. (Ed.). (1976). *History of linguistic thought and contemporary linguistics*. Walter de Gruyter.

⁴ Campbell, L. (2001). The history of linguistics. *The handbook of linguistics*, 81-104. Σελ. 82

⁵ Περιείχε νομοθεσία η οποία μεταξύ άλλων καθόριζε και τις ποινές για διάφορα αδικήματα. Ήταν γραμμένος σε μια στήλη η οποία ήταν εκτεθειμένη δημόσια για την βλέπουν και να τη διαβάσουν όλοι.

Μια άλλη σπουδαία γλωσσική παράδοση κατά την αρχαιότητα υπήρχε στην Ινδία με τη σανσκριτική γλώσσα η οποία είναι η Αρχαία Ινδική Γλώσσα η οποία διδάσκεται ακόμη και σήμερα στην Ινδία ως δεύτερη γλώσσα. Η Σανσκριτική Γλώσσα την περίοδο εκείνη στην Ινδία ήταν η επίσημη γλώσσα την οποία μιλούσαν οι μορφωμένοι και τη χρησιμοποιούσαν στα έργα τους οι λόγιοι. Στην Σανσκριτική ανήκει και η αρχαιότερη γραμματική η οποία γράφτηκε από τον λόγιο Πανίνι και περιείχε κανόνες για την ορθή χρήση της γλώσσας στον προφορικό και γραπτό λόγο. Επίσης άλλο αξιοσημείωτο έργο είναι το Vedas (1200 π.Χ.) το οποίο αποτελεί ένα από τα αρχαιότερα θρησκευτικά κείμενα⁶.

Πολύ σημαντική θέση στην ιστορία της γλώσσας έχει και η Αρχαία Ελληνική Γλώσσα στην οποία σώζονται ακόμη μέχρι σήμερα πολλά και σπουδαία λογοτεχνικά, φιλοσοφικά, ιστορικά, μαθηματικά, φιλολογικά, θρησκευτικά και άλλα γλωσσικά έργα τα οποία μας επιτρέπουν να διακρίνουμε την εξέλιξη της γλώσσας με την πάροδο του χρόνου. Η Αρχαία Ελληνική ομιλούνταν την Αρχαϊκή, την Κλασική και την Ελληνιστική Περίοδο. Πολύ σημαντικό θεωρείται το Ελληνικό Αλφάβητο το οποίο προέρχεται από το Φοινικικό Αλφάβητο με την προσθήκη φωνηέντων. Επίσης στο Ελληνικό Αλφάβητο βασίστηκε και η δημιουργία του Λατινικού Αλφάβητου και άλλων πολλών αλφάβητων⁷. Επιπλέον οι Αρχαίοι Έλληνες έγραφαν και τους αριθμούς με τη βοήθεια των γραμμάτων της αλφαβήτου.

Μνημειώδη έργα στα Αρχαία Ελληνικά θεωρούνται τα έπη του Ομήρου (850 π.Χ.), τα φιλοσοφικά κείμενα με εξέχοντα θέση εκείνα του Πλάτωνα, του Αριστοτέλη και των Στωικών Φιλοσόφων κ.α.. Σημαντικά ζητήματα που προκύπτουν μέσα από διάφορα φιλολογικά και φιλοσοφικά έργα της Αρχαίας Ελληνικής είναι η προέλευση της γλώσσας και τα μέρη του λόγου καθώς επίσης και η σχέση της γλώσσας με την ανθρώπινη σκέψη⁸. Όσον αφορά την προέλευση της γλώσσας κάποιοι φιλόσοφοι υποστήριζαν ότι είναι φυσική και κάποιοι άλλοι ότι είναι συμβατική ενώ υπήρχαν και περιπτώσεις αρχαίων λόγιων που υποστήριζαν μια μέση οδό σύμφωνα με την οποία τα ονόματα των πραγμάτων προήλθαν με φυσικό τρόπο αλλά υπέστησαν κάποια

⁶ Campbell, L. (2001). The history of linguistics. *The handbook of linguistics*, 81-104. Σελ. 82.

⁷ <http://medusa.libver.gr/jspui/bitstream/123456789/3786/6/IstoriaGrafis.pdf>.

⁸ Campbell, L. (2001). The history of linguistics. *The handbook of linguistics*, 81-104. Σελ. 82.

επεξεργασία. Ο Πλάτωνας στο έργο του ασχολήθηκε με την ετυμολογία⁹ και εκτός άλλων προσπάθησε να καθορίσει αν τα ονόματα των πραγμάτων έχουν ανθρώπινη ή Θεϊκή προέλευση και να ορίσει κάποια μέρη του λόγου τα οποία ισοδυναμούν με τα ουσιαστικά και τα ρήματα. Επιπλέον οι Στωικοί Φιλόσοφοι υποστηρίζουν ότι η γλώσσα ανήκει στη φύση.

Ο Αριστοτέλης στο έργο του μελετάει ποια είναι η φύση της γλώσσας και θεωρεί ότι οι προτάσεις είναι της μορφής υποκείμενο/κατηγορημα και ότι διακρίνονται σε τρεις κατηγορίες: ατομικές, καθολικές και μερικές. Ατομική είναι μια πρόταση, αν το υποκείμενο είναι ένα πρόσωπο. Οι γενικές προτάσεις έχουν ως υποκείμενο το όνομα ενός είδους, και οι καθολικές σύμφωνα με τον Αριστοτέλη, αναφέρονται στο σύνολο των ατόμων του είδους¹⁰.

Κατά τη Ρωμαϊκή Περίοδο πολύ σημαντικό ρόλο έπαιξε η Λατινική Γλώσσα και Γραμματεία η οποία είχε σαφή επιρροή από την αντίστοιχη Αρχαία Ελληνική. Μάλιστα οι ίδιοι οι Ρωμαίοι την θεωρούσαν ως μια παρεφθαρμένη μορφή της Αρχαίας Ελληνικής Γλώσσας¹¹. Πολύ σημαντικό έργο αποτελούν οι 18 τόμοι της γραμματικής του Λατίνου λόγιου Πρισκιανού που ασχολούνται με γραμματικούς κανόνες και τη σύνταξη της Λατινικής Γλώσσας. Επίσης πολύ σημαντικό έργο θεωρείται η λατινική γραμματική *Ars Grammatica* του Αίλιου Δονάτου (4^{ος} αιώνας)

1.1.1.2. Μεσαίωνας

Η ευρωπαϊκή πραγματεία την περίοδο αυτή ήταν έντονα επηρεασμένη από την Αρχαία Ελλάδα αλλά και από το έργο του Πρισκιανού και του Αίλιου Δονάτου. Τον 13^ο αιώνα μια ομάδα 30 λόγιων γνωστοί με την ονομασία *Modistae* ή "*Speculative Grammarians*" επηρεασμένοι από το έργο του Αριστοτέλη εισήγαγαν την έννοια της καθολικής γραμματικής, δηλαδή της μιας κοινής για όλες τις γλώσσες γραμματικής η οποία ξαναήρθε στο προσκήνιο από τον Τσόμσκι τη δεκαετία του 60.

⁹ Magoulas, G., Magoula, E., Μαγουλάς, Γ., & Μαγουλά, Ε. (2017). Θέματα Συγχρονικής και Διαχρονικής Γλωσσολογίας. Σελ. 18.

¹⁰ Αναπολιτάνος, Δ., Γαβαλάς, Δ., Δέμης, Α., Δημητρακόπουλος, Κ., Καρασμάνης, Β. (1991). Λογική Θεωρία και Πρακτική. Αθήνα: Διόφαντος.

¹¹ Magoulas, G., Magoula, E., Μαγουλάς, Γ., & Μαγουλά, Ε. (2017). Θέματα Συγχρονικής και Διαχρονικής Γλωσσολογίας. Σελ. 18.

Η Βυζαντινή Γραμματεία υπήρξε έντονα επηρεασμένη από την Αρχαία Ελληνική Γραμματεία και μάλιστα μεγάλη ήταν η χρήση της αττικής διαλέκτου αλλά και της αττικίζουσας η οποία αποτελούσε μια προσπάθεια μίμησης της. Επίσης οι Βυζαντινοί συνέταξαν τους κανόνες οι οποίοι ήταν πίνακες που περιείχαν ονοματικούς και ρηματικούς κλητικούς κανόνες¹². Τον 10^ο αιώνα συναντάμε δημοτικά τραγούδια τα οποία είναι γραμμένα στη δημώδη λαϊκή γλώσσα. Επίσης την περίοδο αυτή παρατηρείται μια σταδιακή εξέλιξη του προφορικού λόγου ο οποίος πλησιάζει το σημερινό σε φωνολογικό επίπεδο. Αλλαγές υπάρχουν και στο μορφοσυντακτικό επίπεδο όσον αφορά την κλιτική δομή των ονομάτων όπου συναντάμε δίπτωτα και τρίπτωτα ουσιαστικά.

1.1.1.3. Αναγέννηση

Κατά την περίοδο της Αναγέννησης τέθηκαν οι βάσεις για την τελική διαμόρφωση των γλωσσών της Ευρώπης στη σημερινή τους μορφή. Όσον αφορά όμως τη γραμματική δεν υπήρξε κάποια σημαντική πρόοδος αλλά περισσότερο επιρροή από την όψιμη Λατινική Γραμματική του Πρισκιανού και του Αίλιου Δονάτου. Επιπλέον σε διάφορες ευρωπαϊκές γλώσσες υπήρξαν μικρές αλλαγές στη γραμματική κυρίως όσον αφορά τα μέρη του λόγου

1.1.1.4. Νεότερη Γλωσσολογία

Οι γραμματικές που συντάχτηκαν το 17^ο και 18^ο αιώνα ακολούθησαν τα πρότυπα των αντίστοιχων γραμματικών του μεσαίωνα, χωρίς να υπάρχει κάποια καινοτομία στο περιεχόμενό τους.

Έως και τα τέλη του 19^{ου} αιώνα πραγματοποιείται συγγραφή γραμματικής για τους λαούς των νέων χωρών που έχουν ανακαλυφθεί. Η συγγραφή των νέων αυτών γραμματικών πραγματοποιείται κυρίως από τους ιεραπόστολους και είναι έντονα επηρεασμένη από τη λατινική γραμματική παράδοση.

¹² Magoulas, G., Magoula, E., Μαγουλάς, Γ., & Μαγουλά, Ε. (2017). Θέματα Συγχρονικής και Διαχρονικής Γλωσσολογίας. Σελ. 33

Σταθμό στη γλωσσολογία αποτελεί το έργο του Ferdinand Saussure όπου βάσει της θεωρίας του δημιουργήθηκε το κίνημα του στρουκτουρισμού ή δομισμού. Επηρεασμένες από το δομισμό είναι οι διάφορες γλωσσολογικές σχολές δημιουργήθηκαν μέχρι το 1950. Πολύ σημαντική υπήρξε και συμβολή του Noam Chomsky από τα μέσα του 20^{ου} αιώνα και μετά, με τη θεωρία της Καθολικής Γραμματικής.¹³

1.1.2. Οι υποκλάδοι της Γλωσσολογίας

Προκειμένου να μπορέσουμε να αναλύσουμε τα κυριότερα χαρακτηριστικά της γλώσσας είναι απαραίτητο να διακρίνουμε τα επίπεδα γλωσσικής ανάλυσης. Τα επίπεδα αυτά είναι : το φωνητικό, το φωνολογικό, το μορφολογικό, το συντακτικό, το σημασιολογικό και το πραγματολογικό. Παρόλο που τα επίπεδα της γλωσσολογικής ανάλυσης αντλούν στοιχεία το ένα από το άλλο, εξετάζονται χωριστά και απαιτούν τις δικές τους αναλύσεις και υπολογισμούς. Γι' αυτό το λόγο το κάθε ένα από αυτά τα επίπεδα αντιστοιχεί σε ξεχωριστό υποκλάδο της γλωσσολογίας και έτσι λοιπόν υπάρχουν:

- Η φωνητική (phonetics) που αντιστοιχεί στο φωνητικό επίπεδο
- Η φωνολογία (phonology) που αντιστοιχεί στο φωνολογικό επίπεδο
- Η μορφολογία (morphology) που σχετίζεται με το μορφολογικό επίπεδο
- Η σύνταξη (syntax) που σχετίζεται με το μορφολογικό επίπεδο
- Η σημασιολογία (semantics) που σχετίζεται με το συντακτικό επίπεδο
- Η πραγματολογία (pragmatics) που σχετίζεται με το πραγματολογικό επίπεδο

Η γλωσσική επιστήμη σε πολλές περιπτώσεις δανείζεται στοιχεία από διάφορες ανθρωπιστικές και θετικές επιστήμες όπως η ιστορία, η φιλοσοφία, η κοινωνιολογία, αλλά και τα μαθηματικά, η τεχνολογία και η πληροφορική¹⁴ κ.α. Η γλωσσολογία

¹³ Magoulas, G., Magoula, E., Μαγουλάς, Γ., & Μαγουλά, Ε. (2017). Θέματα Συγχρονικής και Διαχρονικής Γλωσσολογίας.

¹⁴ De Saussure, F. (1979). *Μαθήματα Γενικής Γλωσσολογίας*. Αθήνα: Εκδόσεις Παπαζήση.

άρχισε να θεωρείται αυτόνομη επιστήμη μόλις στις αρχές περίπου του 20^{ου} αιώνα ¹⁵. Από την περίοδο αυτή και μετά δημιουργήθηκαν πολλοί υποκλάδοι της γλωσσολογίας οι οποίοι συνδυάζουν τη γλωσσολογία με άλλους επιστημονικούς τομείς.

Έτσι λοιπόν οι βασικότεροι επιστημονικοί κλάδοι που έχουν δημιουργηθεί μέσα από αυτό το συσχετισμό με διάφορες κοινωνικές και θετικές επιστήμες είναι: Η Ψυχολογολογία (Psycholinguistics), η Φιλοσοφία της Γλωσσολογίας (Philosophy of Language), η Νευρογλωσσολογία (Neurolinguistics), η Μαθηματική Γλωσσολογία (Mathematical Linguistics), οι Γλωσσικές Θεωρίες (Linguistic Theories), η Γλωσσολογία της Λογοτεχνίας (Linguistics and Literature), η Λεξικογραφία (Lexicography), η Γλωσσική Τεκμηρίωση (Language Documentation), η Απόκτηση της Γλώσσας (Language Acquisition), η Ιστορική Γλωσσολογία (Historical Linguistics), η Γνωσιακή Επιστήμη ή Γνωσιακή Γλωσσολογία (Cognitive Science), η Γενική Γλωσσολογία (General Linguistics), η Δικανική Γλωσσολογία (Forensic Linguistics), η Ανάλυση του Λόγου (Discourse Analysis), η Υπολογιστική Γλωσσολογία (Computational Linguistics), η Εφαρμοσμένη Γλωσσολογία (Applied Linguistics), η Ανθρωπολογική Γλωσσολογία (Anthropological Linguistics), τα συστήματα Γραφής (Writing Systems), η Τυπολογία (Typology), η Διερμηνεία και η Μετάφραση (Translating and Interpreting), τα Σώματα Κειμένου (Text Corpus) και η Κοινωνιογλωσσολογία (Sociolinguistics).

1.2. Η ανάκτηση της πληροφορίας

Σε γενικές γραμμές η ανάκτηση της πληροφορίας αφορά την αναζήτηση πληροφορίας μέσα σε έγγραφα ή ακόμη και την αναζήτηση ενός ολόκληρου εγγράφου. Το εύρος της είναι πολύ μεγάλο καθώς περιλαμβάνει την αναζήτηση δεδομένων σε ψηφιακές βιβλιοθήκες, πληροφοριών στο διαδίκτυο, πληροφοριών κάθε είδους μέσα σε έγγραφα, φίλτρα ταξινόμησης, οπτικοποίηση δεδομένων κ.α. .

Στη συνήθη της μορφή η διαδικασία της ανάκτησης ξεκινάει όταν ο χρήστης συντάξει ένα ερώτημα σε μια μηχανή αναζήτησης. Το σύστημα συγκρίνει το

¹⁵ Ξυδόπουλος, Γ. Ι. (2016). Η γλωσσολογία σήμερα στόχοι, αντιλήψεις εφαρμογές και επαγγελματικές εφαρμογές. *Επιστημονική Επετηρίδα Τμήματος Φιλολογίας*, 35.

ερώτημα το οποίο πραγματοποιεί ο χρήστης με τα έγγραφα μιας συλλογής και επιστρέφει μια ταξινομημένη λίστα που περιέχει τα αποτελέσματα της αναζήτησης.

Η ανάκτηση της πληροφορίας αφορά κυρίως αδόμητου τύπου δεδομένα τα οποία υπάρχουν διάσπαρτα : στο διαδίκτυο και σε σώματα κειμένων, συλλογές εγγράφων, βιβλιοθήκες και βάσεις δεδομένων. Επίσης εκτός από την ανάκτηση αδόμητων δεδομένων, η ανάκτηση πληροφορίας ασχολείται και με την κατηγοριοποίηση και συσταδοποίηση (clustering) σχετικών σε περιεχόμενο δεδομένων σε ομάδες.

Σημαντικός στόχος ενός συστήματος ανάκτησης πληροφορίας είναι η ανάκτηση των πιο σχετικών εγγράφων αναφορικά με το ερώτημα το οποίο έθεσε ο χρήστης. Το πόσο αυτό επιτυγχάνεται καθορίζει την αποτελεσματικότητα ενός συστήματος ανάκτησης της πληροφορίας. Τα μεγέθη που αξιολογούν την επιτυχία και αποτελεσματικότητα ενός τέτοιου συστήματος είναι η ακρίβεια (precision) και η ανάκληση (recall). Ακρίβεια (precision) είναι το ποσοστό των ανακτημένων εγγράφων που είναι συναφή με το ερώτημα που τέθηκε και ισούται με:

- $$\text{Precision} = \frac{\text{Σύνολο σχετικών εγγράφων που επιστρέφονται}}{\text{Σύνολο εγγράφων που επιστρέφονται}}$$

και ανάκληση είναι το ποσοστό των συναφών εγγράφων που ανακτήθηκαν και ισούται με:

- $$\text{Recall} = \frac{\text{Σύνολο σχετικών εγγράφων που επιστρέφονται}}{\text{Σύνολο σχετικών εγγράφων}}$$

Επειδή το διαδίκτυο είναι ένας χώρος με πολύ μεγάλο εύρος, οι εφαρμογές της ανάκτησης της πληροφορίας μπορεί να αφορούν διάφορες θεματικές περιοχές. Έτσι λοιπόν υπάρχει η αναζήτηση σε blog, το φιλτράρισμα πληροφοριών, οι ψηφιακές βιβλιοθήκες, η ανάκτηση φωνής, η ανάκτηση νομικών πληροφοριών, κοινωνική αναζήτηση, κάθετη αναζήτηση, ανάκτηση γεωγραφικών πληροφοριών, ανίχνευση spam μηνυμάτων σε λογαριασμό ηλεκτρονικού ταχυδρομείου, εξόρυξη δεδομένων κ.α. .

1.2.1. Οι μηχανές αναζήτησης

Ενώ παλαιότερα η ανάκτηση της πληροφορίας αφορούσε κυρίως την επιστήμη της βιβλιοθηκονομίας, με την εφεύρεση και εισαγωγή του παγκόσμιου ιστού (World Wide Web) το 1989 σταδιακά άρχισε να αφορά και την πληροφορική. Ένας από τους λόγους που επιτεύχθηκε αυτό είναι το ότι το World Wide Web παρέχει τη δυνατότητα εύκολης πρόσβασης σε μεγάλο όγκο πολυμεσικών ψηφιακών πληροφοριών μέσω των μηχανών αναζήτησης.

Οι μηχανές αναζήτησης είναι γραφικές διεπαφές συστημάτων ανάκτησης πληροφορίας. Αποτελούνται από μεγάλες βάσεις δεδομένων που περιλαμβάνουν δεδομένα και πληροφορίες από το διαδίκτυο. Στην ουσία οι μηχανές αναζήτησης δημιουργούν ένα δικό τους ευρετήριο και από εκεί αναζητούν τις πληροφορίες που έχει ζητήσει ο χρήστης. Οι μηχανές αναζήτησης έχουν αλλάξει τον τρόπο με τον οποίο γίνεται η αναζήτηση καθώς πριν την ευρετηρίαση προηγείται η διαδικασία του crawling κατά την οποία συλλέγονται έγγραφα από τον ιστό και αποθηκεύονται τα αντίγραφα τους σε ειδικούς χώρους¹⁶.

Με τη χρήση των μηχανών αναζήτησης έχει γίνει εφικτή η ανάκτηση μεγάλου πλήθους εγγράφων και γι' αυτό το λόγο το διαδίκτυο έχει γίνει πλέον η μεγαλύτερη πηγή πληροφοριών στην ιστορία. Καθώς με την πάροδο των χρόνων ο παγκόσμιος ιστός επεκτεινόταν, δημιουργήθηκε έντονα η ανάγκη πιο εξειδικευμένων και πολύπλοκων μηχανών αναζήτησης και έτσι σήμερα γίνεται λόγος για social search engines, enterprise search, vertical search engines κ.α.

1.2.2. Από τα πρώτα συστήματα ανάκτησης έως τον σημασιολογικό ιστό

Το 1945 στο έργο του Βαννεβαρ Μπους «Πώς μπορούμε να σκεφτούμε » , εκφράζεται η σκέψη δημιουργίας μιας μηχανής μέσω της οποίας η γνώση θα γινόταν περισσότερο προσβάσιμη στους ανθρώπους. Τη δεκαετία του 60 αναπτύχθηκαν τα πρώτα συστήματα ανάκτησης πληροφορίας για την αναζήτηση σε μικρές συλλογές κειμένων.

¹⁶ Baeza -Yates, R. Ribeiro – Neto , B.(2015). *Ανάκτηση Πληροφορίας Οι έννοιες και η τεχνολογία πίσω από την αναζήτηση*. Θεσσαλονίκη: Εκδόσεις Τζιόλα.

Τα πρώτα συστήματα ανάκτησης πληροφοριών χρησιμοποιήθηκαν από βιβλιοθήκες για αναζήτηση βιβλίων με κριτήρια όπως ο τίτλος, το όνομα του συγγραφέα και το όνομα του εκδοτικού οίκου. Αργότερα με τη πρόοδο που σημειώθηκε στην πληροφορική και την ανάκτηση της πληροφορίας προστέθηκε η δυνατότητα πραγματοποίησης πιο πολύπλοκων ερωτημάτων, χρησιμοποιώντας λογικούς τελεστές όπως π.χ. AND, OR, NOT.

Τη δεκαετία του 80 σημειώθηκε μεγάλη πρόοδος καθώς αναπτύχθηκαν μεγάλα συστήματα ανάκτησης για την εξυπηρέτηση των πληροφοριακών αναγκών εταιρειών. Τη δεκαετία του 90 τα πράγματα άλλαξαν εντελώς με την αναζήτηση πληροφοριών μέσω της υπηρεσίας FTP αλλά και με την καθιέρωση του παγκόσμιου ιστού (World Wide Web). Τη δεκαετία του 2000 λόγω της αύξησης της ταχύτητας του διαδικτύου και της υπολογιστικής ισχύος των υπολογιστών αυξήθηκε πολύ το μέγεθος των πληροφοριών που διακινούνταν μέσω του διαδικτύου. Σταθμός υπήρξε και η δημιουργία της Google τα τέλη της δεκαετίας του 90 και η ταχεία ανάπτυξη της.

Νέα διάσταση στην ανάκτηση της πληροφορίας προσφέρει ο Σημασιολογικός Ιστός. Ο δημιουργός του παγκόσμιου ιστού Tim Berners-Lee το 1999 εισήγαγε για πρώτη φορά την έννοια του σημασιολογικού ιστού ο οποίος αποτελεί την εξέλιξη του παγκόσμιου ιστού. Το βασικό του χαρακτηριστικό είναι ότι κάνει χρήση νοημόνων τεχνικών για την αξιοποίηση του περιεχομένου του προκειμένου να αποτελεί γνώση και όχι απλώς πληροφορία, αξιοποιώντας με αυτόν τον τρόπο το μεγάλο όγκο των δεδομένων που υπάρχει στο διαδίκτυο .

Στο σημασιολογικό ιστό οι πληροφορίες αντικαθίστανται από:

- τις οντολογίες οι οποίες είναι ιεραρχικές δομές αναπαράστασης της γνώσης,
- τα μεταδεδομένα τα οποία είναι δεδομένα που περιγράφουν άλλα δεδομένα
- και τους νοήμονες πράκτορες οι οποίοι είναι λογισμικά τα οποία σαρώνουν το διαδίκτυο και συμβάλουν στην εκτέλεση πολύπλοκων ενεργειών χρησιμοποιώντας γνώση την οποία απέκτησαν.

Κεφάλαιο 2

2.1. Οι Υποκλάδοι της Γλωσσολογίας

Η σχέση της γλωσσολογίας με την ανάκτηση της πληροφορίας είναι ουσιαστικά η σχέση μεταξύ του γραπτού και προφορικού λόγου με την ανάκτηση της πληροφορίας.

Ο ρόλος του γραπτού λόγου είναι πολύ μεγάλης σημασίας στην ανάκτηση της πληροφορίας καθώς το κυριότερο προϊόν της είναι κειμενικής μορφής, ενώ ένα κείμενο διαθέτει τη δική του δομή, τη δική του γλώσσα γραφής, τη δική του σύνταξη, το δικό του ύφος και το δικό του λεξιλόγιο. Επίσης στην αναζήτηση της πληροφορίας προκειμένου να μας επιστραφούν τα σωστά αποτελέσματα πρέπει η διατύπωση να γίνεται με τον κατάλληλο τρόπο που περιλαμβάνει τις κατάλληλες λέξεις και εκφράσεις διατυπωμένες με το σωστό τρόπο.

Μεγάλο ρόλο πλέον παίζει και ο προφορικός λόγος καθώς τα τελευταία χρόνια έχει αναπτυχθεί πάρα πολύ η ανάκτηση και η χρήση της φωνής για την αναζήτηση πληροφοριών και με αυτόν τον τρόπο έχουν δημιουργηθεί πολλές εφαρμογές που διευκολύνουν τον άνθρωπο στην καθημερινότητα του.

Από το 1989 έως σήμερα ο αριθμός των εγγράφων και των πληροφοριών που διακινούνται μέσω διαδικτύου, αυξάνεται συνεχώς. Έτσι λοιπόν υπήρξε η ανάγκη για γλωσσολογική ανάλυση και για τεχνικές που θα έκανα εφικτή την υλοποίηση εφαρμογών επεξεργασίας της φυσικής γλώσσας.

Διάφορες εφαρμογές της ανάκτησης της πληροφορίας είναι επηρεασμένες σε μεγάλο βαθμό από υποκλάδους της γλωσσολογίας και αρκετές φορές είναι απαραίτητη η μελέτη των υποκλάδων αυτών για την υλοποίηση διαφόρων εφαρμογών.

Υπάρχουν πολλοί υποκλάδοι της Γλωσσολογίας οι οποίοι έχουν προταθεί από επιστήμονες και αποτελούν ξεχωριστό πεδίο. Κάποιοι πολύ σημαντικοί από αυτούς αναλύονται και παρουσιάζονται παρακάτω.

2.2. Η Φωνητική

Η φωνητική είναι ο κλάδος της γλωσσολογίας που ασχολείται με την περιγραφή της ομιλίας και της φωνής και γενικότερα των ήχων που παράγει το ανθρώπινο φωνητικό σύστημα. Οι ελάχιστοι γλωσσικοί ήχοι που παράγονται από τον άνθρωπο ονομάζονται φθόγγοι. Οι φθόγγοι δεν έχουν πάντα νόημα και μπορεί να αποτελούνται μόνο από απλούς ήχους. Δύο ή περισσότεροι φθόγγοι αν συνδυαστούν μαζί με τον κατάλληλο τρόπο σχηματίζουν λέξεις οι οποίες έχουν σημασία για τον άνθρωπο. Κάθε ξένη γλώσσα έχει διαφορετικές λέξεις οι οποίες για κάποιαν άλλη ομιλούμενη γλώσσα μπορεί να μην έχουν σημασία. Οι φθόγγοι αναλύονται σε σύμφωνα και φωνήεντα.

Η φωνητική εξετάζει εκτός από την παραγωγή (Αρθρωτική Φωνητική), την πρόσληψη (Ακουστική Φωνητική) και κατανόηση (Αντιληπτική Φωνητική) των φθόγγων¹⁷.

Η μελέτη του τρόπου παραγωγής της ανθρώπινης φωνής είναι χρήσιμη για διάφορους λόγους όπως η μελέτη της προφοράς διαφόρων ξένων γλωσσών, η θεραπεία γλωσσικών προβλημάτων μέσω της λογοθεραπείας και η ανάπτυξη μεθόδων παραγωγής και αναγνώρισης της ομιλίας από ηλεκτρονικούς υπολογιστές¹⁸.

2.2.1. Το ανθρώπινο φωνητικό σύστημα

Με τον όρο ανθρώπινο φωνητικό σύστημα εννοούμε όλα τα όργανα του ανθρώπινου σώματος που συμμετέχουν στην παραγωγή των ήχων και της φωνής. Αυτά είναι:

- Οι πνεύμονες καθώς οι περισσότεροι από τους φθόγγους παράγονται με την εκπνοή αέρα.
- Ο λάρυγγας ο οποίος είναι ένας σωλήνας από μυϊκό ιστό και χόνδρους μέσω του οποίου ο αέρας οδηγείται στους πνεύμονες.
- Οι φωνητικές χορδές οι οποίες βρίσκονται στο λάρυγγα και ανοίγουν όταν αναπνέουμε και κλείνουν όταν τρώμε προκειμένου να εμποδίσουν τη τροφή να εισέλθει στους πνεύμονες. Για την παραγωγή φωνής πρέπει αέρας από τους

¹⁷ <https://el.wikipedia.org/wiki/Φωνητική>

¹⁸ Ladefoged, P. (2007). Εισαγωγή στη φωνητική. Μ. Μπαλτατζάνη (μτφ.). Αθήνα: Εκδόσεις Πατάκη.

πνεύμονες να περάσει από τις φωνητικές χορδές οι οποίες έρχονται κοντά. Όλα τα φωνήεντα και αρκετά σύμφωνα παράγονται με τη βοήθεια των φωνητικών χορδών.

- Το φωνητικό κανάλι το οποίο αποτελείται από τη φαρυγγική, τη στοματική και τη ρινική κοιλότητα.
- Ο φάρυγγας ο οποίος είναι ένας ίνομυώδης σωλήνας που ξεκινάει από το λάρυγγα και καταλήγει στο πίσω μέρος της στοματικής και ρινικής κοιλότητας.
- Η στοματική κοιλότητα η οποία αποτελείται από διάφορα μέρη τα οποία είναι υπεύθυνα για την άρθρωση δηλαδή την προφορά των φθόγγων και των συνδυασμών τους και γι' αυτό ονομάζονται αρθρωτές. Τα μέρη αυτά είναι: η φατνία, ο ουρανίσκος, η υπερώα, η σταφυλή, τα δόντια, τα χείλια, η σιαγόνα και η γλώσσα.
- Τέλος η ρινική κοιλότητα καθώς κάποια σύμφωνα [m,n] όταν παράγονται πραγματοποιείται εκροή αέρα από τη μύτη.¹⁹

Επειδή η ανατομία και το μέγεθος των φωνητικών οργάνων διαφέρει από άνθρωπο σε άνθρωπο καθώς επίσης και για λόγους όπως η ηλικία και το φύλο, κάθε άνθρωπος έχει τη δική του φωνή που είναι μοναδική και διαφέρει από τις υπόλοιπες όσον αφορά το ύψος και τη βραχνάδα.

2.2.2. Διεθνές φωνητικό αλφάβητο

Για την καταγραφή και αναπαράσταση των φθόγγων έχει υιοθετηθεί το Διεθνές Φωνητικό Αλφάβητο το οποίο είναι ίδιο για όλες τις γλώσσες και είναι ανεξάρτητο της ορθογραφίας μιας λέξης. Χρησιμοποιεί κυρίως λατινικούς χαρακτήρες αλλά και κάποιους ελληνικούς και δεν υπάρχει διάκριση μεταξύ κεφαλαίων και πεζών. Η ερμηνεία του κάθε χαρακτήρα ενδέχεται να είναι διαφορετική από γλώσσα σε γλώσσα και με αυτόν τον τρόπο το Διεθνές Φωνητικό Αλφάβητο πετυχαίνει να

¹⁹ Μαγουλάς, Γ., Μαγουλά, Ε., 2017. *Θέματα Συγχρονικής και Διαχρονικής Γλωσσολογίας*. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/6509>

αναπαριστά όλους τους δυνατούς συνδυασμούς φώνησης και άρθρωσης για όλες τις γλώσσες του κόσμου.

Στην φωνητική μετεγγραφή κάθε σύμβολο αντιστοιχεί σ' ένα ήχο. Για την φωνητική μετεγγραφή τις ελληνικής γλώσσας υπάρχουν διαθέσιμα 5 φωνήεντα i, e, a, o, u και 27 σύμφωνα b (μπάλα), d (νταλικά), ɸ (αγκύλωση), g (γκάλοπ), p (παζάρι), t (τασάκι), c (κύριος), k (καράβι), v (βάζο), ð (δάκος), z (ζαρώνω), j (γυρεύω), γ (γαργάρα), f (φασαρία), θ (θεατής), s (σαλάτα), ʃ (χελιδόνι), x (χαρά), m (μάζα), n (νάμα), ɲ (νιότη), dz (τζάμι), ts (τσατσάρα), r (ράμμα), l (λάμπα), ʎ (λιανός), ŋ (άγχος).

2.2.3. Κατηγορίες φωνητικής

2.2.3.1. Αρθρωτική φωνητική

Άρθρωση είναι ο τρόπος με τον οποίο προφέρουμε τους διάφορους φθόγγους οι οποίοι παράγονται από την κίνηση των μερών από το λάρυγγα έως το στόμα και τη μύτη μέσω των οποίων αποβάλλεται ο αέρας. Για την άρθρωση των φωνηέντων πραγματοποιείται ομαλή ροή αέρα από τη φωνητική οδό, αντίθετα η άρθρωση των συμφώνων προκαλείται από εμπόδια που δημιουργούνται από τους αρθρωτές όπως τα χείλη, τα δόντια, τα φαρνία, ο ουρανίσκος κ.α. .

2.2.3.1.1. Άρθρωση συμφώνων

Ο αέρας αυτός συναντά κάποια εμπόδια τα οποία είναι διαφορετικά ανάλογα με το φθόγγο που θέλουμε να προφέρουμε. Ανάλογα τα μέρη της φωνητικής οδού που χρησιμοποιούνται για την παραγωγή ενός φθόγγου τα σύμφωνα της ελληνικής γλώσσας τα διακρίνουμε σε:

- Διχειλικά στα οποία χρησιμοποιούνται το άνω και το κάτω χείλος: [p][b][m]
- Χειλοδοντικά στα οποία χρησιμοποιούνται το κάτω χείλος και τα πάνω δόντια: [f][v]
- Μεσοδοντικά στα οποία χρησιμοποιείται η άκρη της γλώσσας ανάμεσα στα άνω και κάτω δόντια [θ][ð]

- Οδοντικά στα οποία χρησιμοποιείται η άκρη της γλώσσας στα άνω δόντια
- Φατνιακά στα οποία χρησιμοποιείται η άκρη της γλώσσας στη φατνία: [s][z][n][l][r].
- Ουρανικά στα οποία χρησιμοποιείται η ράχη της γλώσσας στον ουρανίσκο : [c][j][ɲ][ç][ʃ][ʎ]
- Υπερωικά στα οποία χρησιμοποιείται το πίσω μέρος της γλώσσας με την υπερώα: [k][x][ɣ][g][ŋ]

Επιπλέον οι φθόγγοι που παράγονται όταν οι φωνητικές πτυχές πάλλονται λέγονται ηχηροί ([m], [n],[b][v][d][z][l][r][j][g][ɣ][ð][ŋ] [ʃ][ʎ] ɲ) ενώ όταν ο αέρας περνάει ελεύθερα είναι άηχοι ([p][f][θ][t][s][k][x][c][ç]

Ως προς το χώρο αντήχησης τα σύμφωνα μπορούν να χωριστούν σε ρινικά: [m], [n],[ɲ] ,[ɲ] και στοματικά: όλα τα μη ρινικά²⁰

2.2.3.1.2. Άρθρωση φωνηέντων

Όσον αφορά την άρθρωση των φωνηέντων μπορούμε να διακρίνουμε τα φωνήεντα σε πρόσθια, οπίσθια και κεντρικά. Τα πρόσθια είναι αυτά τα οποία όταν προφέρονται η γλώσσα βρίσκεται κοντά στα δόντια. Τέτοια φωνήεντα είναι το ι και το ε. Τα οπίσθια φωνήεντα είναι αυτά τα οποία όταν προφέρονται, η γλώσσα βρίσκεται απομακρυσμένη αρκετά από τα δόντια τέτοια παραδείγματα αποτελούν το ο και το ου. Τέλος τα κεντρικά φωνήεντα είναι αυτά τα οποία βρίσκονται σε ενδιάμεση θέση όπως για παράδειγμα το α.

2.2.3.2. Ακουστική φωνητική

Η ακουστική φωνητική περιλαμβάνει τις φάσεις από την παραγωγή ήχου με τη βοήθεια των φωνητικών οργάνων του ανθρώπου, ως τη λήψη του ήχου από τα ακουστικά όργανα του ανθρώπου.

²⁰ Παυλίδου, Θ. Σ. (2008). *Επίπεδα γλωσσικής ανάλυσης*. Θεσσαλονίκη: Ίδρυμα Μανόλη Τριανταφυλλίδη.

Οι φθόγγοι που παράγει το φωνητικό σύστημα του ανθρώπου, μεταδίδονται μέσω ηχητικών κυμάτων. Αυτό σημαίνει ότι κάθε φθόγγος και κατά συνέπεια και κάθε λέξη που αρθρώνεται έχει διαφορετική συχνότητα δηλαδή διαφορετικό αριθμό κύκλων ανά δευτερόλεπτο και διαφορετικό ρυθμό επαναληπτικότητας των ηχητικών κυμάτων.

Κάθε φθόγγος έχει διαφορετική ακουστικότητα η οποία σχετίζεται με το πόσο έντονα ακούγεται ένας ήχος. Η ένταση ενός ήχου εξαρτάται από την ποσότητα του αέρα που αποβάλλουν οι πνεύμονες κατά την παραγωγή ενός φθόγγου. Αν η ποσότητα του αέρα είναι πολύ μεγάλη τότε και η ένταση είναι μεγάλη, αντίθετα αν η ποσότητα του αέρα είναι μικρή, τότε και η ένταση είναι χαμηλή.

Οι ήχοι διακρίνονται σε σύνθετους (αρμονικοί) και τους απλούς. Στους αρμονικούς ήχους οι ταλαντώσεις των ηχητικών κυμάτων έχουν ημιτονοειδή μορφή. Οι αρμονικοί ήχοι δημιουργούνται μέσω νέων κυμάτων που παράγονται από τους απλούς ήχους όταν προσκρούουν σε μια επιφάνεια.

2.2.3.3. Ακροατική φωνητική

Η ακροατική φωνητική έχει σχέση με το πώς αντιλαμβάνονται τα ακουστικά όργανα του ανθρώπου τον ήχο.

Σχετίζεται με την ακουστική φωνητική και με τα ηχητικά κύματα που παράγει το ανθρώπινο φωνητικό σύστημα. Για παράδειγμα ήχους με υψηλή συχνότητα τους αντιλαμβανόμαστε πιο οξείς ενώ ήχους με χαμηλή συχνότητα τους αντιλαμβανόμαστε πιο βαριούς. Επίσης όταν ο ήχος έχει σταθερή ή ομαλά μεταβαλλόμενη συχνότητα τότε χαρακτηρίζεται από καθαρότητα, διαφορετικά τα ακουστικά όργανα του ανθρώπου τον αντιλαμβάνονται με θόρυβο.

Τα ηχητικά κύματα τα οποία αντιστοιχούν στους ήχους που παράγει το φωνητικό σύστημα του ανθρώπου ανακλώνται και δημιουργούν νέα στάσιμα κύματα τα οποία αντιστοιχούν στους αρμονικούς ήχους οι οποίοι καθορίζουν τη χροιά και έκφραση του ήχου. Παρατηρείται ότι ίδιοι φθόγγοι που παράγονται από διαφορετικό άνθρωπο

έχουν διαφορετική χροιά λόγω της διαφορετικής ανατομίας των φωνητικών οργάνων του καθενός. Άρα η χροιά του κάθε φθόγγου δεν είναι πάντα ίδια.

2.2.4 Η χρησιμότητα της Φωνητικής στην Ανάκτηση της Πληροφορίας

Στη φωνητική ανάκτηση της πληροφορίας τα ερωτήματα των χρηστών γίνονται μέσω ανθρώπινης φωνής. Επίσης βασικές γνώσεις από τη φωνητική χρησιμοποιούνται προκειμένου να αναπτυχθούν συστήματα που δέχονται ως είσοδο την ανθρώπινη φωνή. Χαρακτηριστικό παράδειγμα αποτελούν τα συστήματα αναγνώρισης φωνής ή ομιλητή. Τα συστήματα αυτά δίνουν τη δυνατότητα στους χρήστες να αλληλεπιδρούν με κατάλληλο hardware και παρέχεται η δυνατότητα αναγνώρισης της ταυτότητας του χρήστη βάσει συγκεκριμένων χαρακτηριστικών που διαθέτει η φωνή του. Τα χαρακτηριστικά αυτά βασίζονται στη φυσιολογία και ανατομικά χαρακτηριστικά των οργάνων του φωνητικού συστήματος του ανθρώπου όπως το μέγεθος και το σχήμα. Επειδή αυτά διαφέρουν σε κάθε άνθρωπο, κατά συνέπεια και η φωνή του κάθε ανθρώπου είναι μοναδική.

Υπάρχουν δύο κατηγορίες τέτοιων συστημάτων. Στην πρώτη κατηγορία η ταυτοποίηση της φωνής γίνεται μέσω συγκεκριμένης φράσης. Η κατηγορία αυτή φωνητικής αναγνώρισης είναι η πιο συνηθισμένη και παρέχει τις καλύτερες επιδόσεις. Στη δεύτερη η φράση μέσω της οποίας γίνεται η ταυτοποίηση είναι τυχαία.

Η φωνητική αναγνώριση περιλαμβάνει τις διαδικασίες της ταυτοποίησης και της επαλήθευσης. Η ταυτοποίηση είναι ο προσδιορισμός της ταυτότητας του ατόμου και η επαλήθευση είναι η αποδοχή ή απόρριψη της ταυτότητας του χρήστη.

Υπάρχουν διάφορες εφαρμογές στις οποίες χρησιμοποιείται η φωνητική αναγνώριση όπως για έλεγχο πρόσβασης, πρόσβαση σε βάσεις δεδομένων²¹, εξ αποστάσεως πρόσβαση σε υπολογιστές, ξεκλείδωμα κινητού τηλεφώνου μέσω φωνητικού κωδικού κ.α. .

Η εγγραφή σ' ένα τέτοιο σύστημα γίνεται με τη βοήθεια κάποιου μικροφώνου και η φωνητική πληροφορία κατατημάται και αποθηκεύεται. Το σύστημα καταγράφει το

²¹ <https://www.sciencedirect.com/topics/computer-science/speaker-recognition>

φάσμα της φωνής του και μεγάλη σημασία δίνεται σε χαρακτηριστικά της φωνής όπως είναι ο τόνος, το ύψος, το χρώμα και η συχνότητα.

Η διαδικασία ταυτοποίησης σ' ένα τέτοιο σύστημα συνήθως ακολουθεί τα εξής βήματα:

- Αρχικά σκανάρεται η φωνή,
- στη συνέχεια αναλύεται και πραγματοποιείται εξαγωγή χαρακτηριστικών όπως η συχνότητα και ο τόνος της φωνής. Στο στάδιο αυτό πραγματοποιείται κατάτμηση του μηνύματος σε μικρότερα κομμάτια
- Πραγματοποιείται έλεγχος προκειμένου να διαπιστωθεί αν τα κατατμημένα τμήματα του μηνύματος ταιριάζουν με τα αντίστοιχα που είναι αποθηκευμένα και αφορούν το πρότυπο μήνυμα
- Ανάλογα αν υπάρχει αντιστοιχία μεταξύ του πρότυπου μηνύματος και του σκαναρισμένου το σύστημα δέχεται ή απορρίπτει το χρήστη.²²

2.3. Φωνολογία

Η φωνολογία ασχολείται με τον τρόπο με τον οποίο οι ήχοι λειτουργούν και περιγράφονται στα πλαίσια μιας φυσικής γλώσσας. Η φωνολογία μιας φυσικής γλώσσας διαφέρει από την αντίστοιχη μιας άλλης φυσικής γλώσσας. Μια και μόνο διαφοροποίηση σε κάποιο φθόγγο μπορεί να αλλάξει το νόημα μιας λέξης και να δημιουργήσει προβλήματα στη γλωσσική επικοινωνία. Π.χ. στη λέξη «δίνω» αν αλλάξω τον πρώτο φθόγγο με τον φθόγγο «π» προκύπτει η λέξη «πίνω» η οποία έχει εντελώς διαφορετικό νόημα.

Η βασική διαφορά της φωνολογίας με τη φωνητική είναι ότι η δεύτερη αφορά την παραγωγή ήχων και φθόγγων μέσω του ανθρώπινου φωνητικού συστήματος. Η φωνολογία προϋποθέτει τη φωνητική καθώς τα φωνήματα με τα οποία ασχολείται η φωνολογία αποτελούν στην ουσία μια γενικότερη κατηγοριοποίηση των ήχων που παράγει το ανθρώπινο φωνητικό σύστημα.

²² <http://www.biometric-solutions.com/speaker-recognition.html>

2.3.1. Τα Φωνήματα

Τα φωνήματα δεν μπορούν να χωριστούν σε μικρότερες μονάδες²³. Η διαφορά των φωνημάτων από τους φθόγγους είναι ότι οι δεύτεροι είναι οι ελάχιστοι ήχοι που μπορεί κάποιος να αρθρώσει, ενώ τα φωνήματα είναι στην ουσία ο κώδικας με τον οποίο αναπαριστά τους φθόγγους. Επιπλέον οι φθόγγοι ανήκουν στο φυσικό ή επιφανειακό επίπεδο μιας γλώσσας ενώ τα φωνήματα στο φωνηματικό ή αφηρημένο επίπεδο μιας γλώσσας²⁴. Στη γλωσσολογία το κάθε φώνημα συμβολίζεται με το αντίστοιχο σύμβολο εντός «/» π.χ. /ο/ , ενώ ο κάθε φθόγγος εντός «[]» π.χ. [ο].

Κάθε φώνημα διαθέτει συγκεκριμένα χαρακτηριστικά τα οποία σχετίζονται με τα μέρη του φωνητικού συστήματος τα οποία χρησιμοποιούνται για την παραγωγή τους. Ένας από τους ρόλους της φωνολογίας είναι ο εντοπισμός αυτών των χαρακτηριστικών ο οποίος διαφοροποιεί τα φωνήματα.

Τα φωνήματα διαφοροποιούν τις λέξεις και το νόημα τους. Αυτό ίσως είναι περισσότερο ορατό σε περιπτώσεις λέξεων που διαφοροποιούνται μόνο ως προς ένα φώνημα. Π.χ. ρόγα – ρίγα, με – μου, πως – πας. Στην ελληνική γλώσσα τα φωνήματα είναι 23 τα οποία σχετίζονται με φωνητικούς κανόνες. Από αυτά

- τα 5 είναι φωνήεντα : τα κλειστά (ι, ου), τα μέσα (ε, ο) και τα ανοιχτά (α)
- Τα 18 είναι σύμφωνα :τα κλειστά (π, μπ, ντ, τ, κ, γκ), τα τριβώμενα (φ, β, θ, δ, χ, γ), τα ρινικά (μ, ν), τα συριστικά (σ, ζ), το πλευρικό (λ) και το παλλόμενο (ρ).

Μια άλλη ταξινόμηση των συμφώνων είναι: σε διχειλικά (π, μπ, μ), χειλοδοντικά (φ, β), οδοντικά (τ, ντ, θ, δ, σ, ζ, λ), φατνιακά (ρ, ν) και ραχιαία (κ, χ, γκ, γ).²⁵ Επίσης κάποια φωνήματα θεωρούνται ηχηρά και κάποια άλλα άηχα.

2.3.2. Αλλόφωνα

Υπάρχουν περιπτώσεις φωνημάτων τα οποία αποτελούν εναλλακτικές μορφές του ίδιου ήχου. Οι περιπτώσεις αυτές ονομάζονται αλλόφωνα και ουσιαστικά αποτελούνται από φωνήματα που εκδηλώνονται με διαφορετικούς τρόπους²⁶.

²³ Παυλίδου, Θ. (2008). Επίπεδα γλωσσικής ανάλυσης. Αθήνα: Ινστιτούτο Νεοελληνικών Σπουδών (Ιδρυμα Μανόλη Τριανταφυλλίδη).

²⁴ Φιλιπτάκη, Ε. (1992). *Εισαγωγή στη θεωρητική γλωσσολογία*. Αθήνα: Νεφέλη

²⁵ <http://ebooks.edu.gr/modules/ebook/show.php/DSGYM-A112/621/4005,17967/>

²⁶ Φιλιπτάκη, Ε. (1992). *Εισαγωγή στη θεωρητική γλωσσολογία*. Αθήνα: Νεφέλη.

Στην ελληνική γλώσσα έχουμε κάποιες περιπτώσεις αλλόφωνων φωνημάτων. Η πρώτη περίπτωση είναι αυτή του /c/ (κύμα) και του /k/ (κότα) όπου το /c/ το συναντάμε μπροστά από τα φωνήεντα /i, e/ και το /k/ μπροστά από τα φωνήεντα /a, o, u/. Η δεύτερη περίπτωση είναι του /x/ (χάρη, χωριστά, χούφτα) και του /ç/ χυλός, χέρι. Επίσης υπάρχει και η περίπτωση του /γ/ (γάλα, γορίλας, γούρι) και του /j/ (γέρος, γυμνός).

Η σημασία των αλλόφωνων είναι πολύ μεγάλη για τη φωνολογική ανάλυση μιας γλώσσας προκειμένου να αναπαρασταθούν λέξεις και μορφήματα με σύμβολα του διεθνούς φωνητικού αλφάβητου.

2.3.3. Φωνοτακτικοί κανόνες

Σε κάθε γλώσσα υπάρχουν συγκεκριμένοι συνδυασμοί φωνημάτων με τους οποίους σχηματίζονται λέξεις. Οι συνδυασμοί αυτοί και επιπλέον κάποιοι κανόνες σύνταξης λέξεων και συλλαβών έχουν ως αποτέλεσμα τη διαφοροποίηση γλωσσών που διαθέτουν ίδια ή παρόμοια φωνήματα.

Ως παράδειγμα αναφέρονται κάποιοι φωνοτακτικοί κανόνες της ελληνικής γλώσσας, οι οποίοι είναι οι ακόλουθοι:

- Μια λέξη μπορεί να ξεκινάει αλλά δεν επιτρέπεται να τελειώνει με κλειστό σύμφωνο /p, b, t, d, g/. Υπάρχει η λέξη μπάλα, μπελάς. Εξαίρεση αποτελούν λέξεις που έχουν δανειστεί από άλλες γλώσσες (σνομπ, ματ, τοστ).
- Στην αρχή μιας λέξεως επιτρέπονται ακολουθίες τριών συμφώνων μόνο αν πρόκειται για τους εξής συνδυασμούς: /s/ + (άηχο κλειστό /p t k/) ή (τριβόμενο /f/) + υγρό /l/ ή /r/²⁷. Π.χ. σπλάχνα, στρείδι, στρέμμα, στρώμα.
- Υπάρχει δυνατότητα να ξεκινάει μια λέξη από συγκεκριμένους μόνο συνδυασμούς δύο συμφώνων: Π.χ /ps, ks, pr, tr, kr, pl, pn, kl, kn, tm, mn, fr, fl, xr, st/
- Στη μέση μιας λέξης επιτρέπονται οι παραπάνω συνδυασμοί φωνημάτων που ισχύουν για την αρχή μιας λέξης καθώς επίσης και συνδυασμοί τεσσάρων φωνημάτων όπως π.χ (κστρ) εκστρατεία .
- Στο τέλος μιας λέξης δεν επιτρέπονται σύμφωνα εκτός από τα /s/, /n/.

²⁷ http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/glossology/show.html?id=30

2.3.4. Τονισμός

Ο τονισμός αφορά μια συλλαβή σε μια λέξη και μπορεί να παρέχει πληροφορίες που αφορούν το ύψος, την οξύτητα, την ένταση μια συλλαβής σε σχέση με τις υπόλοιπες συλλαβές μιας λέξης. Ο τονισμός διαφέρει από γλώσσα σε γλώσσα. Σε κάποιες γλώσσες όπως την κινέζικη δίνει μουσικότητα στη συλλαβή, ενώ σε κάποιες άλλες όπως την ελληνική γλώσσα ο τονισμός προσδίδει δυναμικότητα σε μια συλλαβή, ενώ υπάρχουν και περιπτώσεις όπου υπάρχει συνύπαρξη μουσικότητας και δυναμικότητας όσον αφορά τον τονισμό όπως στη σουηδική γλώσσα.²⁸

Στην ελληνική γλώσσα όπως και σε άλλες γλώσσες ο επιτονισμός μπορεί να αλλάζει το νόημα μιας λέξης. Π.χ. χάλι – χαλί, γέρος- γερός κτλ.

2.3.5. Σύνδεση της Φωνολογίας με την Ανάκτηση της Πληροφορίας

2.3.5.1. Φωνολογικός Βρόγχος του μοντέλου μνήμης

Οι Baddeley και Hitch προσπάθησαν να προτείνουν το μοντέλο λειτουργίας της ανθρώπινης μνήμης έτσι ώστε να μπορούν να το συσχετίσουν με την κύρια μνήμη. Το μοντέλο που πρότειναν αποτελείται από τα εξής μέρη τα οποία είναι:

- Μια μονάδα που ελέγχει τη ροή των εισερχόμενων πληροφοριών, αντικαθιστά τις παλαιές πληροφορίες με καινούργιες και συντονίζει τα υπόλοιπα 3 συστήματα βραχυπρόθεσμης αποθήκευσης,
- Ένα φωνολογικό βρόγχο
- Ένα σύστημα βραχυπρόθεσμης αποθήκευσης των οπτικών πληροφοριών
- Αργότερα προστέθηκε και ένα βοηθητικό σύστημα προσωρινής αποθήκευσης πληροφοριών

Ο ρόλος του φωνολογικού βρόγχου είναι για τη βραχυπρόθεσμη αποθήκευση φωνολογικών δεδομένων. Αποτελείται από σύστημα βραχυπρόθεσμης αποθήκευσης φωνολογικών πληροφοριών στο οποίο εισέρχονται όλες οι φωνολογικές πληροφορίες που ακούει ένας άνθρωπος και ένα εργαλείο που επιτρέπει την άρθρωση λέξεων και

²⁸ <https://el.wikipedia.org/wiki/Τονισμός>.

φράσεων που προέρχονται από τις πληροφορίες αυτές, δηλαδή ένα σύστημα το οποίο μας επιτρέπει να επαναλαμβάνουμε τις λέξεις που ακούμε.²⁹

Στο φωνολογικό βρόγχο της μνήμης μπορούν να εισέρθουν φωνολογικές πληροφορίες αλλά και οπτικές εικόνες οι οποίες μπορούν να μετατραπούν σε φωνολογικές. Το φωνολογικό σύστημα της μνήμης συμβάλει στο να θυμόμαστε λέξεις και φωνήματα καθώς και τη σωστή σειρά τους για τη δημιουργία σωστά δομημένων προτάσεων. Στην ανθρώπινη μνήμη ο φωνολογικός βρόγχος συνδέεται με το αριστερό ημισφαίριο του ανθρώπινου εγκεφάλου.

Βασικά χαρακτηριστικά του φωνολογικού βρόγχου είναι τα εξής :

- Είναι ευκολότερη η ανάκληση από τη μνήμη μας λέξεων ή φθόγγων που διαφέρουν μεταξύ τους από ότι λέξεων ή φθόγγων που παρουσιάζουν μεγάλη ομοιότητα. Για παράδειγμα είναι ευκολότερο να θυμόμαστε σειρές συμφώνων με διαφορετικό ήχο όπως h,w,y,k,r,l από ότι συμφώνων με ήχους που μοιάζουν όπως b,g,c,v,t ,p.
- Στη μνήμη διατηρείται δυσκολότερα κάτι το οποίο αρθρώνεται δυνατά και είναι σημασιολογικά ασήμαντο. Η διάρκεια της μνήμης εξαρτάται και από το μέγεθος της λέξης καθώς είναι μεγαλύτερη για μικρές μονοσύλλαβες λέξεις όπως : sum, day ,wit από ότι σε πολυσύλλαβες και μεγάλου μεγέθους λέξεις όπως university , aluminum, organization.
- Είναι εφικτή η μεταφορά οπτικών εικόνων σε λέξεις άρα η μετατροπή οπτικής κωδικοποίησης σε φωνητική. Η μεταφορά των οπτικών εικόνων σε λέξεις και φράσεις μπορεί να μη γίνεται σωστά όταν υπάρχουν προβλήματα στην άρθρωση.
- Μια βλάβη στο φωνολογικό σύστημα της μνήμης έχει ως συνέπεια την αδυναμία άρθρωσης από τον άνθρωπο. Μια τέτοια βλάβη μπορεί να οφείλεται καταστάσεις όπως η λεκτική δυσπραξία. Σε περιπτώσεις δυσαρθρίας δηλαδή κακής άρθρωσης λόγω μυϊκών προβλημάτων ο φωνολογικός βρόγχος λειτουργεί σωστά.³⁰

²⁹ https://en.wikipedia.org/wiki/Baddeley%27s_model_of_working_memory

³⁰ Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556-559.

Μέσα από τα πειράματά του οι Gathercole και Baddeley διαπίστωσαν ότι ελλείμματα στη φωνολογική μνήμη έχει ως συνέπεια την ύπαρξη προβλημάτων στην εκμάθηση νέων λέξεων. Ενώ οι BIEDROŃ & SZCZEPANIAK απέδειξαν ότι μια ισχυρή μνήμη εργασίας έχει ως συνέπεια την ευκολότερη εκμάθηση μιας δεύτερης γλώσσας.³¹

2.3.5.2. Συστήματα εισόδου εξόδου ομιλίας

Τα τελευταία χρόνια έντονο είναι το ενδιαφέρον για την ανάπτυξη συστημάτων προφορικού διαλόγου μεταξύ υπολογιστή και ανθρώπου. Οι γνώσεις που πηγάζουν από τη φωνητική είναι πολύ βασικής σημασίας για την ανάπτυξη και βελτιστοποίηση τέτοιων συστημάτων, αλλά ταυτόχρονα απαραίτητες είναι και οι γνώσεις από τη φωνολογία. Τα συστήματα αναγνώρισης ομιλίας χρησιμοποιούν μηχανισμούς αυτόματης μετατροπής της φωνής στις αντίστοιχες λέξεις και απόδοση νοήματος σε αυτές ώστε το σύστημα να ενεργήσει ανάλογα υπακούοντας στην εντολή.

Πολύ σημαντικό είναι να αναφέρουμε ότι η αναγνώριση ομιλίας (speech recognition) δε θα πρέπει να συγχέεται με την αναγνώριση φωνής (voice recognition) όπου δίνεται μεγάλη σημασία στα ανατομικά χαρακτηριστικά και τη φυσιολογία της φωνής ενός συγκεκριμένου ανθρώπου. Στην αναγνώριση ομιλίας μας ενδιαφέρει η μετατροπή της ομιλίας σε κείμενο και δεν μας αφορά η ταυτοποίηση ενός συγκεκριμένου προσώπου.

Τέτοιου είδους διαλογικά συστήματα χρησιμοποιούνται για τηλεφωνικές κρατήσεις, για παροχή οδηγιών – δρομολογίων μέσω τηλεφώνου, για ξεναγήσεις καθώς επίσης και για φωνητική αναζήτηση στον ηλεκτρονικό υπολογιστή και στο διαδίκτυο χωρίς να απαιτείται η χρήση πληκτρολογίου³². Για να υλοποιηθεί ένα τέτοιο σύστημα πρέπει ο υπολογιστής να αναγνωρίσει τους φθόγγους και τις λέξεις και κατόπιν να αποδώσει ένα νόημα στις λέξεις αυτές.

Η Google έχει συμπεριλάβει τη φωνητική αναζήτηση και τις φωνητικές εντολές. Η δυνατότητα αυτή μπορεί να ενεργοποιηθεί ή να απενεργοποιηθεί σε υπολογιστές ή

³¹ <https://workingmemoryrst.wordpress.com/the-phonological-loop/>

³² Κέρδη Jr, JH, Brown, NG, Mezey, Pδ, & Colombo, LM (2003). *US Patent No. 6,636,831*. Washington, DC: Γραφείο Διπλωμάτων Ευρεσιτεχνίας και Εμπορικών Σημάτων των ΗΠΑ.

tablets και με αυτό τον τρόπο μπορούν να διευκολυνθούν άτομα με αναπηρίες ή να εξοικονομηθεί χρόνος.

Βασικό πρόβλημα αποτελεί: η ύπαρξη συντακτικών ή γραμματικών αμφισημιών στις εκφωνήσεις και άρα κρίνεται από το σύστημα αναγκαία η ανάκτηση επιπλέον πληροφοριών είτε μέσω ερωτήσεων στο χρήστη, είτε με σύνδεση με κάποια πηγή όπως κάποια βάση δεδομένων ή το διαδίκτυο. Εκτός από τις αμφισημίες προβλήματα μπορεί να προκύπτουν αν η φωνητική εντολή που δίνει ο χρήστης περνάει από κανάλι με θόρυβο οπότε αυτό έχει ως συνέπεια η αποκωδικοποίηση του σήματος από το σύστημα να γίνει λάθος. Επίσης μια πολύ μεγάλη εκφώνηση μπορεί να δημιουργήσει προβλήματα κατανόησης από τη μεριά του συστήματος.

Υπάρχουν διάφορα τέτοια συστήματα. Σε όλα αυτά μετά τη σύλληψη τη φωνητικού μηνύματος πραγματοποιείται ψηφιοποίηση καθώς η φωνή έχει αναλογική μορφή. Εν συνεχεία ακολουθεί η κατάτμηση του σε μικρότερα μέρη. Τα τμήματα αυτά συγκρίνονται με αποθηκευμένα φωνητικά δεδομένα και στη συνέχεια ελέγχονται όσον αφορά τη γραμματική και τη σύνταξη τους. Το υπολογιστικό σύστημα κατόπιν εκτελεί την εντολή που του δόθηκε, εφόσον η επικοινωνία πραγματοποιήθηκε με επιτυχία και δεν υπήρξαν αμφισημίες.

2.4. Μορφολογία

Η μορφολογία ασχολείται με τις μικρότερες μορφές σημασίας της γλώσσας αλλά και τους κανόνες με τους οποίους συνδυάζονται τα ελάχιστα αυτά κομμάτια της γλώσσας και σχηματίζουν λέξεις³³. Οι μικρότερες αυτές μορφές της γλώσσας ονομάζονται μορφήματα .

Υπάρχουν λέξεις που αποτελούνται από περισσότερα από ένα μορφήματα αλλά υπάρχουν όμως και λέξεις που αποτελούνται από ένα μόρφημα όπως: και, για, χθες,

³³ Μαγουλάς, Γ., Μαγουλά, Ε., 2017. *Θέματα Συγχρονικής και Διαχρονικής Γλωσσολογίας*. [ηλεκτρ. βιβλ.] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Σελ. 158 .Διαθέσιμο στο: <http://hdl.handle.net/11419/6509>

τι, τώρα. Κάθε μόρφημα προσδίδει μια σημασία στην λέξη στην οποία περιέχεται. Έτσι η λέξη σπιτάκι αναλύεται στα μορφήματα: /σπιτ/-/ακ/-/ι/. Το μόρφημα /σπιτ/ δηλώνει κτίριο που προορίζεται για κατοικία, το μόρφημα /ακ/ δηλώνει υποκοριστικό το οποίο έχει την έννοια του μικρού και το μόρφημα /ι/ δηλώνει το μέρος του λόγου στο οποίο ανήκει η λέξη και παρέχει πληροφορίες όπως το γένος, ο αριθμός και η πτώση στην οποία ανήκει η λέξη.

2.4.1. Κατηγορίες μορφημάτων

Τα μορφήματα μπορούν να ταξινομηθούν βάσει διαφόρων κριτηρίων. Ανάλογα με την αυτοδυναμία τους μπορούν να χωριστούν σε :

- Τα ελεύθερα μορφήματα τα οποία από μόνα τους αποτελούν μια λέξη. Π.χ. ξανά, και, τι, χθες κτλ.
- Δεσμευμένα μορφήματα τα οποία δεν εμφανίζονται ποτέ μόνα τους σε μια λέξη αλλά αποτελούν συστατικό μέρος της π.χ. -ος, -ους, σπιτ-, μιλ- κτλπ. Τα δεσμευμένα μορφήματα χωρίζονται στις παρακάτω κατηγορίες:
 1. Τα μορφήματα τα οποία αποτελούν τον βασικό κορμό μιας λέξης και τις αποδίδουν ένα νόημα. Τα μορφήματα αυτά ονομάζονται βάσεις (stems) Π.χ. στη λέξη «βροχή» το μόρφημα /βροχ/ είναι η βάση όπως και στη λέξη «παιδί» το μόρφημα /παιδ/.
 2. Τα μορφήματα τα οποία συνάπτονται στη βάση της λέξης για παραγωγή ή για την κλίση ονομάζονται προσφύματα (affixes) ή παραθύματα ή παραθέματα. Ανάλογα με τη θέση του σε σχέση με τη ρίζα της λέξης τα προσφύματα ονομάζονται : Προθήματα αν προηγούνται της ρίζας π.χ. ξε- μένω, επιθήματα αν έπονται π.χ. γρι -ούλα και ενθήματα αν προστίθεται στο εσωτερικό της ρίζας μιας λέξης για την παραγωγή μιας νέας λέξης π.χ. αναλα-μ-βάνω^{34 35}

³⁴ <https://el.wiktionary.org/wiki/Πρόσφυμα>

³⁵ Φλιατούρας, Α. *Μορφολογική και λεξιλογική ανάλυση της Ελληνικής: Βιβλιογραφικά και ηλεκτρονικά εργαλεία*. [pdf slides] Τμήμα Φιλολογίας Δ.Π.Θ. και στο Τμήμα Μεσογειακών Σπουδών του Παν. Αιγαίου Online: <https://repository.edulll.gr/edulll/retrieve/1389/246.pdf>

Ανάλογα την σημασία τους τα μορφήματα μπορούν να χωριστούν σε :

- Λεξικά μορφήματα τα οποία δηλώνουν ιδιότητα, κατάσταση, πράξη κτλ.
- Γραμματικά μορφήματα τα οποία δηλώνουν τη γραμματική κατηγορία στην οποία ανήκει μια λέξη και σε κάποιες περιπτώσεις μια λέξη προέρχεται από κάποια άλλη .

Π.χ. στη λέξη αθωότητα αναλύεται στο λεξικό μόρφημα /αθω/ και το γραμματικό μόρφημα /ότητα/, στη λέξη τρέχω αποτελείται από το λεξικό μόρφημα /τρεχ/ και το γραμματικό μόρφημα /ω/.

2.4.2 Οι Λέξεις

Βασικό αντικείμενο μελέτης της μορφολογίας είναι και οι λέξεις. Λέξη είναι μια μονάδα του λόγου, εύκολα αναγνωρίσιμη από τους ομιλητές της γλώσσας στην οποία ανήκει. Η κάθε λέξη περιέχει μια σημασία και είναι απαραίτητη στον προφορικό λόγο. Ένα σύνολο λέξεων απαρτίζει μια πρόταση. Σε μια φράση μια λέξη ξεχωρίζει καθώς πριν και μετά από αυτήν υπάρχουν παύσεις.

Οι λέξεις μπορούν να ταξινομηθούν σε κλιτές και άκλιτες. Μια απλή κλιτή λέξη έχει πάντα μια ρίζα και μια κατάληξη. Η ρίζα αποτελεί τον πυρήνα μιας λέξης και ανάλογα την κλίση αλλάζει και η κατάληξη. Ο όρος λέξημα αφορά μια λέξη με μια συγκεκριμένη σημασία η οποία παραμένει ίδια στις διάφορες εκδοχές της κλίσης της. Για παράδειγμα το λέξημα μιλώ έχει κοινή σημασία που υπάρχει σε διάφορες λέξεις όπως μιλάς, μιλά, μιλούσα..

Μια λέξη μπορεί να αποτελείται από μια μορφή ιδίως αν είναι άκλιτη ή ξενόφερτη (και, με, σε, τρακτέρ, μοτέρ) ή από περισσότερες κυρίως όταν είναι κλιτή (παίζω, έπαιξα, έπαιξα κτλ). Υπάρχουν διαφορετικοί τρόποι ανάλυσης μιας λέξης όπως σε ρίζα και κατάληξη, σε συλλαβές και οι τρόποι αυτοί διαφέρουν ανάλογα τη γλώσσα που εξετάζουμε.

2.4.3. Βασικές μορφολογικές διαδικασίες

Υπάρχουν διάφοροι τρόποι δημιουργίας λέξεων όταν οι λέξεις δεν αποτελούν ελεύθερα μορφήματα. Αυτοί είναι οι παρακάτω:

- **Η σύνθεση** η οποία είναι η δημιουργία ενός λεξήματος συνδυάζοντας δύο ήδη υπάρχοντα λεξήματα. Π.χ. λαθροκυνηγός, ελαφρόμυαλος κτλ. Στην ελληνική γλώσσα για τη σύνθεση μπορεί να χρησιμοποιείται κάποιες φορές το φωνήεν –ο όπως στη λέξη μεγαλοδικηγόρος, το φωνήεν –η θανατηφόρος ή ακόμη να χρησιμοποιούνται τα λεξήματα χωρίς να παρεμβάλλεται κάποιος ενδιάμεσος χαρακτήρας π.χ. πολύχρωμος.
- **Η παραγωγή** κατά την οποία έχουμε συνδυασμό λεξικών και γραμματικών μορφημάτων προκειμένου να σχηματιστεί μια νέα λέξη. Π.χ. από-κληρ-ώνω, μετά-μορφ –ώνω, νε-ότητ-α.
- **Η κλίση** και οι αντίστοιχες λέξεις ονομάζονται κλιτές. Στις λέξεις αυτές πραγματοποιείται συνένωση του θέματος μ' ένα γραμματικό επίθεμα. Από το γραμματικό μόρφημα προκύπτουν πληροφορίες όπως το γένος, ο αριθμός και η πτώση στην περίπτωση των ουσιαστικών και τον επιθέτων ενώ στην περίπτωση των ρημάτων τον αριθμό, το πρόσωπο και τον χρόνο στις περιπτώσεις των ρημάτων.
- **Η Μετατροπή** κατά τη οποία έχουμε αλλαγή γραμματικής κατηγορίας σε μια λέξη. Π.χ. ζώη→ζω, κυνηγώ→ κυνηγός κτλ.

2.4.4. Περιστολή (Stemming)

Η διαδικασία του stemming σχετίζεται με την αφαίρεση από μια λέξη των προσφυμάτων που αφορούν την γραμματική κατηγορία και αριθμό της λέξης έτσι ώστε να απομένει μόνο η βάση ή ρίζα μιας λέξης.

Η διαδικασία αυτή είναι πολύ σημαντική στην ανάκτηση της πληροφορίας και πολλές φορές περιλαμβάνεται σε εφαρμογές επεξεργασίας φυσικής γλώσσας. Η ανάλυση μιας λέξης και η αποκοπή των καταλήξεων λύνει πολλά προβλήματα στην αναζήτηση πληροφορίας, επιστρέφει περισσότερα αποτελέσματα και περιλαμβάνει

όλες τις πιθανές εκδοχές μιας λέξης. Για παράδειγμα μια αναζήτηση του όρου «υπολογιστής» πρέπει να λαμβάνει υπόψη και τους όρους υπολογιστές, υπολογιστών, υπολογιστή οι οποίοι έχουν ίδια ρίζα /υπολογιστ/. Επίσης η διαδικασία του stemming είναι πολύ χρήσιμη στη αυτοματοποιημένη σύγκριση διαφορετικών κειμένων.

Γι' αυτό το λόγο οι μηχανές αναζήτησης χρησιμοποιούν αλγόριθμους stemming προκειμένου να επιτυγχάνεται επιστροφή περισσότερων συναφών αποτελεσμάτων. Υπάρχουν διάφοροι αλγόριθμοι αποκοπής λέξεων οι οποίοι χρησιμοποιούνται. Κάποιοι από τους πιο σημαντικούς είναι οι παρακάτω:

- Ο αλγόριθμος του Porter ο οποίος σχεδιάστηκε για την αφαίρεση καταλήξεων από λέξεις στην αγγλική γλώσσα. Αλγόριθμος αυτός περιέχει διάφορους κανόνες που αφορούν τις καταλήξεις που μπορεί να έχουν οι λέξεις στην αγγλική γλώσσα και ακολουθεί κάποια βήματα τα οποία πρέπει να εκτελεστούν στη σειρά.

Στον αλγόριθμο του Porter στο πρώτο βήμα για κάθε λέξη εκτελούνται οι εξής μετατροπές:

A

sses → ss
ies → i
ss → ss
s → ∅

B

(*φωνήεν*) ing → ∅
(*φωνήεν*) ed → ∅

Στο δεύτερο βήμα εκτελούνται οι παρακάτω μετατροπές:

ational → ate
izer → ize
ator → ate

- Ο αλγόριθμος του Lovins ο οποίος περιέχει ένα σύνολο από κανόνες και μια λίστα από καταλήξεις. Σύμφωνα με τον αλγόριθμο αυτό πρώτα πρέπει να αφαιρεθεί η μεγαλύτερη κατάληξη.
- Η μέθοδος n-gram σύμφωνα με την οποία δίνεται ο αριθμός n και βάση αυτού υπολογίζονται τα μοναδικά ζεύγη n χαρακτήρων . Για παράδειγμα έστω ότι έχουμε τις λέξεις «computation, computing» για n=2 περιέχει τα εξής ζεύγη :
 - για τη λέξη computation :co-om-mp-pu-ut-ta-at-ti-io-on.
 - για τη λέξη computing: co-om-mp-pu-ut-ti-in-ng.

2.4.5. Υπολογιστική μορφολογία

Ο αυτοματοποιημένος τρόπος διαχωρισμού μιας λέξης στις μορφολογικές πληροφορίες που προκύπτουν αποτελεί την αυτοματοποιημένη μορφολογία της οποίας η σημασία είναι πολύ μεγάλη σε διάφορες εφαρμογές όπως για παράδειγμα στη μηχανική μετάφραση. Επιπλέον πολύ σημαντική είναι και η αυτοματοποιημένη παραγωγή και ανάλυση μιας λέξης . Π.χ.

Μορφολογική ανάλυση: books: books+N+Pl

Μορφολογική Παραγωγή: book+N+Pl: books

2.5. Η Σύνταξη

Η σύνταξη ασχολείται με τη μελέτη όλων των κανόνων και δομών βάσει των οποίων καθορίζονται όλοι οι πιθανοί συνδυασμοί των λέξεων και των μορφημάτων προκειμένου να σχηματιστούν προτάσεις.

Το συντακτικό μιας γλώσσας είναι μέρος της γραμματικής της και ακολουθεί κάποιες γραμματικές δομές. Η πρόταση: «Η Μαρία διαβάζει το βιβλίο» είναι μια γραμματικά ορθή πρόταση όπου το υποκείμενο «Μαρία» συμφωνεί στον αριθμό και το πρόσωπο με το ρήμα «διαβάζει». Αντίθετα η πρόταση : «Η Μαρία διαβάζουν το βιβλίο» είναι μια μη γραμματική πρόταση καθώς ο αριθμός του υποκειμένου δεν αντιστοιχεί με τον αριθμό του ρήματος. Από τα δύο παραπάνω παραδείγματα

προκύπτει ότι ένας βασικός κανόνας είναι το υποκείμενο να συμφωνεί σε αριθμό με το ρήμα .

Η δομή σε μια πρόταση είναι ιεραρχική υπό την έννοια ότι δεν είναι τυχαία. Αυτό σημαίνει ότι κάθε λέξη είναι πολύ βασικό να γράφεται στη σωστή θέση ή σε συμφωνία με άλλες λέξεις ανάλογα με τους κανόνες της κάθε γλώσσας. Για παράδειγμα από την πρόταση «Η Μαρία το βιβλίο διαβάσει θέλει να», δε προκύπτει κάποιο νόημα. Άρα υπάρχουν συγκεκριμένοι τρόποι για να γραφτεί μια φράση. Η παραπάνω πρόταση αποκτά νόημα όταν η σειρά των λέξεων είναι: «Η Μαρία θέλει να διαβάσει το βιβλίο». Την ίδια ακριβώς σημασία θα είχε η πρόταση αν η σειρά των λέξεων ήταν : «Το βιβλίο, η Μαρία θέλει να διαβάσει» ή «Η Μαρία, το βιβλίο θέλει να διαβάσει».

Βέβαια για να υπάρχει νόημα σε μια πρόταση δε χρειάζεται μόνο να ακολουθεί κάποιες γραμματικές δομές, και κάποιους αποδεκτούς γραμματικούς συνδυασμούς λέξεων, αλλά να έχει και κάποιο νόημα. Στο παράδειγμα του Chomsky «Colorless green ideas sleep furiously» βλέπουμε ότι παρόλο που η πρόταση είναι γραμματικά ορθή, δεν είναι σωστή σημασιολογικά.

2.5.1. Συστατικά μιας πρότασης

Για να μπορέσουμε να διατυπώσουμε μια πρόταση στην ελληνική γλώσσα, αλλά και στις υπόλοιπες γλώσσες, χρειαζόμαστε τουλάχιστον δύο βασικούς όρους: το υποκείμενο και το κατηγορήμα.

Το υποκείμενο είναι η λέξη που φανερώνει ποιος ενεργεί, βρίσκεται σε μια κατάσταση ή δέχεται μια ενέργεια³⁶ . Το κατηγορήμα αποτελείται από το ρήμα της πρότασης και τις λέξεις που το συνοδεύουν. Οι λέξεις που συνοδεύουν το ρήμα μπορεί να είναι αντικείμενα ή κατηγορούμενα . Αντικείμενο είναι η λέξη η οποία δηλώνει το πρόσωπο, το ζώο ή το πράγμα στο οποίο μεταβαίνει ή στο οποίο αναφέρεται η ενέργεια του υποκειμένου. Για παράδειγμα στην πρόταση «Ο Γιώργος πίνει γάλα». Η λέξη «γάλα» είναι το αντικείμενο της πρότασης. Κατηγορούμενο είναι η λέξη που δίνει μια ιδιότητα στο υποκείμενο ή στο αντικείμενο μιας πρότασης. Η λέξη αυτή μπορεί να είναι επίθετο, ουσιαστικό, αντωνυμία, αριθμητικό, μετοχή

³⁶ [https://el.wikipedia.org/wiki/Υποκείμενο_\(γραμματική\)](https://el.wikipedia.org/wiki/Υποκείμενο_(γραμματική))

εμπρόθετο ή μια άλλη πρόταση. Για παράδειγμα στην πρόταση «Η Μαρία είναι όμορφη». Η λέξη «όμορφη» είναι κατηγορούμενο καθώς προσδίδει ένα χαρακτηριστικό στο υποκείμενο.

Ένας άλλος τρόπος ανάλυσης μιας πρότασης είναι σε ονοματικές φράσεις (ΟΦ) και ρηματικές φράσεις (ΡΦ). Σε περιπτώσεις που ονοματικές φράσεις συνδυάζονται με προθέσεις προκύπτει μια άλλη κατηγορία φράσεων οι οποίες είναι οι προθετικές φράσεις. Για παράδειγμα στην πρόταση «Πήγα για καφέ με τους φίλους μου », η πρόθεση «με» συνδυάζεται με την ΟΦ «τους φίλους μου» και έτσι προκύπτει η προθετική φράση (ΠρΦ) «με τους φίλους μου».

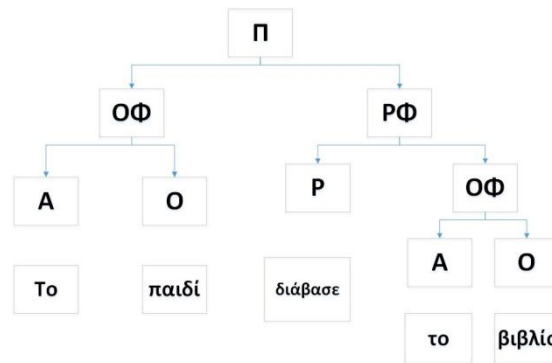
Μια ονοματική φράση έχει ως κεντρική λέξη ένα ουσιαστικό ή άλλο μέρος του λόγου (όπως αντωνυμία , επίθετο ή αριθμητικό) που μπορεί να λειτουργήσει ως ουσιαστικό. Για παράδειγμα στη φράση : «Κάποιος κάθισε δίπλα μου» , η λέξη «κάποιος» μπορεί να θεωρηθεί ως ονοματική φράση. Επίσης ως ονοματική φράση μπορεί να θεωρηθεί και ένας συνδυασμός λέξεων που μπορεί να περιέχει ουσιαστικό, επίθετο, αντωνυμία και άρθρο. Έτσι στην πρόταση «Η αδελφή μου αγόρασε ένα παλτό» ο συνδυασμός άρθρου, ουσιαστικού και αντωνυμίας «Η αδελφή μου» αποτελεί ονοματική φράση. Μια ονοματική φράση μπορεί να λειτουργήσει ως υποκείμενο, αντικείμενο, κατηγορούμενο, ως προσδιοριστικό ή συμπλήρωμα πρότασης.

Υπάρχει επίσης και η ρηματική φράση που αποτελείται από ένα ρήμα και τις λέξεις που το συνοδεύουν οι οποίες μπορεί να είναι αντικείμενο ή κατηγορούμενο.

2.5.2. Δενδρική δομή πρότασης

Η ανάλυση μιας πρότασης στα συστατικά της στοιχεία μπορεί να αναπαρασταθεί με διάφορους τρόπους. Ένας από αυτούς είναι η χρήση δεντρικών δομών. Μια τέτοια δομή περιγράφεται με τη χρήση μιας ρίζας, κόμβων και κλαδιών.

Με τη χρήση δεντρικών δομών μπορούμε να αναλύσουμε μια πρόταση σε ονοματικές φράσεις και ρηματικές φράσεις. Ένα παράδειγμα φαίνεται παρακάτω:



Στην πρόταση του παραδείγματος τα συστατικά είναι τα εξής:

- το βιβλίο
- διάβασε
- διάβασε το βιβλίο
- Το παιδί
- Το παιδί διάβασε το βιβλίο.

Η παραπάνω ιεραρχία μπορεί να εκφραστεί και με τη χρήση αγκυλών ως εξής:

[π[ΟΦ[Α ΤΟ]][Ο παιδί]][ΡΦ[Ρ διάβασε][ΟΦ[Α το][Ο βιβλίο]]].

2.5.3. Μορφοσύνταξη

Η σύνταξη αφορά τον τρόπο με τον οποίο είναι οργανωμένες οι λέξεις σε μια πρόταση ενώ η μορφολογία αφορά την εσωτερική οργάνωση των λέξεων. Αυτό σημαίνει ότι οι γραμματικές πληροφορίες που προκύπτουν σε μια πρόταση πρέπει να σχετίζονται με τις γραμματικές πληροφορίες που προκύπτουν από τα μορφήματα που αφορούν τις καταλήξεις των κλιτών λέξεων. Έτσι λοιπόν προκύπτει ότι υπάρχει στενή σχέση μεταξύ της σύνταξης και της μορφολογίας.³⁷

2.5.4. Η γραμματική

Πολύ σημαντική για τη σύνταξη είναι η γραμματική ανάλυση με την οποία προσδιορίζονται οι πιθανές αποδεκτές δομές μιας γλώσσας. Οι γραμματικοί κανόνες είναι της μορφής $A \rightarrow B$ όπου κάθε κανόνας που ταιριάζει με το αριστερό στοιχείο, αντικαθίσταται από το περιεχόμενο του δεξιού μέρους.

³⁷http://www.greeklanguage.gr/greekLang/modern_greek/tools/lexica/glossology/show.html?id=120

Ο Νόαμ Τσόμσκι το 1956 ταξινόμησε τις γραμματικές σε μια ιεραρχία ανάλογα τους κανόνες παραγωγής τους. Μια τυπική γραμματική αποτελείται από το σύνολο $G(V, T, P, S)$ όπου S συμβολίζει την αρχική κατάσταση, V ένα πεπερασμένο σύνολο από μη τερματικά στοιχεία τα οποία περιγράφουν τις δομές μιας γλώσσας, T ένα πεπερασμένο σύνολο από τερματικά στοιχεία τα οποία απαρτίζουν το λεξικό μιας γλώσσας, P ένα πεπερασμένο σύνολο από κανόνες παραγωγής οι οποίοι ορίζουν τους επιτρεπτούς συνδυασμούς των λέξεων μιας γλώσσας.

Το σύνολο $G(V, T, P, S)$ εκτός από την περιγραφή της γραμματικής μιας φυσικής γλώσσας, περιγράφει επίσης την γραμματική μιας γλώσσας προγραμματισμού ή μιας γλώσσας δημιουργίας ερωτημάτων για μια βάση δεδομένων.

Ο Νόαμ Τσόμσκι ταξινόμησε τις γραμματικές σε μια ιεραρχία αποτελούμενη από τέσσερις κατηγορίες οι οποίες συνοπτικά είναι οι ακόλουθες

- Γενικές γραμματικές οι οποίες είναι αναγνωρίσιμες από μηχανές Turing. Η μορφή των κανόνων παραγωγής είναι $\alpha \rightarrow \beta$ με $\alpha \neq \epsilon$ όπου ϵ είναι το κενό σημείο και $\alpha, \beta \in (T \cup P)$.
- Γραμματικές με συμφραζόμενα για τις οποίες οι κανόνες παραγωγής είναι της μορφής $\alpha \rightarrow \beta$ με $|\alpha| \leq |\beta|$ και $\alpha, \beta \in (T \cup P)$ και επιτρέπεται $S \rightarrow \epsilon$.
- Γραμματικές δίχως συμφραζόμενα οι οποίες αποτελούνται από ένα σύνολο κανόνων που μπορούν να παράγουν όλες τις συμβολοσειρές μιας γλώσσας. Η αρχική κατάσταση S ονομάζεται και παραγωγός κατάσταση γιατί από αυτήν παράγονται όλες οι συμβολοσειρές. Οι γραμματικές αυτές αποτελούνται από αναδρομικούς κανόνες βάσει των οποίων δημιουργούνται οι συμβολοσειρές με αναδρομικό τρόπο και έναν τερματικό κανόνα ο οποίος τερματίζει την παραγωγή της συμβολοσειράς.
- Κανονικές γραμματικές οι οποίες έχουν τη μορφή $X \rightarrow \sigma$ και $X \rightarrow \sigma Y$ όπου X, Y είναι μεταβλητές και $\sigma \in S$ είναι σύμβολα του αλφάβητου ή η κενή συμβολοσειρά

Στην επιστήμη της πληροφορικής υπάρχουν γλώσσες προγραμματισμού οι οποίες ελέγχουν τη λειτουργία των υπολογιστών. Οι γλώσσες αυτές διαθέτουν τη δική τους γραμματική και δικούς τους συντακτικούς κανόνες. Οι γραμματικές αυτές

ονομάζονται τυπικές γραμματικές και περιέχουν το δικό τους αλφάβητο το οποίο μπορεί να περιέχει γράμματα, σύμβολα και αριθμούς.

Για τον έλεγχο της γραμματικής και συντακτικής εγκυρότητας των προτάσεων που γράφουμε υπάρχουν προγράμματα τα οποία ονομάζονται αναλυτές (parsers) τα οποία δέχονται ως είσοδο διαδοχικές εντολές κώδικα και σκοπό έχουν να τις χωρίσουν σε συντακτικά τμήματα τα οποία πρέπει στη συνέχεια να εξεταστούν σε σημασιολογικό επίπεδο. Η όλη διαδικασία θυμίζει τη διαδικασία συντακτικής ανάλυσης μιας φυσικής γλώσσας.

Υπάρχουν δύο κατηγορίες συντακτικών αναλυτών οι top down και οι bottom up. Οι top down αναλυτές δημιουργούν συντακτικά δέντρα ξεκινώντας από τη ρίζα και συνεχίζοντας προς τα κάτω, ενώ στους bottom up η σάρωση γίνεται από κάτω προς τα πάνω.

Συντακτικούς αναλυτές διαθέτουν και συστήματα διαχείρισης βάσεων δεδομένων, προκειμένου να είναι εφικτή η ανάπτυξη γλώσσας ερωτημάτων τα οποία έχουν ως σκοπό την ανάκτηση συγκεκριμένων δεδομένων από τη βάση.

2.5.5. Αυτόματη επεξεργασία φυσικής γλώσσας

Για τη μηχανική επεξεργασία της γλώσσας πολύ σημαντική είναι η δυνατότητα αυτοματοποιημένου διαχωρισμού των προτάσεων καθώς επίσης και η συντακτική ανάλυση που αποσκοπεί στον καθορισμό της σειράς και των πιθανών συνδυασμών με τους οποίους εμφανίζονται διάφορα μέρη του λόγου σε μια πρόταση.

Σε αυτή την περίπτωση οι συντακτικοί αναλυτές εξετάζουν το κείμενο εισόδου όσον αφορά σε σχέση με τους γραμματικούς κανόνες που διαθέτουν προκειμένου να δημιουργήσουν συντακτικές δομές.

2.6. Η Σημασιολογία

Σημασιολογία είναι ο υποκλάδος της γλωσσολογίας ο οποίος εξετάζει τη σημασιολογική δομή μιας γλώσσας³⁸. Με άλλα λόγια ασχολείται με το σημασιολογικό περιεχόμενο των μορφημάτων, των λέξεων και των φράσεων μιας γλώσσας και αποτελεί ένα από τα επίπεδα της γλωσσικής ανάλυσης.

Ο όρος σημασία σχετίζεται με την ιδιότητα και τη μορφή των πραγμάτων. Τα βασικά είδη σημασίας είναι : η λεξική, η προτασιακή και η κειμενική σημασία. Σύμφωνα με τον Ullman η σημασία είναι η σχέση που έχει το σημαίνον με το σημαινόμενο δηλαδή τα ονόματα με τις έννοιες, ενώ κατά τους Ogden και Richards η σημασία αφορά την ερμηνεία της σχέσης μεταξύ σημαίνοντος και σημαινόμενου³⁹.

2.6.1. Κατηγορίες σημασιολογίας

Ανάλογα με διάφορα κριτήρια μπορούμε να διακρίνουμε διάφορες κατηγορίες στη σημασιολογία.

Μια διάκριση είναι σε διαχρονική και σε συγχρονική. Η διαχρονική σημασιολογία εξετάζει τις μεταβολές στη σημασία των λέξεων. Οι μεταβολές αυτές σχετίζονται με κοινωνικούς, πολιτικούς και οικονομικούς κυρίως λόγους. Οι μεταβολές στη σημασία των λέξεων γίνονται με την πάροδο του χρόνου και μπορεί να είναι πολλών ειδών όπως:

- Επέκταση της σημασίας π.χ. : το ρήμα παιδεύω παλαιότερα είχε τη σημασία «ανατρέφω παιδιά» ενώ σήμερα σημαίνει «διδάσκω» «εκπαιδεύω»
- μπορεί όμως να συμβαίνει και περιορισμός της σημασίας μιας λέξης
- και μεταφορά π.χ. πόδι «ανθρώπου» - πόδι «τραπεζιού»

Η συγχρονική σημασιολογία επικεντρώνει το ενδιαφέρον της στην σύγχρονή προσέγγιση της γλώσσας και εξετάζει τη σχέση μεταξύ της σημασίας και της μορφής των λέξεων καθώς και το ρόλο της σημασίας στην γραμματική μιας λέξης.

³⁸ Μπαμπινιώτης, Γ.(1985). *Εισαγωγή στη Σημασιολογία*. Αθήνα. Σελ. 7.

³⁹ Μπαμπινιώτης, Γ.(1985). *Εισαγωγή στη Σημασιολογία*. Αθήνα. Σελ. 22.

Η λεξική σημασιολογία σχετίζεται με την απόδοση σημασίας λημμάτων σ' ένα λεξικό μέσω σύνθετων δομών. Π.χ. για τη λέξη καταδιώκω υπάρχει η αντίστοιχη ερμηνεία: «ακολουθώ κάποιον, τρέχω πίσω από αυτόν για να τον συλλάβω ή για να τον σκοτώσω»⁴⁰.

2.6.2. Σημασία λέξης

Η σημασία των λέξεων είναι καθοριστική για το νόημα που έχει μια πρόταση. Η σημασία μιας λέξης μπορεί να είναι είτε λεξική, είτε γραμματική⁴¹. Η λεξική σημασία αφορά πράγματα και καταστάσεις της εξωγλωσσικής πραγματικότητας, ενώ η γραμματική σημασία διαθέτει εσωγλωσσική θέση μιας λέξης σε προτάσεις⁴² δηλαδή σχετίζεται με γραμματικές πληροφορίες όπως ο χρόνος, το πρόσωπο και ο αριθμός ενός ρήματος, το γένος, ο αριθμός και η πτώση ενός ουσιαστικού, η οριστικότητα που δηλώνεται με τα άρθρα ή σύνδεση⁴³.

2.6.3. Σημασιολογικές σχέσεις λέξεων

Ανάμεσα στις λέξεις και τις σημασίες υπάρχουν κάποιες σχέσεις. Υπάρχει η περίπτωση της πολυσημίας όταν σε μια λέξη αποδίδονται περισσότερα του ενός νοήματα. Π.χ. για τη λέξη χώρα αποδίδονται οι παρακάτω σημασίες:⁴⁴

- (ιστορία) οι αγροτικές περιοχές μιας πόλης-κράτους, η ύπαιθρος, σε αντίθεση προς το άστυ.
- τόπος που κατοικείται από ένα συγκεκριμένο λαό, η πατρίδα
- κράτος, κρατική οντότητα. Π.χ. *άνοιξαν τα σύνορα με τη γειτονική μας χώρα*.

⁴⁰ http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/search.html?lq=%CE%BA%CE%B1%CF%84%CE%B1%CE%B4%CE%B9%CF%8E%CE%BA%CF%89&dq=

⁴¹ Φιλιππάκη, Ε.Υ. (1992). *Εισαγωγή στη θεωρητική γλωσσολογία*. Αθήνα: Νεφέλη. σελ. 273.

⁴² Παυλίδου, Θ. (2008). *Επίπεδα γλωσσικής ανάλυσης*. Αθήνα: Ινστιτούτο Νεοελληνικών Σπουδών (Ιδρυμα Μανόλη Τριανταφυλλίδη).

⁴³ http://www.greek-language.gr/digitalResources/modern_greek/tools/lexica/glossology_edu/lemma.html?id=80&ur=4

⁴⁴ <https://el.wiktionary.org/wiki/χώρα>.

- (για πολλά νησιά) ορεινή πρωτεύουσα του νησιού, εκεί που ήταν συνήθως το κάστρο του και η έδρα της ηγεσίας. Π.χ. *μόνο στη χώρα μπορείς να το βρεις αυτό*.
- περιοχή του ανθρώπινου σώματος. Π.χ. *ηβική χώρα, κροταφική χώρα*

Άλλη περίπτωση αποτελεί η **ομωνυμία**. Στην περίπτωση της ομωνυμίας δύο ή παραπάνω λέξεις οι οποίες έχουν το ίδιο σημαίνουν αλλά διαφορετικά σημεινόμενα. Π.χ. η λέξη «χοίρος» και «χήρος» έχουν διαφορετικό σημεινόμενο, παρόλο που είναι ομόηχες, άρα είναι ομώνυμες καθώς διαθέτουν διαφορετικό σημεινόμενο⁴⁵.

Επίσης υπάρχει και η περίπτωση της **συνωνυμίας** όπου μια σημασία μπορεί να αποδίδεται σε δύο ή και παραπάνω λεκτικούς τύπους. Π.χ. για τη λέξη αδέξιος υπάρχουν και οι λέξεις : άγαρμος και άτσαλος⁴⁶ στις οποίες μπορεί να αποδοθεί η ίδια σημασία.

Όσον αφορά τις λεξικές σημασίες υπάρχουν οι σχέσεις υπωνυμίας και αντωνυμίας. **Υπώνυμο** είναι μια λέξη που υπάγεται σ' ένα ευρύτερο σύνολο. Π.χ. η λέξη μαργαρίτα είναι υπώνυμο της λέξης λουλούδι. **Αντώνυμο** μιας λέξης είναι μια άλλη λέξη με αντίθετη έννοια.

2.6.4. Η σημασιολογική συνάφεια στην ανάκτηση της πληροφορίας

Πολύ σημαντική είναι και η δυνατότητα κατανόησης της σημασίας του περιεχομένου του ιστού από τους υπολογιστές ώστε σε κάθε αναζήτηση να επιστρέφονται τα πιο κατάλληλα σημασιολογικά αποτελέσματα⁴⁷. Για την αναζήτηση μπορεί να χρησιμοποιείται ένας όρος μπορεί όμως και μια φράση. Το σύστημα ανάκτησης πρέπει να είναι δομημένο με τέτοιο τρόπο ώστε να μπορεί εντοπίσει τα κατάλληλα κάθε φορά έγγραφα.

⁴⁵ Παυλίδου, Θ. (2008). Επίπεδα γλωσσικής ανάλυσης. Αθήνα: Ινστιτούτο Νεοελληνικών Σπουδών (Ιδρυμα Μανόλη Τριανταφυλλίδη). Σελ. 87.

⁴⁶ http://www.neurolingo.gr/el/online_tools/lexiscope.htm?term=αδέξιος

⁴⁷ Πούλος, Μ., 2015. *Σημασιολογική επεξεργασία της πληροφορίας*. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Σελ. 23.

Πολύ σημαντική για την επιστροφή σημασιολογικά συναφών αποτελεσμάτων είναι η απόδοση βάρους. Η πιο συχνά χρησιμοποιούμενη μέθοδος είναι ο υπολογισμός του μεγέθους tf-idf . Ο παράγοντας tf σχετίζεται με συχνότητα εμφάνισης του όρου της αναζήτησης στα διάφορα έγγραφα και ο παράγοντας idf τα επίπεδα ανάκλησης και ακρίβειας⁴⁸.

2.6.5. Αλγόριθμοι κατάταξης

Τη δυνατότητα ανάκτησης συναφών αποτελεσμάτων στην αναζήτηση πρακτικά την παρέχουν οι αλγόριθμοι κατάταξης που χρησιμοποιεί η κάθε μηχανή αναζήτησης. Ο αντίστοιχος αλγόριθμος της Google ήταν ο page rank.

Η κατάταξη σημασιολογικής συνάφειας των ιστοσελίδων είναι ένας αριθμός ο οποίος όσο μεγαλύτερος είναι, τόσο μεγαλύτερη είναι η συνάφεια των αποτελεσμάτων της αναζήτησης σε σχέση με το ερώτημα που έχει υποβληθεί. Τα κριτήρια της αξιολόγησης είναι πολλά και κάποια από αυτά είναι: οι λέξεις κλειδιά στην ετικέτα του τίτλου (title tag) του κώδικα της ιστοσελίδας, η ύπαρξη λέξεων κλειδιών στο description tag, οι επικεφαλίδες, η θέση των λέξεων κλειδιών μέσα στο περιεχόμενο του ιστοτόπου, το sitemap, οι εσωτερικοί υπερσύνδεσμοι και γενικώς όλα τα meta tags που χρησιμοποιούνται για την περιγραφή των χαρακτηριστικών μιας ιστοσελίδας όπως meta name και meta content, κ.α.⁴⁹ Βέβαια τα τελευταία χρόνια αλγόριθμοι page rank έχουν αρχίσει να εγκαταλείπονται μάλιστα η Google το 2014 σταμάτησε να ανανεώνει το εργαλείο κατάταξης των ιστοσελίδων της.

Ο νέος τρόπος αξιολόγησης της σημασιολογικής συνάφειας των αποτελεσμάτων των μηχανών αναζήτησης είναι το Domain Authority το οποίο υπολογίζει σε μια κλίμακα από 0 έως το 100 τα αποτελέσματα της αναζήτησης, προσμετρώντας τουλάχιστον 40 παράγοντες συνάφειας για μια ιστοσελίδα. Το Domain Authority έχει τη δυνατότητα να ανιχνεύσει το δημιουργό μιας ιστοσελίδας και την γεωγραφική τοποθεσία από τη

⁴⁸ Πούλος, Μ., 2015. *Σημασιολογική επεξεργασία της πληροφορίας*. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Σελ. 27.

⁴⁹ <https://backlinko.com/google-ranking-factors>

οποία προέρχεται. Αυτό δίνει τη δυνατότητα ανώνυμα και αναξιόπιστα αντικείμενα να βαθμολογούνται με 0 σκορ⁵⁰.

2.6.6. Σημασιολογικές Μηχανές Αναζήτησης

Η υπολογιστική σημασιολογία αναφέρεται στην αυτόματη σημασιολογική ανάλυση προτάσεων φυσικής γλώσσας. Η επόμενη γενιά των έξυπνων εφαρμογών απαιτεί υψηλότερο επίπεδο σημασιολογικής ανάλυσης από ότι έχουμε σήμερα⁵¹.

Σε αυτό μπορεί να βοηθήσει ο σημασιολογικός ιστός που βασίζεται στις σχέσεις που υπάρχουν μεταξύ των διαδικτυακών πόρων και στην ύπαρξη των μεταδεδομένων. Στις σημασιολογικές μηχανές αναζήτησης η αναζήτηση γίνεται σ' ένα σημασιολογικό δίκτυο και εκτός από έναν όρο αναζήτησης, μπορούν να δοθούν τα κατάλληλα αποτελέσματα και σε ερωτήσεις όπως «Ποιος είναι ο πρωθυπουργός της Γαλλίας;» ή «Ποια είναι η πρωτεύουσα της Ισλανδίας;». Στην ουσία οι σημασιολογικές μηχανές αναζήτησης του WEB 3.0 διαθέτουν διαφορετική νοημοσύνη από τις αντίστοιχες μηχανές του διαδικτύου. Γι' αυτό το λόγο στις σημασιολογικές μηχανές αναζήτησης παίζει μεγάλο ρόλο η σημασιολογική ανάλυση των φράσεων ώστε να γίνεται κατανοητό το νόημά τους.

2.7. Πραγματολογία

Πραγματολογία ορίζεται ο κλάδος της γλωσσολογίας που ασχολείται με τη χρήση της γλώσσας σε πραγματικές συνθήκες επικοινωνίας⁵² και γι αυτό το λόγο διαφέρει από τη σημασιολογία. Στην πραγματολογία το ενδιαφέρον επικεντρώνεται σε εκφωνήματα και γλωσσικές εκφράσεις που σκοπό έχουν τη χρήση της γλώσσας για επικοινωνιακούς κυρίως λόγους.

⁵⁰ Satow, M., & Widman, J. M. (2013). *U.S. Patent No. 8,521,743*. Washington, DC: U.S. Patent and Trademark Office.

⁵¹ <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1162/handouts/Computational-Semantics.pdf>

⁵² Μπέλλα, Σ.(2015). *Πραγματολογία, Από τη γλωσσική επικοινωνία στη γλωσσική διδασκαλία*. Αθήνα: Gutenberg. Σελ. 13.

Για να πραγματοποιηθεί η επικοινωνία χρειάζεται τουλάχιστον ένας πομπός, ένας δέκτης και η χρήση ενός κοινού κώδικα επικοινωνίας. Στη πραγματολογία μας αφορά ο γλωσσικός κώδικας επικοινωνίας που χρησιμοποιείται. Η όλη διαδικασία είναι ένα σύνολο ενεργειών κωδικοποίησης και αποκωδικοποίησης ενός μηνύματος. Η πραγματολογία σχετίζεται με τη σημασία που έχει μια λέξη ή μια έκφραση για τον ομιλητή και πως την εκλαμβάνει ο ακροατής⁵³.

Απαραίτητη προϋπόθεση για τη πραγματολογική γνώση είναι η γνώση του περικειμένου το οποίο είναι το περιβάλλον της γλωσσικής χρήσης και διακρίνεται σε : γλωσσικό περικείμενο, σε περιστασιακό περικείμενο και γνωστικό περικείμενο. Το γλωσσικό περικείμενο περιλαμβάνει τα συμφραζόμενα μιας επικοινωνίας, το περιστασιακό περικείμενο περιλαμβάνει τις γνώσεις των συνομιλητών σχετικά με χρονικά και τοπικά δεδομένα που αφορούν συγκεκριμένη συνομιλία και το γνωστικό περικείμενο περιλαμβάνει τις γνώσεις του κάθε ομιλητή που σχετίζονται με την προσωπικότητα και την ταυτότητα του συνομιλητή⁵⁴.

2.7.1. Προβλήματα που δυσχεραίνουν την επικοινωνία

Κατά την επικοινωνία μπορεί να προκύψουν κάποια προβλήματα τα οποία δυσχεραίνουν την όλη διαδικασία. Αυτά είναι⁵⁵ : η ύπαρξη συμφραζομένων και αμφισημιών, το πρόβλημα της αναφοράς, της κυριολεκτικής χρήσης μιας λέξης, της μη ρητής επικοινωνίας και της μη επικοινωνιακής πρόθεσης του ομιλητή.

- Ένα σημαντικό θέμα είναι **η ύπαρξη αμφισημιών**. Η πραγματολογία σχετίζεται άμεσα με τη συμφραστική σημασία μιας λέξης καθώς τα συμφραζόμενα σε μια φράση καθορίζουν το νόημα της όπως και το νόημα των λέξεων και βοηθούν στην επίλυση νοηματικών ασαφειών. Γι' αυτό το λόγο στην πραγματολογία το ενδιαφέρον επικεντρώνεται στις σημασιολογικές αμφισημίες και όχι στις γραμματικές. Η άρση τέτοιων αμφισημιών γίνεται

⁵³ <http://www1.aegean.gr/social-anthropology/canakis/Courses%20offered/Pragmatics/Yule-chapter%201.htm>

⁵⁵ Μπέλλα, Σ.(2015). *Πραγματολογία, Από τη γλωσσική επικοινωνία στη γλωσσική διδασκαλία*. Αθήνα: Gutenberg. Σελ. 25.

λαμβάνοντας υπόψη τα συμφραζόμενα πριν και μετά την αμφισημία. Τα συμφραζόμενα αλλιώς ονομάζονται γλωσσικό περικείμενο ή γλωσσικό περιβάλλον. Σε περιπτώσεις όπου στην επικοινωνία υπάρχει κάποια λέξη ή έκφραση η οποία μπορεί να έχει διαφορετικές χρήσεις και σημασίες, είναι απαραίτητο να λαμβάνονται υπόψη τα συμφραζόμενα της πρότασης προκειμένου να κατανοήσει ο ακροατής το νόημα και να μπορέσει να πραγματοποιηθεί η επικοινωνία. Για παράδειγμα η λέξη «καλάμι» έχει τις παρακάτω σημασίες⁵⁶:

- (βοτανική) πολυετές υδροχαρές φυτό με ευλύγιστο κούφιο ξυλώδη βλαστό που φτάνει σε μεγάλο ύψος
- ο ξυλώδης αποξηραμένος βλαστός αυτού του φυτού ή ένα κομμάτι του
- μακρύ και λεπτό κυλινδρικό σύνεργο του ψαρέματος, από την άκρη του οποίου κρέμεται μια πετονιά
- (ανατομία) (λαϊκότροπο) το μπροστινό μέρος του οστού της κνήμης
- (αργκό): αυτοσχέδιος αργιλές που φέρει καλάμι αντί μαρκούτσι
- Επίσης υπάρχει και η φράση «έχει καβαλήσει το καλάμι» η οποία σημαίνει ότι κάποιος έχει πολύ μεγάλη ιδέα για τον εαυτό του.

Έτσι λοιπόν στις φράσεις «πήρε το καλάμι για να ψαρέψει» και «καθάρισε το χωράφι και έκοψε όλα καλάμια» τα συμφραζόμενα είναι αυτά που καθορίζουν την ερμηνεία της λέξης «καλάμι».

Βέβαια αμφίσημη δεν είναι πάντα μια λέξη καθώς υπάρχει περίπτωση να υπάρχουν αμφίσημες προτάσεις. Για παράδειγμα η φράση «Η Μαρία έφυγε από το σπίτι» μπορεί να σημαίνει ότι άλλαξε σπίτι, βγήκε έξω από το σπίτι για κάποια δουλεία, πήγε ταξίδι. Σε τέτοια περίπτωση πρέπει να λαμβάνονται υπόψη οι φράσεις πριν και μετά τη συγκεκριμένη καθώς και κάποιες φορές το σύνολο του κειμένου.

- **Το πρόβλημα της αναφοράς.** Σε κάποιες προτάσεις ή εκφράσεις είναι δύσκολο να προσδιορίσουμε ένα συγκεκριμένο υποκείμενο ή αντικείμενο επειδή λείπουν οι σχετικές πληροφορίες. Για παράδειγμα στην πρόταση «Ο

⁵⁶ <https://el.wiktionary.org/wiki/καλάμι>

παπός του χωριού» δεν προσδιορίζεται ποιο είναι το χωριό ή το όνομα του παπά άρα αν δεν υπάρχουν συμφραζόμενα προκύπτει πρόβλημα αναφοράς.

- **Πρόβλημα της μη κυριολεκτικής χρήσης μιας λέξης.** Το πρόβλημα αυτό προκύπτει όταν μια λέξη δε χρησιμοποιείται με την κυριολεκτική της σημασία. Π.χ. τέρας μαθήσεως, ανυπόμονη καρδιά, αβάσταχτος πόνος, βασανισμένη ζωή. Σε τέτοιες περιπτώσεις παρουσιάζονται προβλήματα κατά τη μηχανική μετάφραση και κάποιες φορές δεν αποδίδεται πλήρως το νόημα της φράσης.
- **Το πρόβλημα της μη ρητής επικοινωνίας.** Πολλές φορές κατά την επικοινωνία το μήνυμα το οποίο θέλει να μεταδώσει ο πομπός το εκφράζει με περισσότερα λόγια από ότι πρέπει. Για παράδειγμα η φράση «θα ήθελα οπωσδήποτε κάποια στιγμή , αν μπορείς να περάσεις από το σπίτι μου» θα μπορούσε να γραφτεί : «θα ήθελα να περάσεις από το σπίτι».
- **Το πρόβλημα της επικοινωνιακής πρόθεσης του ομιλητή.** Ο πομπός ενός μηνύματος μπορεί να θέλει να εκφράσει μια επιθυμία, μια υπόσχεση, μια απειλή κ.α. . Πολλές φορές κάποιες προτάσεις μπορεί να ερμηνευτούν με ορισμένο τρόπο από τον ακροατή. Βέβαια μια φράση δεν μπορεί να ερμηνευτεί το ίδιο βάσει του παραδοσιακού επικοινωνιακού μοντέλου.

2.7.2. Γλωσσικές προσεγγίσεις

Η πραγματολογία περιλαμβάνει κάποιες γλωσσικές προσεγγίσεις όπως η «θεωρία των γλωσσικών πράξεων», «η θεωρία της ευγένειας» και «η θεωρία των συνομιλιακών υπονοημάτων».

- **Οι γλωσσικές πράξεις** είναι πράξεις οι οποίες πραγματοποιούνται μέσω του λόγου όπως χαιρετισμός, απειλή, παράκληση, συμβουλή, εντολή και αποτελούνται από εκφωνήματα τα οποία έχουν επιτελεστικό χαρακτήρα.

- **Τα συνομιλιακά υπονοήματα:** σχετίζονται με το 3 γεγονός ότι πολλές φορές σε μια επικοινωνία υπάρχει διαφορά μεταξύ αυτού που λέγεται και αυτού που εννοείται. Για παράδειγμα έστω ότι έχουμε δύο συνομιλητές Α,Β και το παρακάτω μεταξύ τους διάλογο:
Α. Θέλεις να πάμε για φαγητό;
Β. Έχω πονοκέφαλο.
Λόγω των συμφραζομένων η απάντηση του Β έστω και αν μην είναι ρητά αρνητική, υπονοεί άρνηση. Μπορεί εκ πρώτης όψεως η απάντηση να φαίνεται άσχετη, όμως η σημασία των λέξεων καθορίζεται από τα συμφραζόμενα όλων των συμμετεχόντων στην επικοινωνία ομιλητή και ακροατή.
- **Το φαινόμενο της γλωσσικής ευγένειας** σχετίζεται με την αμφίδρομη επικοινωνία μεταξύ πομπού και δέκτη. Η ευγενική γλωσσική συμπεριφορά είναι γενικώς μια αποδεκτή συμπεριφορά. Μπορεί να γίνεται για διάφορους σκοπούς όπως παράκληση, ως ένδειξη σεβασμού στο πρόσωπο του συνομιλητή, λόγω της ιδιοσυγκρασίας του ομιλητή κ.α.

2.7.3. Ο ρόλος της Πραγματολογίας στην ανάκτηση της πληροφορίας

Στην πραγματολογία δίνεται μεγάλη σημασία στην συνδετικότητα και το νοηματικό συσχετισμό των προτάσεων ενός κειμένου. Γι' αυτό το λόγο η πραγματολογική ανάλυση είναι πολύ χρήσιμη σε διάφορες εφαρμογές ανάκτησης πληροφοριών όπως η αυτόματη περίληψη κειμένων, το word sense analysis, η αυτόματη κατηγοριοποίηση κειμένου, τη μηχανική μετάφραση κ.α. .

2.7.3.1 Αυτόματη περίληψη κειμένου

Η αυτόματη περίληψη είναι παρά πολύ χρήσιμη σε διάφορες περιπτώσεις όπως δημοσιογραφικές αναφορές, νομικά κείμενα, περιλήψεις μεγάλων σε έκταση κειμένων κ.α.

Το αυτόματο σύστημα περίληψης πρέπει να είναι σχεδιασμένο με τέτοιο τρόπο ώστε να ικανοποιεί τις απαιτήσεις του χρήστη και το επιθυμητό ύφος της περίληψης.

Η αυτόματη περίληψη περιλαμβάνει δύο βασικούς τύπους : την εξαγωγή (extraction) και την αφαίρεση (abstraction). Στην εξαγωγή το περιεχόμενο της περίληψης εξάγεται από το αρχικό κείμενο χωρίς να υπάρχουν τμήματα τα οποία δεν εμφανίζονται στο πρωτότυπο κείμενο. Αντίθετα στην περίληψη κάνοντας χρήση την τεχνική της αφαίρεσης περιλαμβάνονται και κάποια τμήματα που δεν εμφανίζονται στο αρχικό κείμενο αυτούσια.

Υπάρχουν διάφορες εφαρμογές αυτόματης περίληψης οι οποίες συνήθως χρησιμοποιούν τεχνικές μηχανικής μάθησης, γλωσσολογικές τεχνικές επεξεργασίας φυσικής γλώσσας. Εντοπίζουν τις λέξεις κλειδιά σ' ένα κείμενο, προσπαθούν να εντοπίσουν τις αντωνυμίες και να καταλάβουν που αναφέρονται και αφαιρούν τις λέξεις και φράσεις που θεωρούνται περιττές όπως π.χ. τους συνδέσμους. Στο τέλος εκτελούνται κάποιες διορθώσεις που αφορούν τη συνοχή των προτάσεων έτσι ώστε το κείμενο να είναι περισσότερο ευανάγνωστο.

Πολύ σημαντικό είναι να εντοπιστούν οι σημαντικότερες για το νόημα προτάσεις σ' ένα κείμενο και γι' αυτό το λόγο υπάρχουν κανόνες αξιολόγησης του βάρους για κάθε πρόταση. Στους κανόνες αυτούς συμπεριλαμβάνεται η ομοιότητα με τον τίτλο, η ύπαρξη αντωνυμιών, η ύπαρξη λέξεων με μεγάλη συχνότητα στο κείμενο κ.α. .

Στο διαδίκτυο υπάρχουν πολλές εφαρμογές αυτόματης περίληψης σε διάφορες γλώσσες, για κείμενα και email. Κάποια από αυτά είναι το resoomer⁵⁷ , το smmry⁵⁸ , το copernic⁵⁹ κ.α.

2.7.3.2. Αποσαφήνιση νοήματος διαφορούμενων λέξεων (word sense disambiguation)

Ένα μέρος της ανάκτησης της πληροφορίας αφορά την αναζήτηση πληροφορίας σε κείμενο ή σε συλλογές κειμένων. Η αναζήτηση αυτού του τύπου πολλές φορές μπορεί

⁵⁷ <https://resoomer.com/en/>

⁵⁸ <https://smmry.com/>

⁵⁹ <https://www.copernic.com/en/products/summarizer/>

να είναι πιο εξειδικευμένη και να αφορά συγκεκριμένη ορολογία. Την αναζήτηση αυτή την δυσκολεύει το γεγονός ότι υπάρχουν λέξεις οι οποίες μπορεί να έχουν παραπάνω από μια σημασίες. Για τον ομιλητή είναι εύκολο να κατανοήσει το πραγματικό νόημα μιας διφορούμενης λέξης ή πρότασης, όμως δεν είναι το ίδιο για τον υπολογιστή και κατά συνέπεια για τη μηχανική μετάφραση.

Η αποσαφήνιση σημασιολογικών αμφισημιών (WSD) μπορεί να επιτευχθεί μέσω των κατάλληλων υπολογιστικών μοντέλων. Απαραίτητη είναι η βοήθεια των κατάλληλων σωμάτων κειμένων με τη βοήθεια των οποίων φαίνονται οι διάφορες σημασίες της κάθε λέξης και μπορεί να υπολογιστεί η συχνότητα της κάθε σημασίας.

Η αποσαφήνιση σημασιολογικών αμφισημιών βρίσκει πεδίο εφαρμογής σε διάφορους τομείς της γλωσσολογικής έρευνας. Για παράδειγμα στη μηχανική μετάφραση καθώς και σε περιπτώσεις λέξεων με πολλαπλές σημασίες χρειάζεται να λαμβάνουμε υπόψη τα συμφραζόμενα. Επίσης λαμβάνονται υπόψη στην κατασκευή μηχανών αναζήτησης επειδή σε περιπτώσεις λέξεων με πολλαπλά νοήματα λαμβάνονται πρώτα υπόψη η πιο δημοφιλείς ερμηνείες. Επίσης είναι χρήσιμη και σε διαφορές εφαρμογές εξόρυξης γνώσης⁶⁰.

Υπάρχουν τρία είδη προσεγγίσεων αποσαφήνισης σημασιολογικών αμφισημιών : Αυτές είναι : η knowledge based προσέγγιση, η προσέγγιση μέσω επιβλεπόμενης μάθησης και τέλος η προσέγγιση μέσω μη επιβλεπόμενης μάθησης.

2.7.3.2.1 Knowledge based προσέγγιση

Η knowledge- based προσέγγιση βασίζεται στη χρήση κατάλληλων ηλεκτρονικών λεξικών όπως το Wordnet. Υπάρχουν 4 περιπτώσεις knowledge- based προσέγγισης.

- Η πρώτη χρησιμοποιεί τον αλγόριθμο LESK ο οποίος βασίζεται στη φιλοσοφία ότι σε μια πρόταση ή ένα κείμενο οι λέξεις βασίζονται πάνω σ' ένα συγκεκριμένο θέμα. Ο αλγόριθμος LESK συγκρίνει για μια αμφίσημη λέξη τους διαθέσιμους ορισμούς της στο ηλεκτρονικό λεξικό με τα συμφραζόμενα.

⁶⁰ Navigli, R. (2009). Word sense disambiguation: A survey. *ACM computing surveys (CSUR)*, 41(2), 10.

- Στη δεύτερη περίπτωση χρησιμοποιούνται μέτρα σημασιολογικής ομοιότητας από τα οποία προκύπτει μια τιμή που καθορίζει το βαθμό σημασιολογικής ομοιότητας μιας λέξης σ' ένα κείμενο με τις διάφορες σημασίες που υπάρχουν σε ένα κατάλληλο ηλεκτρονικό λεξικό.
- Ένας άλλος τρόπος είναι με την εύρεση πιθανών σημασιολογικών σχέσεων που μπορεί να υπάρχουν μεταξύ των λέξεων με σκοπό να ερευνηθεί η συχνότητα εμφάνισης αυτών των ζευγών λέξεων σε σώματα κειμένου. Π.χ. χορδή - κιθάρα, δέντρο – δάσος, ρούχο – ύφασμα.
- Ο τελευταίος τρόπος είναι ο εξής: για κάθε πιθανό νόημα που μπορεί να έχει μια λέξη υπολογίζεται η συχνότητα εμφάνισης της σε σώματα κειμένου. Σε αυτή την περίπτωση επιλέγονται τα νοήματα με τη μεγαλύτερη συχνότητα εμφάνισης.

2.7.3.2.2. Μέσω επιβλεπόμενης μάθησης

Στην επιβλεπόμενη μάθηση στόχος είναι η αποσαφήνιση της σημασιολογικής αμφισημίας με βάση κάποια δεδομένα εκπαίδευσης. Τα δεδομένα αυτά αποτελούνται από παραδείγματα τα οποία μπορεί αν προέρχονται από τα κατάλληλα λεξικά και τα χρησιμοποιεί ο ταξινομητής προκειμένου να εκπαιδευτεί. Για την άρση αμφισημιών με επιβλεπόμενη μάθηση γίνεται μεταξύ άλλων χρήση :

- Δέντρων απόφασης και λιστών απόφασης όπου στην ουσία αποτελούνται από δομές του τύπου «if – then – else»
- Του αλγορίθμου Naïve Bayes. Ο αλγόριθμος Naïve Bayes είναι πιθανοτικός και κατάλληλος για μεγάλα σύνολα δεδομένων. Ο βασικός τύπος του αλγορίθμου είναι:

$$P(S_i | f_1, f_2, \dots, f_n) = \frac{P(f_1, f_2, \dots, f_n | S_i)P(S_i)}{P(f_1, f_2, \dots, f_n)}$$

, όπου S_i είναι οι διάφορες σημασίες μιας λέξης και $f_1, f_2, \dots, f_n$ είναι τα συμφραζόμενα του κειμένου

- Ο αλγόριθμος SVM ο οποίος διαχωρίζει σε δύο ομάδες τα δεδομένα (Θετικά-Αρνητικά παραδείγματα)

2.7.3.2.3. Η μη επιβλεπόμενη μάθηση

Στην μη επιβλεπόμενη μάθηση το σύστημα πρέπει από μόνο του να βγάλει συσχετίσεις από το σύνολο των δεδομένων για τα οποία δεν είναι γνωστή η κλάση . Χρησιμοποιείται για μεγάλο όγκο δεδομένων όπου οι άλλες μέθοδοι είναι περισσότερο δαπανηρές.

2.8. Κοινωνιογλωσσολογία

Κοινωνιογλωσσολογία είναι ο υποκλάδος της γλωσσολογίας που μελετάει τη σχέση που έχει η κοινωνία με τη διαμόρφωση της γλώσσας των μελών ενός κοινωνικού συνόλου. Ειδικότερα επιδιώκει να εντοπίσει τις επιρροές που έχουν οι γλωσσικές διάλεκτοι με το κοινωνικό περιβάλλον στο οποίο λαμβάνουν χώρα καθώς επίσης και τις ομοιότητες, τις διαφορές και σχέσεις με άλλες διαλέκτους και γλώσσες με τις οποίες έρχεται σε επαφή η γλώσσα ενός κοινωνικού συνόλου. Στην κοινωνιογλωσσολογία το ενδιαφέρον επικεντρώνεται όχι στο γραπτό αλλά στον προφορικό λόγο καθώς είναι περισσότερο αυθόρμητος και αυθεντικός και μαρτυρά στοιχεία για την ταυτότητα του ομιλούντος.

Σχετικός είναι και ο όρος «κοινωνιολογία της γλώσσας» η οποία μελετάει τη σχέση που έχει η γλώσσα με στοιχεία που αφορούν την ταυτότητα και την κοινωνική υπόσταση του ομιλούντος όπως: η κοινωνική ομάδα στην οποία ανήκει, το μορφωτικό επίπεδο, την ηλικία και το φύλο του. Στην κοινωνιογλωσσολογία το ενδιαφέρον δεν επικεντρώνεται στις γλωσσικές δομές αλλά στη μελέτη διαφόρων γλωσσικών φαινομένων που προκύπτουν ανάλογα το κοινωνικό περιβάλλον.

2.8.1. Η Γλωσσική Ποικιλία

Ένα από τα χαρακτηριστικά της γλώσσας το οποίο προκύπτει μέσα από τη διαφορετικότητα των ομιλητών είναι η γλωσσική ποικιλία. Πολλές γλώσσες περιλαμβάνουν γλωσσικές ποικιλίες οι οποίες τοποθετούνται μεταξύ κοινωνικών ομάδων ή γεωγραφικών περιοχών. Πολλές φορές στοιχεία, λέξεις και εκφράσεις από γλωσσικές ποικιλίες, μπορεί να εμπλουτίζουν το λεξιλόγιο, τη γραμματική και τη σύνταξη μιας γλώσσας .

Πολλές φορές, χρησιμοποιούμε διαφορετικό λεξιλόγιο, εκφράσεις και σύνταξη για να απευθυνθούμε σ' έναν ενήλικα , σ' έναν έφηβο, σ' ένα παιδί, σ' έναν ηλικιωμένο. Ομοίως με διαφορετικό τρόπο και χρησιμοποιώντας απλές λέξεις και εκφράσεις, πρέπει να εκφράσουμε ένα αίτημα μας σε άτομα διαφορετικού μορφωτικού επιπέδου. Επίσης ανάλογα τις περιστάσεις θα πρέπει να χρησιμοποιήσουμε διαφορετικό τρόπο για να διατυπώσουμε ένα αίτημά μας ή για να εκφράσουμε την άποψη μας .

Η γλωσσική ποικιλία διακρίνεται σε γεωγραφική και κοινωνική. Από τη γεωγραφία μπορεί να προκύπτουν διαφορές στους ομιλητές από τις οποίες προκύπτουν οι διάλεκτοι μιας γλώσσας και τα γλωσσικά ιδιώματα. Η διάλεκτος στην ουσία είναι μια παραλλαγή μιας γλώσσας με πολλές ομοιότητες με αυτή όσον αφορά το λεξιλόγιο και τη γραμματική, όμως δεν αποτελεί την εθνική γλώσσα μιας χώρας. Αξιοσημείωτη είναι η φράση του γλωσσολόγου Μαξ Βάινραϊχ (Max Weinreich): Γλώσσα είναι μια διάλεκτος με στρατό και ναυτικό⁶¹. Χαρακτηριστικό παράδειγμα είναι η καστιλιάνικη γλώσσα στην Ισπανία η οποία είναι κατανοητή σε μεγάλο βαθμό από διαλέκτους όπως τα Αστουριανά, τα Αραγονέζικα, τα Γαλικιανά, τα Λαντίνο, Λεονές, Μιραντές, αλλά και την πορτογαλική γλώσσα. Το ίδιο ισχύει και με τις διαλέκτους της ελληνικής γλώσσας όπως για παράδειγμα αυτές που προέρχονται από τη Μικρά Ασία (Καππαδόκιη, Ποντιακή), την Κυπριακή και την Κρητική.

Επίσης στη γεωγραφική γλωσσική ποικιλία ανήκουν και τα γλωσσικά ιδιώματα τα οποία είναι γλωσσικές παραλλαγές μιας γλώσσας σε μορφολογικό ή φωνητικό επίπεδο. Στην ελληνική γλώσσα χαρακτηριστικά παραδείγματα είναι τα γλωσσικά ιδιώματα των Δωδεκανήσων όπου παρατηρείται σίγηση των συμφώνων [ν], [γ] και

⁶¹ <https://el.wikipedia.org/wiki/Διάλεκτος>.

[δ] ανάμεσα από δύο φωνήεντα σε διάφορες λέξεις. Έτσι η λέξη σίδερο αντιστοιχεί στο ιδίωμα σίερο κτλ.

Η κοινωνική γλωσσική ποικιλία αφορά τους διάφορους τρόπους με τους οποίους μια γλώσσα χρησιμοποιείται από διάφορες κοινωνικές ομάδες ή για συγκεκριμένο σκοπό. Η κοινωνική γλωσσική ποικιλία αφορά τη φωνολογικές και μορφολογικές διαφοροποιήσεις που σχετίζονται με παράγοντες όπως η ηλικία, το μορφωτικό επίπεδο, το κοινωνικό status κ.α.

Υπάρχουν περιπτώσεις και συνθηματικών γλωσσών όπως οι αργκό οι οποίες ομιλούνται από διάφορες κοινωνικές ομάδες όπως ανθρώπους της νέας γενιάς, περιθωριακές ομάδες ατόμων κ.α. Οι αργκό είναι αποτέλεσμα της επιθυμίας κοινωνικών ομάδων να κατασκευάσουν τη δική τους διάλεκτο μέσα από κωδικοποιημένες λέξεις. Για παράδειγμα σε νέους ανθρώπους η λέξη «βράζω» σημαίνει είμαι ευερέθιστος, η λέξη «φορτώνω» σημαίνει αρχίζω να νευριάζω και η φράση «δεν παλεύεται» σημαίνει δεν αντέχεται.

2.8.2. Η Γλωσσική Επικοινωνία

Προκειμένου να προταθούν και να ερευνηθούν οι διάφορες λειτουργίες μιας γλώσσας σε διάφορες κοινωνικές ομάδες, πολλοί γλωσσολόγοι έχουν προτείνει διάφορα επικοινωνιακά σχήματα. Στα περισσότερα επικοινωνιακά σχήματα που έχουν προταθεί περιλαμβάνουν πομπό, δέκτη, το περιεχόμενο του μηνύματος, το δίαυλο επικοινωνίας και τη μετάδοση του μηνύματος.

Ο πρώτος γλωσσολόγος που αναφέρθηκε σε συστήματα επικοινωνίας ήταν ο Ferdinand de Saussure ο οποίος διέκρινε τρεις όψεις στα συστήματα επικοινωνίας και αναφέρθηκε στη διάκριση μεταξύ «λόγου» και «ομιλίας». Λόγος είναι η ικανότητα ενός ανθρώπου να επικοινωνεί χρησιμοποιώντας τον έναρθο λόγο και επικοινωνία είναι η χρήση του συστήματος που κάνει ένας ομιλητής προκειμένου να επικοινωνήσει χρησιμοποιώντας τη γλώσσα.

2.8.3. Γλωσσικές Ανισότητες

Σύμφωνα με τη σύγχρονη γλωσσολογία υπάρχει ισότητα μεταξύ όλων των διαλέκτων και καμία γλώσσα ή διάλεκτος δεν είναι ανώτερη από κάποια άλλη. Μια γλώσσα η διάλεκτος μπορεί να κυριαρχήσει και να γίνει η επίσημη γλώσσα ενός κράτους για καθαρά οικονομικούς, πολιτικούς και στρατιωτικούς λόγους ή γιατί ομιλείται από την πλειοψηφία των πολιτών μιας χώρας. Η γλώσσα αυτή λέγεται πρότυπη γλώσσα ή νόρμα.

Παρόλα αυτά από πολλούς μη γλωσσολόγους πολλές φορές έχει εκφραστεί η άποψη ότι κάποια γλώσσα ή διάλεκτος είναι ανώτερη από κάποια άλλη. Το γεγονός αυτό αποτελεί ένδειξη γλωσσικής ανισότητας η οποία πολλές φορές είναι αποτέλεσμα κοινωνικών ανισοτήτων. Για παράδειγμα μια διάλεκτος που ομιλείται σε πιο οικονομικά υποβαθμισμένες περιοχές ενός κράτους είναι δύσκολο να καθιερωθεί ως η επίσημη γλώσσα του κράτους.

Οι γλωσσικές ανισότητες προκαλούν προβλήματα και στην εκπαίδευση καθώς κατά πολλούς η σχολική αποτυχία σε μεγάλο βαθμό οφείλεται στην ύπαρξη κοινωνικών ανισοτήτων καθώς σε πολλές περιπτώσεις μαθητές που προέρχονται από οικογένειες της εργατικής τάξης έχουν χαμηλότερες επιδόσεις και γραμματικό συντακτικό υπόβαθρο σε σχέση με μαθητές που προέρχονται από οικογένειες με υψηλό μορφωτικό επίπεδο ή από την αστική τάξη.

2.8.4. Social Information Retrieval

Η κοινωνική ανάκτηση της πληροφορίας διαφέρει από την παραδοσιακή προσέγγιση στο γεγονός ότι το περιεχόμενο της αναζήτησης προέρχεται μέσα από χρήστες μιας κοινότητας⁶². Υπάρχει επίσης και η δυνατότητα άντλησης εικόνων, βίντεο, κειμένων και αρχείων και από κοινωνικά δίκτυα όπως facebook, twitter, instagram κτλ.⁶³.

⁶² Horowitz, D., & Kamvar, S. D. (2010, April). The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web* (pp. 431-440). ACM.

⁶³ https://en.wikipedia.org/wiki/Social_search

Η ιδέα της κοινωνικής ανάκτησης της πληροφορίας προέρχεται από το γεγονός ότι τα νέα από άτομα τα οποία γνωρίζουμε και εμπιστευόμαστε είναι πιο έγκυρα⁶⁴. Έτσι λοιπόν δίνεται η δυνατότητα για την ανακάλυψη ταξιδιωτικών προορισμών, χώρων διασκέδασης, μουσικής, πώλησης προϊόντων που προτείνονται από ομάδες ατόμων που εμπιστευόμαστε.

Η κοινωνική ανάκτηση της πληροφορίας βασίζεται στα κοινωνικά μέσα δικτύωσης και στους σχολιασμούς των χρηστών τους μέσα από διαλόγους και hashtags. Οι πληροφορίες αυτές μπορούν να δώσουν απαντήσεις στα διάφορα ερωτήματα και να προτείνουν ιστοτόπους καθώς μέσα από τα σχόλια υπάρχει η δυνατότητα παρουσίασης με συνοπτικό τρόπο του περιεχομένου μιας ιστοσελίδας και της δημοφιλίας της⁶⁵ σε μια κοινότητα ανθρώπων.

2.8.4.1. Κοινωνικές μηχανές αναζήτησης

Οι κοινωνικές μηχανές αναζήτησης στην ουσία δέχονται ερωτήσεις και παρέχουν τις κατάλληλες απαντήσεις μέσω της σύνδεσης τους με τον κατάλληλο πόρο . Για την κατασκευή τους δεν χρειάζονται η αλγόριθμοι κατάταξης των παραδοσιακών μηχανών αναζήτησης . Στην μηχανή αναζήτησης «Aardvark» γίνεται χρήση crawler , ευρετηρίου και αναλυτή ερωτημάτων (query analyzer) προκειμένου να αναλυθούν τα ερωτήματα των χρηστών και να εκτιμηθούν οι πληροφοριακές ανάγκες τους. Επίσης μια κοινωνική μηχανή αναζήτησης περιλαμβάνει ακόμη μια συνάρτηση βαθμολόγησης (ranking function) και την κατάλληλη διεπαφή για τον χρήστη⁶⁶.

Πολλές εταιρείες όπως η Google και η Bing έχουν ασχοληθεί με την κοινωνική αναζήτηση . Η εφαρμογή «google discover» ή «google feed» σαν σκοπό έχει τη συγκέντρωση των πιο δημοφιλών πληροφοριών που προέρχονται από κοινωνικά μέσα. Μια άλλη εφαρμογή της google η «Google Social Search» δίνει τη δυνατότητα στο χρήστη να επιλέξει μέσα από έξι διαφορετικά κοινωνικά δίκτυα (facebook,

⁶⁴ Horowitz, D., & Kamvar, S. D. (2010, April). The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web* (pp. 431-440). ACM.

⁶⁵ Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., & Su, Z. (2007, May). Optimizing web search using social annotations. In *Proceedings of the 16th international conference on World Wide Web* (pp. 501-510). ACM.

⁶⁶ Horowitz, D., & Kamvar, S. D. (2010, April). The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web* (pp. 431-440). ACM.

twitter, linkedin, instagram, pinterest, google+) προκειμένου να αντληθούν οι πληροφορίες για να δοθούν απαντήσεις στα ερωτήματα τους. Άλλη σημαντική εφαρμογή είναι το «Social Searcher» το οποίο δίνει τη δυνατότητα να κρατάει ενήμερους τους χρήστες τους μέσα από ηλεκτρονικά μηνύματα σχετικά με τους πιο δημοφιλείς χρήστες και τα αρχεία που αναρτούν⁶⁷

2.9. Η Φιλοσοφία της Γλώσσας

Η φιλοσοφία της γλώσσας είναι το πεδίο της φιλοσοφίας που αφορά την εξήγηση της σημασίας της γλώσσας, του νοήματος των όρων της και προσπαθεί να δώσει απαντήσεις στο ερώτημα από τι αποτελείται και τι περιέχει η κατανόηση μιας γλώσσας και η γλωσσική επικοινωνία. Ασχολείται με άλλα λόγια με το περιεχόμενο και τη φύση του νοήματος και όχι με το ίδιο το νόημα, καθώς δεν αφορά τη φιλοσοφία της γλώσσας τι σημαίνει μια έκφραση ή μια πρόταση μιας οποιασδήποτε γλώσσας.

Σύμφωνα με κάποιους φιλοσόφους η γνώση του νοήματος μιας λέξης σημαίνει και γνώση των κανόνων χρήσης της λέξης αυτής⁶⁸. Εκτός από τις λέξεις, δομικούς κανόνες έχουν και οι φράσεις.

Σύμφωνα με πολλούς φιλοσόφους υπάρχει μια κοινή γλώσσα για όλους η οποία μετασχηματίζεται στις ομιλούσες γλώσσες. Αυτή η γλώσσα θεωρείται ιδεώδης - καθαρή, χωρίς τις ασάφειες και αμφισημίες οι οποίες υπάρχουν στις άλλες γλώσσες. Πολύ μεγάλης σημασίας είναι η ύπαρξη μιας κοινής γλώσσας η οποία μπορεί να περιγράψει τη λογική μορφή μιας πρότασης. Ο φιλόσοφος Frege προσπάθησε να κατασκευάσει μια τέτοιου είδους γλώσσα αυτή με τη χρήση διαφόρων συμβολισμών.

⁶⁷ <https://www.social-searcher.com/>

⁶⁸ Searle, J. R., & Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.

2.9.1. Το περιεχόμενο της φιλοσοφίας της γλώσσας

Η φιλοσοφία της γλώσσας αποτελεί έναν πολύ ευρύ κλάδο αλλά χωρίς ξεκάθαρο εννοιολογικό πεδίο καθώς με στον όρο αυτό περιλαμβάνονται πολλά διαφορετικά αντικείμενα εκ των οποίων κάποια σχετίζονται και με άλλους ξεχωριστούς κλάδους της γλωσσολογίας όπως η ψυχολογολογία, η νευρογλωσσολογία και η γενική γλωσσολογία. Κάποια από τα θέματα που εξετάζει η φιλοσοφία της γλώσσας είναι⁶⁹:

- Η φύση της γλώσσας δηλαδή τι είναι η γλώσσα, ποιους σκοπούς εξυπηρετεί καθώς και την προέλευσή της
- Οι σκέψεις και αντιλήψεις φιλοσόφων για τη γλώσσα,
- Οι διάφοροι στοχασμοί οι οποίοι σχετίζονται με τη γλώσσα και τους οποίους προσπαθούν να εξηγήσουν οι φιλόσοφοι.
- Διάφορες συζητήσεις σχετικά με τεχνικά θέματα της γλώσσας όπως π.χ. η σχέση ενός κύριου ονόματος με την κυριολεκτική σημασία του.
- Η φιλοσοφία της καθημερινής γλώσσας που μιλούν οι άνθρωποι, δηλαδή της καθομιλουμένης, η οποία δεν έχει τυπική μορφή.
- Η γενική γλωσσολογία, η οποία όμως θεωρείται από πολλούς ως ξεχωριστό αντικείμενο.
- Διάφορα ερωτήματα σχετικά με τη γλώσσα τα οποία δεν έχουν ακόμη απαντηθεί με ολοκληρωμένο τρόπο.
- Διάφορες αναφορές στους αρχαίους φιλοσόφους και τις απόψεις τους και στοχασμούς τους για τη γλώσσα.

2.9.2. Ιστορική αναδρομή

Η φιλοσοφία της γλώσσας έχει τις ρίζες της στην Αρχαία Ελλάδα. Βέβαια κατά την αρχαιότητα οι Έλληνες είχαν μια πιο διαφορετική προσέγγιση για τη γλώσσα η οποία όμως είναι χρήσιμη και σήμερα. Οι Αρχαίοι Έλληνες θεωρούσαν τον άνθρωπο ως «έλλογο ον» και σύμφωνα με τον Πλάτωνα κάθε φώνημα εκφράζει ένα συναίσθημα, ενώ μέσα από τα έργα του ασχολήθηκε με το πώς προσδιορίζονται τα ονόματα των πραγμάτων. Επίσης ο Αριστοτέλης κατηγοριοποίησε όλα τα πράγματα με βάση το

⁶⁹ Auroux, S.(2005). Η φιλοσοφία της γλώσσας. Αθήνα: Μεταίχμιο.

είδος και το γένος και ασχολήθηκε με τη λογική. Θεώρησε ότι ο άνθρωπος από πολύ μικρή ηλικία έχει τη δυνατότητα να σκέφτεται λογικά και όταν τεθεί ένας προβληματισμός, μπορεί να λυθεί μέσα από μια αλληλουχία σκέψεων.

Η κανονική φιλοσοφία της γλωσσολογίας προέρχεται από την Ινδική Φιλοσοφία κατά τη διάρκεια του Μεσαίωνα. Για τα νοήματα της γλώσσας ενδιαφέρθηκαν και οι φιλόσοφοι της Ευρώπης κατά τη διάρκεια του Μεσαίωνα και ορισμένοι από αυτούς θεωρούσαν τη λογική ως μια επιστήμη της γλώσσας. Επίσης αναλύθηκαν σε μεγάλο βαθμό τα φαινόμενα του απροσδιόριστου νοήματος (vagueness) και του διφορούμενου νοήματος (ambiguity) και αναπτύχθηκε η μελέτη των κατηγορηματικών λέξεων και των ιδιοτήτων τους.

Κατά την Αναγέννηση κάποιοι σημαντικοί γλωσσολόγοι οραματίστηκαν τη δημιουργία μιας φιλοσοφικής γλώσσας παγκοσμίου εμβέλειας η οποία θα συνέβαλε στην μείωση της ασυνεννοησίας που προκαλεί η διαφορετικότητα των διαφόρων γλωσσών.

Στα τέλη του 19^{ου} αιώνα υποστηρίχθηκε από τους ειδικούς ότι η γλώσσα πρέπει να παίζει σημαντικό ρόλο στην φιλοσοφία και ότι η φιλοσοφία θα πρέπει να ασχοληθεί με τη γλώσσα. Έτσι λοιπόν η φιλοσοφία της γλώσσας έγινε πολύ διαδεδομένη στα πλαίσια της αναλυτικής φιλοσοφίας και στις αρχές του 20^{ου} αιώνα πραγματοποιήθηκε γλωσσολογική στροφή καθώς δόθηκε μεγάλη έμφαση από τη φιλοσοφία και τους φιλοσόφους στη γλώσσα.

Στην αναλυτική φιλοσοφία που κυριάρχησε τον 20ο αιώνα, δίνεται έμφαση στην ανάλυση των πραγμάτων βασιζόμενη στη χρήση της λογικής, την ανάλυση των νοημάτων και σε μικρότερο βαθμό βασιζόμενη σε μαθηματικούς κανόνες και κανόνες της φυσικής. Η γλώσσα στην αναλυτική φιλοσοφία εξετάζεται ως ξεχωριστή θεωρία.

2.9.3. Οι λειτουργίες της γλώσσας

Η χρήση της γλώσσας δίνει τη δυνατότητα επικοινωνίας μεταξύ των ανθρώπων και αποτελεί μέσω της σκέψης μέσω του ανθρώπινου νου αλλά και της εξωτερίκευσής των σκέψεων και των ιδεών μέσω του γραπτού και του προφορικού λόγου. Από την αρχαιότητα απασχόλησε τους Έλληνες φιλοσόφους η σχέση μεταξύ σκέψης και

γλώσσας. Ο Πλάτωνας υποστήριξε ότι όταν σκέπτεται ένας άνθρωπος πραγματοποιεί διάλογο με τον εαυτό του και συνεπώς υπάρχει σχέση σκέψης και γλώσσας. Ο αρχαίος φιλόσοφος Γοργίας υποστήριξε ότι «ο λόγος προκύπτει από τις εντυπώσεις που κάνουνε πάνω μας τα αισθητά» Τον 17^ο αιώνα στη θεωρία των φιλόσοφων Leibniz και Hobbes διακρίνεται το γεγονός ότι μια λέξη είναι αποτέλεσμα έκφρασης της σκέψης⁷⁰.

Πολλοί άλλοι γλωσσολόγοι έχουν προσπαθήσει να διατυπώσουν όλες τις δυνατές λειτουργίες και χρήσεις της γλώσσας. Κάποιες χαρακτηριστικές περιπτώσεις λειτουργιών έχουν αναφερθεί οι ακόλουθες :

- Την εκφραστική λειτουργία μέσω της οποίας εκφράζονται τα ανθρώπινα συναισθήματα
- Συνδηλωτική λειτουργία προσανατολισμένη προς τον παραλήπτη (προστακτική)
- Επισημαντική λειτουργία (κραυγή ενός πουλιού προς σήμανση κινδύνου)
- Περιγραφική λειτουργία
- Μεταγλωσσική λειτουργία επικεντρωμένη στον κώδικα
- Ποιητική λειτουργία επικεντρωμένη στο μήνυμα
- Αναφορική λειτουργία η οποία αφορά τη μετάδοση των πληροφοριών

2.9.4. Σύνδεση της φιλοσοφίας της γλώσσας με την ανάκτηση της πληροφορίας

Λόγω του μεγάλου εύρους και θεματολογίας της φιλοσοφίας της γλώσσας θα μπορούσαμε να βρούμε αρκετά κοινά στοιχεία και εφαρμογές με την ανάκτηση της πληροφορίας. Σίγουρα η προσπάθεια του ανθρώπου να κατασκευάσει συστήματα τα οποία έχουν τη δυνατότητα να δίνουν απαντήσεις σε ερωτήματα ή να εξάγουν συμπεράσματα, φανερώνει ότι τα συστήματα αυτά διαθέτουν κάποια ευφυΐα. Μάλιστα έχει καθιερωθεί ως ξεχωριστός επιστημονικός κλάδος η Τεχνητή Νοημοσύνη η οποία περιλαμβάνει όλα τα συστήματα που προσπαθούν να μιμηθούν την ανθρώπινη νοημοσύνη και μέσω αυτής της δυνατότητας να καταφέρουν να επιλύσουν διάφορα προβλήματα.

⁷⁰ Auroux, S.(2005). Η φιλοσοφία της γλώσσας. Αθήνα: Μεταίχμιο. Σελ. 138.

Μια από τις εφαρμογές της ανάκτησης της πληροφορίας η οποία προσπαθεί να μιμηθεί την ανθρώπινη νοημοσύνη και να παρέχει απαντήσεις είναι η «Απάντηση Ερωτήσεων- Question Answering». Τα συστήματα αυτά παρέχουν απαντήσεις σε ερωτήσεις που κάνουν οι χρήστες μέσω κάποιας φυσικής γλώσσας. Απαραίτητη προϋπόθεση είναι η ύπαρξη κάποιας βάσης γνώσης από την οποία μπορεί το σύστημα να αντλεί τις απαντήσεις. Η βάση γνώσης μπορεί να αποτελείται από δομημένες ή αδόμητες συλλογές εγγράφων καθώς και από ιστοσελίδες του διαδικτύου.

Υπάρχουν και συστήματα ερωτήσεων – απαντήσεων που περιλαμβάνουν κλειστού τύπου ερωτήσεις και βρίσκουν εφαρμογή σε πιο εξειδικευμένα πεδία και επιστήμες (π.χ. ιατρική και μηχανολογία).

Δυνατότητα παροχής απαντήσεων σε πολλές περιπτώσεις παρέχει και το Web 3.0 του σημασιολογικού ιστού , καθώς μέσω διαφόρων μηχανών αναζήτησης μπορούν να δοθούν απαντήσεις σε διάφορες ερωτήσεις.

2.10. Η Ψυχολογία της Γλώσσας

Η Ψυχολογία της γλώσσας είναι ένας σχετικά νέος κλάδος της γλωσσολογίας ο οποίος καθιερώθηκε μετά το Β' Παγκόσμιο Πόλεμο αν και το αντικείμενο της είχε απασχολήσει ερευνητές και στα τέλη του 19^{ου} αιώνα.

Η ψυχογλωσσολογία περιλαμβάνει όλες τις ψυχολογικές διεργασίες που συμμετέχουν στη γλώσσα . Οι ψυχογλωσσολόγοι εξετάζουν την κατανόηση, την παραγωγή και τη μνήμη της γλώσσας⁷¹. Επιπλέον ασχολούνται με το πώς ο άνθρωπος αποκτάει τη γλώσσα και την αλληλεπίδραση της γλώσσας με άλλα ψυχολογικά συστήματα.

Η ομιλία ενός ανθρώπου εξωτερικεύει τις σκέψεις του και τον ψυχισμό του και άρα είναι ένας τρόπος πρόσβασης στο χώρο του ασυνείδητου ενός ανθρώπου. Μέσα από τη γλώσσα φανερώνονται ο συναισθηματικός κόσμος και οι επιθυμίες ενός ανθρώπου.⁷²

⁷¹ Harley.T.(2008). *Η Ψυχολογία της Γλώσσας*.Θεσσαλονίκη: Univercity Studio Press.

⁷² http://www.komvos.edu.gr/glwssa/odigos/thema_a5/a_5_thema.htm

2.10.1. Η Απόκτηση γλώσσας

Η εκμάθηση μιας γλώσσας είναι αποτελεσματικότερη όσο μικρότερη είναι η ηλικία του ανθρώπου καθώς με την πάροδο του χρόνου η ιδιότητα αυτή φθίνει. Έτσι λοιπόν υπάρχει μια συγκεκριμένη περίοδος της ζωής του ανθρώπου η οποία θεωρείται η ιδανική για την εκμάθηση γλώσσας, πράγμα το οποίο σημαίνει ότι αν κατά την περίοδο αυτή αν δεν αποκτηθούν γλωσσικές ικανότητες από έναν άνθρωπο, είναι δύσκολο έως αδύνατο να γίνει αυτό αργότερα. Αυτό συμβαίνει γιατί κατά την περίοδο αυτή αναπτύσσεται ο ανθρώπινος εγκέφαλος και τα δύο ημισφαίρια του εγκεφάλου εξειδικεύονται με διάφορες ιδιότητες και λειτουργίες.

2.10.2. Σχέση σκέψης – γλώσσας

Μεγάλο ενδιαφέρον υπάρχει από τους γλωσσοψυχολόγους για τη σχέση μεταξύ γλώσσας και σκέψης. Οι σημαντικότερες θεωρίες που έχουν διατυπωθεί είναι των συμπεριφοριστών αλλά και των εποικοδομιστών με κύριους εκπρόσωπους τους Piaget και Vygotsky.

Σύμφωνα με το συμπεριφορισμό η σκέψη είναι λόγος και μάλιστα τα παιδιά λένε φωναχτά τις σκέψεις τους. Κάποιοι ερευνητές στο παρελθόν μετά από πειράματα υποστήριξαν ότι εντόπισαν ηλεκτρική δραστηριότητα σε μυς του λαιμού όταν έλεγαν σ' έναν άνθρωπο να σκεφτεί. Με άλλα λόγια η σκέψη ισοδυναμεί με διέγερση των μυών του φωνητικού συστήματος του ανθρώπου.

Σύμφωνα όμως με τους Piaget και Vygotsky η γλώσσα είναι κάτι παραπάνω από κίνηση των μυών του λάρυγγα. Σύμφωνα με τον Piaget η γνωστική ανάπτυξη καθορίζει τη γλωσσική ανάπτυξη και η γλώσσα φανερώνει το γνωστικό επίπεδο ενός ανθρώπου. Επίσης ο εγωκεντρικός λόγος τον οποίο χρησιμοποιούν τα μικρά παιδιά οφείλεται στην έλλειψη κοινωνικοποίησης. Σύμφωνα με τον Chomsky η γλωσσική ικανότητα είναι ανεξάρτητη της γνωστικής ικανότητας καθώς ο άνθρωπος δε γεννιέται με γνωστικές ικανότητες αλλά με ευφυΐα. Η γνωστική ικανότητα είναι κάτι το οποίο έρχεται αργότερα στον άνθρωπο.

Κατά τον Vygotsky ενώ γλωσσική και η γνωστική ικανότητα ξεκινούν ως διαφορετικές λειτουργίες στη συνέχεια υφίσταται μια αλληλεξάρτηση. Αυτό σημαίνει

ότι αρχικά κατά την περίοδο που ονομάζεται προλογική, οι ήχοι που παράγουν τα παιδιά δε σχετίζονται με τη σκέψη και κατ' αυτή την περίοδο η σκέψη είναι μη λεκτική. Από την ηλικία των 3 ετών η σκέψη ξεκινάει να γίνεται λεκτική και ξεκινούν οι εσωτερικοί διάλογοι στον άνθρωπο. Επίσης η γλωσσική ανάπτυξη καθορίζει και τη γνωσιακή ανάπτυξη.

Ενώ η γλώσσα καθορίζει τη γνωστική λειτουργία και τις δομές της σκέψης σύμφωνα με τους Sapir και Whorf. Η γλώσσα σχετίζεται με τον τρόπο που αντιλαμβανόμαστε τον κόσμο και τον τρόπο που θυμόμαστε. Διαφορετικές γλώσσες παράγουν διαφορετικές γνωστικές δομές .

2.10.3. Κατανόηση γλώσσας

Ο λόγος μπορεί να είναι γραπτός ή προφορικός. Ο προφορικός είναι πιο σύντομος καθώς ενώ ο γραπτός είναι μόνιμος. Υπολογίζεται ότι ο άνθρωπος μπορεί να συλλάβει μέχρι 20 φωνήματα το δευτερόλεπτο. Η κατανόηση του λόγου περιλαμβάνει νοητικές, σωματικές και ψυχολογικές διαδικασίες.⁷³

Ο ακροατής εκτός από φωνήματα θα πρέπει να αναγνωρίζει και λέξεις. Είναι ευκολότερη η αναγνώριση λέξεων μέσα σε μια πρόταση με νόημα από ότι σε μη νοηματικές προτάσεις. Από τον προφορικό λόγο αυτό το οποίο συγκρατεί ο ακροατής είναι κυρίως το νόημα και όχι η σειρά των λέξεων. Μεγάλο ρόλο παίζουν και τα συμφραζόμενα τα οποία βοηθούν στο σχηματισμό ολοκληρωμένου νοήματος.

Κάποια από τα βασικά στοιχεία που παίζουν ρόλο στην κατανόηση του προφορικού λόγου είναι: η γνώση του αντικειμένου, η ύπαρξη οπτικής επαφής με αυτόν ο οποίος απαγγέλει, η δυνατότητα διαλόγου, η δυνατότητα ο λόγος να είναι συνεχής ή διακοπτόμενος και ο επιτονισμός σε κάποια βασικά μέρη των προτάσεων, η ταχύτητα ροής του λόγου και η καθαρότητα της άρθρωσης.

⁷³ http://www.greek-language.gr/greekLang/modern_greek/foreign/education/proposals/01a.html

2.10.4. Η Παραγωγή της γλώσσας

Για τη παραγωγή προφορικού λόγου διενεργούνται κάποιες εσωτερικές διαδικασίες όπως:

- η λεκτικοποίηση η οποία είναι η μεταφορά όσων σκεφτόμαστε με μορφή λόγου συνδυάζοντας το κατάλληλο λεξιλόγιο.
- Η φωνολογική κωδικοποίηση που σχετίζεται με τη σωστή σειρά των ήχων στο λόγο και η ανάλυση των δισταγμών που έχει να κάνει με την κατανομή των παύσεων στο λόγο
- Και ο συντακτικός σχεδιασμός

Πρέπει επίσης να λαμβάνονται υπόψη κάποια συγκεκριμένα θέματα. Ένα βασικό ζήτημα είναι τα λεκτικά λάθη τα οποία γίνονται κατά την παραγωγή του λόγου. Τα λάθη αυτά μπορεί να αφορούν ένα λάθος ή περισσότερα λάθος γράμματα τα οποία αλλάζουν όλο το νόημα σε μια φράση. Έτσι λοιπόν τα λεκτικά λάθη επίσης μπορεί να εντοπίζονται σε επίπεδο φωνολογικών χαρακτηριστικών, φωνήματος, συλλαβής μορφήματος λέξης και φράσης⁷⁴.

Εξηγήσεις πάνω στην παραγωγή των λεκτικών λαθών έχουν προσπαθήσει να δώσουν διάφοροι ψυχολόγοι και γλωσσολόγοι. Κατά τον Φρόυντ οφείλονται λόγω του υποσυνείδητου και των απωθημένων σκέψεων οι οποίες όμως υπάρχουν στο υποσυνείδητο. Με την άποψη αυτή διαφώνησαν άλλοι σπουδαίοι ψυχολόγοι και ψυχογλωσσολόγοι όπως ο Ρούντολφ Μερινγκέρ ο οποίος θεώρησε τα λεκτικά λάθη ως εισβολές γραμμάτων στα χείλη μας, ενώ ο ψυχογλωσσολόγος Ρομπ Χαρτσούικερ θεώρησε τέτοιου είδους λάθη ως εντελώς αθώα. Η κούραση, τα νεύρα, το άγχος και η γρήγορη ροή του λόγου μπορεί να είναι παράγοντες που συμβάλλουν στην παραγωγή λεκτικών λαθών.

Η ορθή παραγωγή λόγου πρέπει να λαμβάνει υπόψη κάποιες εξαρτήσεις μεταξύ κατηγοριών λέξεων καθώς τα άρθρα πρέπει να συμφωνούν κατά αριθμό, γένος και πτώση με ουσιαστικά, επίσης εξαρτήσεις υπάρχουν στην ελληνική αλλά και σε άλλες γλώσσες μεταξύ των αντωνυμιών και των ρημάτων.

⁷⁴ Harley.T.(2008). *Η Ψυχολογία της Γλώσσας*. Θεσσαλονίκη: Univercity Studio Press. Σελ. 472.

2.10.5. Η μνήμη της γλώσσας

Οι άνθρωποι έχει αποδειχθεί ότι χρησιμοποιούν το ίδιο λεξιλόγιο, τις ίδιες συντακτικές και προτασιακές δομές. Αυτό το οποίο αλλάζει στη ουσία είναι το νόημα. Πολύ σημαντικό ρόλο παίζει η ύπαρξη σχετικού υπόβαθρου καθώς και η ύπαρξη συνδέσμων του παλιού κείμενου στο νέο.

2.10.6. Πρόβλεψη λέξεων (Word Prediction)

Η πρόβλεψη των λέξεων είναι μια λειτουργία επεξεργαστών κειμένου προκειμένου να είναι δυνατή η πληκτρολόγηση λιγότερων γραμμάτων. Εκτός από τους επεξεργαστές κειμένου η λειτουργία αυτή υπάρχει και σε μηχανές αναζήτησης.

Στην ουσία μόλις ο χρήστης πληκτρολογήσει τα πρώτα γράμματα της λέξης ή του όρου αναζήτησης, η εφαρμογή προσπαθεί να μαντέψει τη λέξη ώστε να μην είναι απαραίτητη η πληκτρολόγηση ολόκληρης της λέξης. Στην ουσία η εφαρμογή αυτή απομνημονεύει λέξεις τις οποίες έχουμε είδη πληκτρολογήσει ή όρους τους οποίους έχουμε ξανά αναζητήσει στο διαδίκτυο προηγούμενες φορές. Υπάρχουν διαφορετικές μέθοδοι πρόβλεψης ανάλογα τις ανάγκες της εφαρμογής⁷⁵ η οποία μπορεί να είναι η πληκτρολόγηση γραπτού μηνύματος, η σύνταξη κειμένου, η αναζήτηση στο διαδίκτυο.

Με τη δυνατότητα αυτή είναι εφοδιασμένα και πολλά κινητά τηλέφωνα νέας γενιάς και ψηφιακές συσκευές (tablets) προκειμένου να εξυπηρετούν τους χρήστες στην πληκτρολόγηση γραπτών μηνυμάτων.

Αυτή η λειτουργία πολλές φορές μπορεί να φανεί χρήσιμη γιατί στην πληκτρολόγηση προσφέρει μεγαλύτερη ταχύτητα. Επίσης μπορεί να βοηθήσει άτομα με μειωμένες ικανότητες πληκτρολόγησης λόγω κινητικών προβλημάτων που δεν έχουν τη δυνατότητα χρησιμοποίησης των κλασικών πληκτρολογίων και κάνουν χρήση εξειδικευμένων συσκευών εισόδου όπως trackballs, mouthsticks⁷⁶.

Επίσης μπορεί να βοηθήσει παιδιά με προβλήματα γραφής όπως δυσλεκτικά παιδιά. Γι' αυτό το λόγο υπάρχουν ειδικά λογισμικά για Android και Ios συσκευές τα οποία

⁷⁵ Miller, J. W. (1998). *U.S. Patent No. 5,805,911*. Washington, DC: U.S. Patent and Trademark Office.

⁷⁶ http://www2.edc.org/ncip/library/wp/What_is.htm

εκτός από την πρόβλεψη λέξεων προσφέρουν και φωνητική υποστήριξη για τη διευκόλυνση δυσλεκτικών παιδιών και παιδιών με προβλήματα όρασης. Χαρακτηριστικό παράδειγμα είναι το λογισμικό spellbetter⁷⁷ ή το iWordQ⁷⁸.

2.11. Η Νευρογλωσσολογία

Η νευρογλωσσολογία ασχολείται με τη μελέτη των νευρικών μηχανισμών του ανθρώπινου εγκεφάλου οι οποίοι συμβάλλουν στην ανάπτυξη γλωσσικών ικανοτήτων στον άνθρωπο. Συνδέεται και με την ψυχολογία καθώς αντικείμενό της είναι η μελέτη της λειτουργίας του εγκεφάλου και σχετίζεται με την απόκτηση, κατανόηση και παραγωγή του λόγου.

2.11.1. Η Ανατομία του εγκεφάλου

Ο εγκέφαλος αποτελείται από δύο ημισφαίρια τα οποία επικοινωνούν μεταξύ τους. Στα δύο ημισφαίρια υπάρχουν συγκεκριμένα σημεία στα οποία αντιστοιχούν συγκεκριμένες δεξιότητες και αντιληπτικές ικανότητες του ανθρώπου. Στους περισσότερους δεξιόχειρες αλλά και σε αριστερόχειρες ανθρώπους, στο αριστερό ημισφαίριο λαμβάνουν χώρα λεκτικές, αναλυτικές και λογικές νοητικές διεργασίες όπως η γραφή και ο λόγος. Επίσης το αριστερό ημισφαίριο είναι υπεύθυνο για την αίσθηση του χρόνου και την ανάπτυξη μαθηματικών, γλωσσικών και λογικών ικανοτήτων του ανθρώπου.

Το δεξί ημισφαίριο είναι υπεύθυνο για τα οπτικά ερεθίσματα, την αναγνώριση εικόνων προσώπων και αντικειμένων, την επικοινωνία, την ανάπτυξη συναισθημάτων, τη φαντασία και γενικώς για την ανάπτυξη καλλιτεχνικής έκφρασης.

Το όλο σύστημα χωρίζεται σε πέντε βασικά τμήματα τα οποία είναι: το επάνω μέρος του εγκεφάλου, η παρεγκεφαλίδα η οποία βρίσκεται κάτω από το πίσω μέρος του εγκεφάλου, ο φλοιός του εγκεφάλου, ο νωτιαίος μυελός και τέλος από το στέλεχος το

⁷⁷ <https://www.spellbetterapp.com/>

⁷⁸ <https://apps.apple.com/us/app/iwordq-us/id557929840>

οποίο συνδέει το φλοιό του εγκεφάλου με το νωτιαίο μυελό. Η ραχιαία άποψη του εγκεφάλου αποτελείται από περιοχές που χωρίζονται από ρωγμές (έλικα - gyri) και μικρά αυλάκια ανάμεσα από κάθε έλικα (αύλακα - sulci).

Επίσης ο ανθρώπινος εγκέφαλος αποτελείται από νευρικά κύτταρα τα οποία ονομάζονται νευρώνες και συνδέονται μεταξύ τους. Υπάρχουν περίπου 100 δισεκατομμύρια νευρώνες στον άνθρωπο, δηλαδή νευρικά κύτταρα τα οποία δέχονται πληροφορίες, τις επεξεργάζονται και τις μεταβιβάζουν σε άλλους νευρώνες. Οι νευρώνες μπορούν να επικοινωνούν μεταξύ τους μέσω ηλεκτρονικών σημάτων. Τα σήματα αυτά μεταβιβάζονται από νευρώνα σε νευρώνα με τη βοήθεια των νευροδιαβιβαστών και με τη βοήθεια μιας χημικής μετάδοσης.

2.11.2. Η Παιδική ηλικία

Ο άνθρωπος όταν γεννιέται δε διαθέτει δύο εγκεφαλικά ημισφαίρια αλλά ένα και ο διαχωρισμός τους πραγματοποιείται κατά την παιδική ηλικία. Όσον αφορά τη γλώσσα τα δύο ημισφαίρια είναι ίδια κατά τη γέννηση ενός παιδιού αλλά σταδιακά πραγματοποιείται απόκτηση νευρώνων. Μια βλάβη στο αριστερό ημισφαίριο του εγκεφάλου μπορεί να προκαλέσει απώλεια λόγου η οποία μπορεί να επανέρθει αν η βλάβη συμβεί κατά την παιδική ηλικία, αλλά μπορεί να είναι μόνιμη αν συμβεί σε ενήλικους. Αυτό συμβαίνει γιατί κατά την παιδική ηλικία οι λειτουργίες του ενός ημισφαιρίου μπορούν να μεταφερθούν στο άλλο ημισφαίριο.

Η βρεφική ηλικία είναι μια κρίσιμη περίοδος και ιδιαίτερα η ηλικία έως το τέλος του πρώτου έτους καθώς τότε χτίζονται οι νευρολογικές βάσεις της λογικής σκέψης και ευφυΐας. Επίσης στη βρεφική ηλικία ο εγκέφαλος πραγματοποιεί τη διασύνδεση των νευρώνων και η δομή του εγκεφάλου ενός βρέφους, αλλάζει συνεχώς και συμβαδίζει με την αλλαγή και την ανάπτυξη των γλωσσικών δυνατοτήτων του παιδιού.

Σημαντικό ρόλο παίζει το περιβάλλον στο οποίο μεγαλώνει ένα παιδί καθώς οι λέξεις που ακούει αποθηκεύονται στους νευρώνες του εγκεφάλου του και μπορεί αργότερα να τους χρησιμοποιήσει. Άρα όσο περισσότερα λόγια ακούει ένα μικρό παιδί τόσο πιο πλούσιες θα είναι οι γλωσσικές του δυνατότητες και ενδεχομένως και η ευφυΐα του.

2.11.3. Οι Λειτουργίες του εγκεφάλου

Οι περιοχές του αριστερού ημισφαιρίου που σχετίζονται με τη γλώσσα είναι η περιοχή Broca και η περιοχή Wernicke. Διάφοροι επιστήμονες έχουν προσπαθήσει να εντοπίσουν ποια τμήματα του εγκεφάλου είναι υπεύθυνα για συγκεκριμένες λειτουργίες.

Σύμφωνα με το Μοντέλο Γλωσσικής Επεξεργασίας Wernicke-Lichtheim υπάρχουν τρεις βασικές περιοχές οι οποίες είναι : η περιοχή Wernicke που είναι υπεύθυνη για την αποθήκευση νοητικών και σημασιολογικών αναπαραστάσεων και για διάφορες λεξικές πληροφορίες. Επίσης υπάρχει η περιοχή Broca που είναι υπεύθυνη για την κατασκευή ακολουθιών λέξεων και τέλος μια τοξοειδής δεσμίδα η οποία είναι υπεύθυνη για τη μεταφορά σημασιολογικών και λεξικών πληροφοριών. Το μοντέλο αυτό δεν είναι απολύτως ακριβές ως προς την περιγραφή των περιοχών που συμμετέχουν στη παραγωγή και κατανόηση της γλώσσας.

Στον εγκέφαλο υπάρχουν νευρώνες με σκοπό την αναπαράσταση εννοιών και κάποιες άλλες δομές που έχουν ως σκοπό τη δημιουργία λέξεων και προτάσεων και ανάμεσα τους υπάρχει ένα διαμεσολαβητικό υπόστρωμα . Οι νευρώνες που βρίσκονται στο αριστερό ημισφαίριο εξυπηρετούν τη διεκπεραίωση λέξεων και προτάσεων. Για την παραγωγή και την κατανόηση της γλώσσας υπεύθυνες είναι οι περιοχές Broca στον μετωπιαίο λοβό και η περιοχή Wernicke στον κροταφικό λοβό. Η γλωσσική ικανότητα προϋποθέτει τη συνεργασία πολλών μερών του φλοιού. Τα δίκτυα που συνδέουν τους νευρώνες του εγκεφάλου μας αποτυπώνουν ερεθίσματα από το περιβάλλον πράγμα το οποίο μας κάνει πιο ευφυείς και δημιουργικούς και όσο μεγαλύτερος είναι ο αριθμός των νευρωνικών διασυνδέσεων τόσο πιο ευφυές είναι ένα άτομο.

Με τη χρήση ηλεκτρονικών υπολογιστών και του κατάλληλου τομογράφου είναι δυνατόν να εντοπιστούν τα σημεία του ανθρώπινου εγκεφάλου που συμμετέχουν όταν ένας άνθρωπος π.χ. κατονομάζει ένα αντικείμενο καθώς και ποια σημεία του εγκεφάλου συμμετέχουν και σε άλλες γλωσσικές δραστηριότητες.

2.11.4. Γλωσσικές Διαταραχές και Νευρογλωσσολογία

Κάθε γλωσσική διαταραχή μπορεί να αντιστοιχεί σε διαφορετικές περιοχές εγκεφαλικής βλάβης. Για τη διάκριση των διαταραχών σε κατηγορίες κάποια χρήσιμα κριτήρια θα μπορούσαν να είναι : η άρθρωση, η κατανόηση λέξεων και προτάσεων, η παραγωγή λέξεων και προτάσεων η ανάγνωση και η γραφή. Τυπικά σύνδρομα μπορεί να παρουσιάζουν συμπτώματα όπως κακή κατανόηση κειμένου, μη ρέοντα λόγια, προβλήματα και αδυναμίες στη γραφή. Στο σύνολο τους τα σύνδρομα αυτά ονομάζονται αφασικά.

Σε κάποια αφασικά σύνδρομα ο ασθενής μπορεί να έχει γνώση και συναίσθηση του προβλήματος του και ενδεχομένως να υποφέρει από κατάθλιψη. Μετά από σωστή διάγνωση και με κατάλληλο τρόπο αντιμετώπισης ο οποίος μπορεί να περιλαμβάνει και λογοθεραπεία ο ασθενής μπορεί να επανέρθει.

Για να ερευνήσουμε τη σχέση μεταξύ γλώσσας και εγκεφάλου μπορούμε να κάνουμε καταγραφή της εγκεφαλικής δραστηριότητας καθώς συμβαίνουν διάφορες νοητικές λειτουργίες. Για να διερευνήσουμε την εγκεφαλική δραστηριότητα αρκεί να δούμε τη ροή του αίματος στον εγκέφαλο και ένας άλλος τρόπος είναι τα μοντέλα πυροδότησης των νευρώνων δημιουργώντας ηλεκτρονικά σήματα.

Τις περιοχές του εγκεφάλου που ενεργοποιούνται όταν κάνουμε χρήση της γλώσσας μπορούμε να τις εντοπίσουμε και με τη χρήση νευροαπεικονίσεων. Μέθοδοι νευροαπεικονίσεων είναι: τομογραφία εκπομπής ποζιτρονίων (PET), μαγνητική απεικόνιση (fMRI) , το ηλεκτροεγκεφαλογράφημα (EEG), και το μαγνητοεγκεφαλογράφημα (MEG). Στο ηλεκτροεγκεφαλογράφημα (EEG) οι κυματομορφές που προκύπτουν αποθηκεύονται σε ηλεκτρονικό υπολογιστή.

2.11.5. Μηχανισμοί σύνθεσης ομιλίας

Τα τελευταία χρόνια έχουν κατασκευαστεί συσκευές παραγωγής ομιλίας, κατάλληλες για άτομα παραπληγικά άτομα που δεν έχουν ικανότητα ομιλία. Οι συσκευές αυτού του είδους συνδέονται με τους νευρώνες του εγκεφάλου και διαβάζουν τη σκέψη του ασθενούς. Στη συνέχεια πραγματοποιείται παραγωγή ήχου μέσω κατάλληλων ηχείων.

2.11.6. Τεχνητά νευρωνικά δίκτυα

Ένα κύκλωμα διασυνδεδεμένων βιολογικών νευρώνων αποτελεί ένα νευρωνικό δίκτυο. Σκοπός των τεχνητών νευρωνικών δικτύων είναι η μίμηση της λειτουργίας των βιολογικών νευρώνων μέσω της ανάπτυξης αλγορίθμων. Τα νευρωνικά δίκτυα παρέχουν υπολογιστική ισχύ πολύ υψηλού επιπέδου προκειμένου να εκτελεστεί ένα υπολογιστικό πρόβλημα ή να κατασκευαστεί κάποια εξειδικευμένη εφαρμογή.

Ένα τέτοιο δίκτυο αποτελείται από σύνολα τριών ειδών νευρώνων: τους νευρώνες εισόδου, τους νευρώνες εξόδου και τους υπολογιστικούς νευρώνες. Οι νευρώνες εισόδου και εξόδου δεν εκτελούν υπολογιστικές πράξεις εν αντιθέσει με τους υπολογιστικούς νευρώνες. Ο κάθε νευρώνας αποτελείται από πολλές εισόδους και μια μόνο έξοδο. Κάθε είσοδος έχει μια τιμή βάρους μέσω της οποίας μπορούμε να δούμε πόσο στενά είναι συνδεδεμένοι οι δύο νευρώνες που συνδέονται με το βάρος αυτό. Η τιμή του βάρους κυμαίνεται από -1 έως 1. Τα νευρωνικά δίκτυα έχουν την ιδιότητα μπορούν να εκπαιδεύονται και να ανταλλάσσουν δεδομένα μεταξύ τους με σκοπό τη σταδιακή σύλληψη της πληροφορίας και της ανάκτησής της κατόπιν.⁷⁹

Τα νευρωνικά δίκτυα εκτελούν παράλληλες επεξεργασίες και μια συγκεκριμένη εργασία κατανέμεται σε διάφορα τμήματα του δικτύου, επίσης ο κάθε νευρώνας διαθέτει μεγάλο αριθμό διασυνδέσεων. Έτσι λοιπόν εξασφαλίζονται εκτός από υψηλή υπολογιστική ισχύ και πολύ μεγάλες ταχύτητες. Επιπλέον μια πληροφορία αποθηκεύεται σε κάποιον νευρώνα και όχι σε μια μνήμη όπως στους ηλεκτρονικούς υπολογιστές.

Παρόλο που τα νευρωνικά δίκτυα βρίσκουν εφαρμογή σε πάρα πολλούς τομείς, θεωρείται ότι είναι ένα αντικείμενο το οποίο δεν έχει διερευνηθεί και αξιοποιηθεί ακόμη πλήρως. Κάποιοι από τα πεδία στα οποία χρησιμοποιείται είναι: η ρομποτική, η αναγνώριση και η παραγωγή ομιλίας, τα χρηματοοικονομικά, οι τηλεπικοινωνίες, η ιατρική, η βιολογία⁸⁰ κ.α.

⁷⁹ Πούλος, Μ., 2015. *Σημασιολογική επεξεργασία της πληροφορίας*. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών

⁸⁰ Λιβιέρης, Ι. (2008). *Αποτίμηση μεθόδων εκπαίδευσης τεχνητών νευρωνικών δικτύων και εφαρμογές* (Doctoral dissertation).

2.12.Μαθηματική Γλωσσολογία

Βασικός σκοπός της μαθηματικής γλωσσολογίας είναι η τυποποίηση των γλωσσών και των γραμματικών τους με τη βοήθεια μαθηματικών μεθόδων όπως αλγορίθμων , της άλγεβρας , της θεωρίας των συνόλων, των πεπερασμένων αυτόματων κ.α. Η μαθηματική γλωσσολογία αφορά τη μελέτη των ιδιοτήτων και της δομής μιας γλώσσας και την έκφρασή της με μαθηματικά μοντέλα.

Περιπτώσεις όπου μπορούν να χρησιμοποιηθούν μαθηματικά μοντέλα στη γλωσσολογία είναι ο υπολογισμός του συνόλου των λέξεων που αναγνωρίζουμε ή του συνόλου των λέξεων που μεταφράζουμε σωστά. Άρα αντικείμενα του ενδιαφέροντος είναι συλλογές λέξεων σε μια φυσική γλώσσα, η συλλογή προτάσεων, η συλλογή νοημάτων κτλ. Επίσης προκειμένου να μπορέσουμε να περιγράψουμε τη δομή των μερών του λόγου, χρησιμοποιούμε ειδικές συντακτικές δομές , τις οποίες μπορούμε να τις αναπαραστήσουμε με διαγράμματα, γράφους και δέντρα.

Εφαρμογή τα μαθηματικά βρίσκουν και σε διάφορους τομείς της γλωσσολογίας όπως η σύνταξη, η μορφολογία, η φωνολογία κ.α. Παρόλα αυτά θα μπορούσαμε να πούμε ότι η μαθηματική γλωσσολογία ανήκει περισσότερο στα εφαρμοσμένα μαθηματικά από ότι στη γλωσσολογία.

2.12.1. Οι γραμματικές

Κεντρική θέση στη μαθηματική γλωσσολογία έχουν οι γραμματικές οι οποίες περιέχουν το σύνολο των κανόνων που διαθέτει μια γλώσσα και κάνουν δυνατή την προσομοίωση των πιο ουσιωδών πτυχών της λειτουργίας της προσπαθώντας να υποκαταστήσουν το σύστημα ανάλυσης της δομής των λέξεων.

Την γραμματική την κατέχουν όλοι οι ομιλητές ακόμη και οι αγράμματοι και τα μικρά παιδιά που ομιλούν μια γλώσσα.

2.12.1.1. Γενετική – μετασχηματιστική γραμματική

Η γενετική γραμματική προσπαθεί να ερμηνεύσει τη δομή μιας γλώσσας και αναφέρεται στην υποσυνείδητη γνώση του γλωσσικού μας συστήματος από έναν ομιλητή, η οποία του δίνει τη δυνατότητα να παράγει και να κατανοεί πολλές προτάσεις οι οποίες είναι δομημένες με σωστό τρόπο. Η ικανότητα του ανθρώπου να παράγει άπειρες προτάσεις είναι έμφυτη και σε αυτό οφείλεται το γεγονός ότι τα μικρά παιδιά μαθαίνουν εύκολα τη γραμματική της γλώσσας τους.

Ο Chomsky θεωρεί ότι υπάρχουν κάποιες καθολικές δομές και χαρακτηριστικά στα οποία υπακούουν όλες οι γλώσσες όσο και αν διαφέρουν μεταξύ τους σε όλα τα επίπεδα της γλωσσολογικής ανάλυσης. Έτσι λοιπόν η γενετική γραμματική εξετάζει τη δομή των προτάσεων και προσπαθεί να βρει δομικά σχήματα για όλα τα είδη των προτάσεων και ουσιαστικά αποτελεί έναν αλγόριθμο παραγωγής σειρών στοιχείων και χαρακτήρων που ανήκουν σε μια γλώσσα. Η γενετική γραμματική επιτρέπει το σχηματισμό σωστών προτάσεων σε οποιαδήποτε γλώσσα.⁸¹

Οι γενετικές γραμματικές πρέπει να περιλαμβάνουν πάντα μια αρχική κατάσταση S , κάποια στοιχεία όπως λέξεις, γράμματα και σύμβολα κτλ, και δομικούς κανόνες. Κάνοντας επιλογή κανόνων μπορούμε να σχηματίσουμε στοιχειοσειρές. Πιο αναλυτικά σύμφωνα με τον Νόαμ Τσόμσκι μια γενετική γραμματική αποτελεί μια διατεταγμένη τετράδα του τύπου $G = (N, \Sigma, P, S)$ και περιλαμβάνει :

- Ένα πεπερασμένο σύνολο N μη τελικών συμβόλων
- Ένα πεπερασμένο σύνολο Σ τελικών συμβόλων για το οποίο ισχύει $\Sigma \cap N = \emptyset$
- Ένα πεπερασμένο σύνολο P κανόνων
- Μια αρχική κατάσταση $S \in N$

2.12.1.2. Κατηγοριακή Γραμματική

Στη μαθηματική γλωσσολογία πολύ σημαντικός είναι ο ρόλος των κατηγοριακών γραμματικών για τη σύνταξη των φυσικών γλωσσών. Με τη χρήση συγκεκριμένων τύπων και την εκτέλεση μιας σειράς βημάτων, μια κατηγορηματική γραμματική

⁸¹ Ρούσσου, Ά. 2015. Γλώσσα και γραμματική. [Κεφάλαιο Συγγράμματος]. Στο Ρούσσου, Ά. 2015. *Σύνταξη*. [ηλεκτρ. βιβλ.] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. κεφ 1. Διαθέσιμο στο: <http://hdl.handle.net/11419/582>

αποφασίζει εάν μια φράση ανήκει σε μια γλώσσα. Όλα τα γραμματικά συστατικά μιας πρότασης συσχετίζονται με κατηγορίες οι οποίες ορίζουν αν τα μέρη από τα οποία αποτελείται μια φράση είναι σωστά συνδυασμένα⁸².

Οι κατηγοριακές γραμματικές περιλαμβάνουν λεξικά τα οποία περιέχουν ένα σύνολο τύπων σε κάθε σύμβολο και κάποιους συμπερασματικούς κανόνες που καθορίζουν τον τρόπο με τον οποίο τα συντακτικά συστατικά συνδυάζονται μεταξύ τους σε μια πρόταση. Σημαντική διαφορά σε σχέση με άλλες γραμματικές είναι ότι οι κανόνες συμπερασμού είναι της μορφής $A \rightarrow B$ και μπορεί να έχουν δύο συναρτησιακούς τύπους έναν που εφαρμόζεται στο αριστερό μέρος του κανόνα και έναν που εφαρμόζεται στο δεξί.

Ένα από τα απλά είδη κατηγοριακής γραμματικής είναι η Γραμματική AB η οποία αποτελείται από την πλειάδα $(\Sigma, \text{Prim}, S, \triangleleft)$. Όπου Σ είναι ένα πεπερασμένο σύνολο συμβόλων, Prim είναι ένα σύνολο λέξεων και $\text{Tr}(\text{Prim})$ είναι ένα σύνολο τύπων και $S \in \text{Tr}(\text{Prim})$. Ο χαρακτήρας ' $\triangleleft$ ' συμβολίζει είναι το λεξικό το οποίο συσχετίζει τύπους με σύμβολα. Παράδειγμα κατηγοριακής γραμματικής για την αγγλική γλώσσα είναι το σύνολο (N, NP, S) όπου το N τα ουσιαστικά, το NP τις ονομαστικές προτάσεις και το S τις προτάσεις, τα επίθετο έχει τύπο N/N , τα αμετάβατα ρήματα $NP \backslash S$ και τα μεταβατικά ρήματα $(NP \backslash S) / NP$ ⁸³.

2.12.2. Σύνδεση της μαθηματικής γλωσσολογίας με την ανάκτηση της πληροφορίας

Η μαθηματική γλωσσολογία αποτελεί συνδυασμό της γλωσσολογίας με τα μαθηματικά και την επιστήμη των υπολογιστών. Στην ουσία αποτελεί υποκατηγορία της Τεχνητής Νοημοσύνης. Σχετίζεται με τον προγραμματισμό των ηλεκτρονικών υπολογιστών καθώς προκειμένου να παραχθεί μια φυσική γλώσσα απαιτείται μαθηματική τυποποίηση κανόνων και δομών.

Πολύ σημαντικός είναι ο ρόλος της μαθηματικής γλωσσολογίας και στη μηχανική μετάφραση φυσικών γλωσσών, η οποία αποτελεί υπηρεσία που παρέχουν οι πιο δημοφιλείς web browsers. Μάλιστα η υπολογιστική γλωσσολογία γεννήθηκε από

⁸² Steedman, M. (1992). Categorical grammar. *Technical Reports (CIS)*, 466.

⁸³ Steedman, M., & Baldridge, J. (2011). Combinatory categorical grammar. *Non-Transformational Syntax: Formal and Explicit Models of Grammar*, 181-224.

τους Αμερικάνους στην προσπάθειά τους να πραγματοποιούν αυτόματες μεταφράσεις ρωσικών επιστημονικών κειμένων. Οι υπολογιστές δεν μπορούν να επεξεργάζονται γράμματα και λέξεις, μπορούν όμως να επεξεργάζονται αριθμούς και να πραγματοποιούν μαθηματικές πράξεις. Η αναπαράσταση γραμματικών κανόνων και δομών με μαθηματικά κάνει εφικτή την επεξεργασία φυσικής γλώσσας από τους ηλεκτρονικούς υπολογιστές.

2.13. Λεξικογραφία

Λεξικογραφία είναι ο κλάδος της γλωσσολογίας ο οποίος ασχολείται με τον τρόπο σύνταξης ενός λεξικού. Χωρίζεται σε δύο κλάδους: την πρακτική λεξικογραφία η οποία ασχολείται με την εγγραφή και την έκδοση των λεξικών και ο δεύτερος κλάδος είναι η θεωρητική λεξικογραφία η οποία σχετίζεται με την ανάλυση των διαφόρων εννοιών.

Σύμφωνα με τον Noam Chomsky «Για κάθε φυσική γλώσσα υπάρχει λεξικό το οποίο αποτελείται από το σύνολο των μορφημάτων μιας γλώσσας».⁸⁴

2.13.1. Λεξικό

Όσον αφορά την επιστήμη της γλωσσολογίας το λεξικό θα μπορούσε να οριστεί ως το σύνολο των λέξεων που έχει στη διάθεσή της για επικοινωνία μια γλωσσική κοινότητα και περιέχει διάφορες πληροφορίες όπως:

- Γραμματικές πληροφορίες που αφορούν τη γραμματική κατηγορία στην οποία ανήκει το λήμμα καθώς επίσης και την κλίση της λέξης αν ανήκει σε κλιτό μέρος του λόγου.
- Συντακτικές πληροφορίες, όπως η θέση της λέξης μέσα στο λόγο, σύμφωνα με τους κανόνες του συντακτικού.
- Σημασιολογικές, όπως η περιγραφή της σημασίας του λήμματος είτε κυριολεκτική είτε, μεταφορική,

⁸⁴ Τσίγκου, Μ. (2011). *Γλωσσολογία & Μετάφραση*. Αθήνα: Δίαυλος.

- φωνολογικές και μορφολογικές, που αφορούν το συλλαβισμό και τα μορφήματα από τα οποία αποτελείται η λέξης
- ετυμολογικές, ορθογραφικές που έχουν να κάνουν με την προέλευση την πρώτη μορφή και την αρχική σημασία της λέξης καθώς και την εξέλιξή της με τη πάροδο του χρόνου. Π.χ. για τη λέξη σπίτι η ετυμολογία είναι: < μεσαιωνική ελληνική σπίτιν < όσπίτιν < ελληνιστική κοινή όσπίτιον < λατινική hospitium < hospes
- Φράσεις-λεξικολογικές συνάψεις (αναγκαστική προσγείωση, κατώτατος μισθός),
- περιφραστικές λέξεις (π.χ. φακός επαφής, ηλεκτρονικός υπολογιστής), παροιμίες (γελάει καλύτερα όποιος γελάει τελευταίος),
- ιδιωματισμοί (π.χ. του έδωσε τα παπούτσια στο χέρι) ή φράσεις από την επιστημονική ορολογία (π.χ. αντιποίηση αρχής)

2.13.2. Είδη λεξικών

Κάποια από τα βασικά είδη λεξικών είναι τα παρακάτω:

- **Γλωσσικά Λεξικά** τα οποία περιγράφουν και παραθέτουν μεταγλωσσικές πληροφορίες (γραμματικές, ορθογραφικές, ερμηνευτικές κτλ.) για λέξεις. Τα γλωσσικά λεξικά μπορεί να είναι μονόγλωσσα ή πολύγλωσσα στα οποία προσφέρεται μετάφραση των λημμάτων σε μια ή παραπάνω γλώσσες.
- Επίσης τα γλωσσικά λεξικά μπορεί να είναι **ερμηνευτικά ή αλλιώς σημασιολογικά** τα οποία για κάθε λήμμα παραθέτουν την αντίστοιχη ερμηνεία μέσα από αναλυτικές περιγραφές
- **Εννοιολογικά λεξικά** στα οποία παρέχονται λέξεις που περιγράφουν το λήμμα καθώς και οι σχέσεις που έχει κάθε λήμμα με αυτές τις λέξεις. Χαρακτηριστικά παραδείγματα αποτελούν οι θησαυροί οι οποίοι παρουσιάζουν σχέσεις μεταξύ των λέξεων και παρέχουν συγγενικές με το λήμμα λέξεις.
- **Τα εγκυκλοπαιδικά Λεξικά** τα οποία δίνουν πληροφορίες για κάποια λήμματα τα οποία μπορεί να είναι επιστημονικού και γενικότερα γνωστικού

περιεχομένου. Τα εγκυκλοπαιδικά λεξικά μπορεί να είναι γενικού περιεχομένου ή ειδικού περιεχομένου π.χ. Λεξικό όρων πληροφορικής, λεξικό ιατρικής ορολογίας κτλ.

2.13.3. Ο ρόλος των λεξικών στην ανάκτηση της Πληροφορίας

Το διαδίκτυο σήμερα αποτελεί μεγάλο αποθετήριο ηλεκτρονικών λεξικογραφικών όρων. Τα είδη ηλεκτρονικών λεξικών είναι δύο: εκείνα που φιλοξενούνται σε ιστοτόπους και για τα οποία δεν αντίστοιχη έντυπη μορφή και αυτά για τα οποία υπάρχει αντίστοιχο έντυπο λεξικό. Υπάρχουν πολλοί ιστότοποι που φιλοξενούν θησαυρούς, γλωσσάρια, λεξικά, εκφράσεων, εγκυκλοπαιδικά λεξικά κ.α..

Η ηλεκτρονική λεξικογραφία μπορεί να επιφέρει πολλά πλεονεκτήματα τόσο στη διαχείριση και στην κατάρτιση του λεξικογραφικού υλικού⁸⁵ και σίγουρα μεγάλη ευκολία. Κάποια από τα πλεονεκτήματα που παρέχει είναι τα ακόλουθα:

- ταχύτητα πρόσβασης στους χρήστες χάριν την αμεσότητα του διαδικτύου.
- Επίσης χάρις τις δυνατότητες του Web 2.0 οι διαδικτυακοί χρήστες μπορούν να έχουν διαδραστική χρήση σε τέτοιους ιστοτόπους⁸⁶ προσθέτοντας λήμματα και αλλάζοντας ή εμπλουτίζοντας το περιεχόμενό τους.
- Βασικό πλεονέκτημα είναι και ο τρόπος αναζήτησης καθώς υπάρχουν πολλοί εναλλακτικοί τρόποι ανάλογα τον ιστότοπο που φιλοξενεί το λεξικό. Π.χ αναζήτηση με ολόκληρη τη λέξη, αναζήτηση με αυτόματη συμπλήρωση της λέξης με την οποία μπορεί ο χρήστης πληκτρολογεί την αναζητούμενη λέξη και ταυτόχρονα να εμφανίζεται ένας κατάλογος με διάφορες πιθανές λέξεις που μπορεί να ταιριάζουν (π.χ. Wikipedia). Επίσης υπάρχει αναζήτηση με επιπλέον χρήση χαρακτήρων μπαλαντέρ όπως το «*» που δηλώνει έναν ή περισσότερους οποιουδήποτε χαρακτήρες και το «?» που δηλώνει έναν οποιοδήποτε χαρακτήρα. Επιπλέον πολλές μηχανές αναζήτησης μπορούν να εντοπίσουν το λήμμα ανεξάρτητα αν δεν είναι σωστή η ορθογραφία του ή

⁸⁵ Pantazara, A. V., Mantzari, E., Πανταζάρα, Α. Β., & Μάντζαρη, Ε. (2015). Αναζήτηση σε σώματα κειμένων.

⁸⁶ Lawrence, S., & Giles, C. L. (1999). Accessibility of information on the web. *Nature*, 400(6740), 107.

ανεξάρτητα από την κλίση του. Τέλος υπάρχει και η αναζήτηση με χρήση φίλτρων που μας επιτρέπει να κάνουμε μια πιο εξειδικευμένη αναζήτηση.⁸⁷

2.13.3.1. Wordnet

Το WordNet είναι μια μεγάλη λεξικογραφική βάση δεδομένων για την αγγλική γλώσσα η οποία δημιουργήθηκε από το Πανεπιστήμιο του Princeton στα μέσα της δεκαετίας του 80'. Περιλαμβάνει ουσιαστικά, ρήματα, επίθετα και επιρρήματα. Για κάθε λήμμα περιλαμβάνει τις πιθανές διαφορετικές ερμηνείες που μπορεί να υπάρχουν και είναι κάτι περισσότερο από λεξικό καθώς πραγματοποιεί ομαδοποίηση των λέξεων σε σύνολα συνωνύμων⁸⁸.

Λόγω της δομής του το WordNet χρησιμοποιείται σε εφαρμογές υπολογιστικής γλωσσολογίας και επεξεργασίας φυσικής γλώσσας.

2.13.3.2. Χρήση λεξικών σε εφαρμογές

Σε πολλές εφαρμογές του Παγκόσμιου Ιστού είναι βασικής σημασίας η χρήση των κατάλληλων ηλεκτρονικών λεξικών. Για την κατασκευή εφαρμογών όπως π.χ. το sentimental analysis στο οποίο είναι απαραίτητη η χρήση λεξικών συναισθημάτων προκειμένου να είναι δυνατή η ταξινόμηση tweet, μηνυμάτων, ιστοσελίδων. Επίσης για τη μηχανική μετάφραση είναι απαραίτητη η χρήση κατάλληλων ηλεκτρονικών λεξικών τα οποία επιτρέπουν την αντιστοίχιση λέξεων και φράσεων από μια γλώσσα σε μια άλλη.

⁸⁷ Pantazara, A. V., Mantzari, E., Πανταζάρα, Α. Β., & Μάντζαρη, Ε. (2015). Αναζήτηση σε σώματα κειμένων, σελ. 8-16.

⁸⁸ <https://wordnet.princeton.edu/>

2.14. Η Τεκμηρίωση της Γλώσσας

Οι γλώσσες που ομιλούνται στον πλανήτη υπολογίζεται ότι είναι περίπου 6900. Παρόλα αυτά εκτιμάται ότι κάθε δύο εβδομάδες μια γλώσσα εξαφανίζεται⁸⁹. Επίσης πολλές γλώσσες κινδυνεύουν να εξαφανιστούν λόγω της μείωσης του πληθυσμού τους. Χαρακτηριστικό παράδειγμα είναι η γλώσσα Ουρούμ την οποία τη συναντάμε στον Καύκασο. Στις αρχές της δεκαετίας του 1980 η γλώσσα αυτή ομιλούνταν από περίπου 30000 κατοίκους, ενώ σήμερα απομένουν περίπου 1500 άτομα⁹⁰ που μιλούν τη συγκεκριμένη γλώσσα. Σε μια άλλη γλώσσα που ονομάζεται Ubuhi και που ομιλούνταν πάλι στην περιοχή του Καυκάσου, το τελευταίο άτομο που τη μιλούσε πέθανε το 1992⁹¹. Σε κίνδυνο βρίσκονται πολλές γλώσσες στις χώρες της Βόρειας Αμερικής και της Αυστραλίας. Χαρακτηριστικό παράδειγμα αποτελεί η γλώσσα Άνω Τανάνα στην Αλάσκα την οποία την μιλούν περίπου 25 άτομα. Επίσης υπάρχουν γλώσσες στην Ασία στην περιοχή των Ιμαλαίων τις οποίες τις μιλούν 8 άτομα ενώ το πρόβλημα αυτό είναι υπαρκτό και σε χώρες της Ευρώπης όπως για παράδειγμα στις Σκανδιναβικές χώρες.

Όταν εξαφανίζεται μια γλώσσα εξαφανίζεται και ένα κομμάτι της παγκόσμιας πολιτιστικής κληρονομιάς αφού παράλληλα χάνονται στοιχεία που αφορούν την ιστορία του, τα ήθη και τα έθιμα. Αναγκαία λοιπόν κρίνεται «Τεκμηρίωση της Γλώσσας» και γι' αυτό υπάρχει υποπεδίο της γλωσσολογίας που στοχεύει στην καταγραφή μιας γλώσσας ή μιας διαλέκτου.

Μια γλωσσική τεκμηρίωση, θεωρείται μια διαρκή, πολλαπλών χρήσεων καταγραφή μιας γλώσσας⁹². Η καταγραφή αυτή γίνεται για πολλούς λόγους όπως :

- προκειμένου να μην εξαφανιστεί και να διασωθεί η γλώσσα αυτή,

⁸⁹ <http://www.kathimerini.gr/782251/article/epikairothta/episthmh/mia-glwsa-pe8ainei-ka8e-dyo-evdomades-logw-anapty3hs>

⁹⁰ Ries, V., Skopeteas, S., Turan, E., & Nahrman, K. (2014). Discovering the prehistory of multilingual situations in the lexicon. An empirical study on the Caucasian Urum vocabulary. *Linguistik online*, 64(2).

⁹¹ Solé, R. V., Corominas-Murtra, B., & Fortuny, J. (2010). Diversity, competition, extinction: the ecophysics of language change. *Journal of The Royal Society Interface*, 7(53), 1647-1664.

⁹² Gippert, J., Himmelmann, N., & Mosel, U. (Eds.). (2006). *Essentials of language documentation* (Vol. 178). Walter de Gruyter.

- προκειμένου να χρησιμοποιηθεί από την επιστημονική κοινότητα για μελέτη και έρευνα,
- για ιστορικούς και πολιτιστικούς λόγους όπως για παράδειγμα η καταγραφή της ιστορίας, των εθίμων και της κουλτούρας μιας κοινότητας ανθρώπων.

Η τεκμηρίωση της γλώσσας είναι πολύ σημαντική για γλώσσες που κινδυνεύουν να χαθούν, για γλώσσες φυλών, κοινοτήτων ανθρώπων και μειονοτήτων που κινδυνεύουν να αφομοιωθούν στο γλωσσικό σύστημα των χωρών τους. Για τοπικές διαλέκτους οι οποίες με την πάροδο του χρόνου εξαφανίζονται λόγω της αστυφιλίας κ.α.

2.14.1. Τα δεδομένα

Στην διαδικασία της καταγραφής της γλώσσας απαραίτητη προϋπόθεση είναι η ενεργός συμμετοχή των ατόμων που μιλούν και γνωρίζουν τη γλώσσα με τη βοήθεια των οποίων πραγματοποιείται η συγκέντρωση όσο περισσότερων πρωτογενών δεδομένων, τα οποία προσφέρουν μια επαρκή περιγραφή για τη γλώσσα. Μια τέτοια περιγραφή θα πρέπει να περιλαμβάνει την καταγραφή των λέξεων και τη σημασία τους, την περιγραφή της γραμματικής και τη χρήσης της καθώς και διάφορα άλλα στοιχεία όπως εκφράσεις, ιδιωματισμούς, που αφορούν την προφορά, καθώς επίσης και ιστορίες, συνήθειες, μύθους και άλλα δεδομένα που αφορούν την κουλτούρα και τον πολιτισμό της κοινότητας που μιλάει τη γλώσσα

Αρχικά θα πρέπει να συγκεντρωθούν τα πρωταρχικά δεδομένα τα οποία θα πρέπει να είναι σε κατάλληλη μορφή αρχείων ήχου ή video που θα περιέχουν συνεντεύξεις, συζητήσεις και αφηγήσεις των ατόμων που ομιλούν τη γλώσσα. Μόλις πραγματοποιηθεί το στάδιο της καταγραφής, θα πρέπει τα δεδομένα αυτά αφού απομαγνητοφωνηθούν, να οργανωθούν με δομημένο τρόπο, π.χ. σε ψηφιακές βάσεις δεδομένων.

Πολύ σημαντικό σε μια τεκμηρίωση γλώσσας είναι να παρατηρηθεί και η γλωσσική συμπεριφορά: δηλαδή ο τρόπος με τον οποίο οι άνθρωποι πραγματικά επικοινωνούν μεταξύ τους. Πρέπει λοιπόν να παρατηρηθεί η καθημερινότητα, τα έθιμα, οι συνήθειες και οι συζητήσεις των ατόμων που μιλούν τη γλώσσα. Όλη αυτή η

καταγραφή πρέπει να αφορά ένα αντιπροσωπευτικό δείγμα καθώς είναι αδύνατο να καταγραφούν τα πάντα.

Από όλη αυτή την καταγραφή πολλές φορές μπορούν να προκύψουν χρήσιμα μεταδεδομένα που φανερώνουν τον πολιτισμό και τις συνήθειες των κοινοτήτων που μιλούν τη γλώσσα. Τέτοια μεταγνώση μπορεί να αφορά: τις τοπικές ταξινομίες για τα φυτά, τα ζώα και τα μουσικά όργανα και στυλ, και άλλα αντικείμενα, εκφράσεις για αριθμούς κ.α.⁹³. Επίσης από τα μεταδεδομένα μπορεί να προέλθουν επιπλέον στοιχεία που αφορούν τη γραμματική της εν λόγω γλώσσας ή τη διαμόρφωση της φωνολογίας, της μορφολογίας και στη διαμόρφωση λεξικών. Συνήθως είναι σε μορφή αρχείων κειμένου αλλά μπορεί να πραγματοποιηθεί και σε διαφορετικού τύπου αρχεία.

2.14.2. Η σχέση με την ανάκτηση της πληροφορίας

Χάριν της ψηφιακής τεχνολογίας η γλωσσολογική τεκμηρίωση έχει αποκτήσει πολλές δυνατότητες. Μπορεί να πραγματοποιηθεί απευθείας με ψηφιακά μέσα ή με αναλογικά και στη συνέχεια ακολουθεί η ψηφιοποίηση των αρχείων αυτών.

Για την όλη διαδικασία απαιτούνται συγκεκριμένα βήματα τα οποία σύμφωνα με τον Peter K. Austin είναι τα ακόλουθα⁹⁴:

- Η εγγραφή αρχείων video, ήχου, εικόνας και κειμένου που περιέχουν, συζητήσεις, συνεντεύξεις, ομιλίες κτλ.
- Η ψηφιοποίηση δεδομένων που βρίσκονται σε αναλογική μορφή. Π.χ. η μετατροπή κειμένου σε ψηφιακή μορφή.
- Ανάλυση των δεδομένων που περιλαμβάνει μεταφράσεις, εύρεση μεταδιδόμενων κ.α.
- Ομαδοποίηση των αρχείων σε καταλόγους και ορισμός δικαιωμάτων πρόσβασης και χρήσης σε αυτά
- Δημοσίευση και διανομή του υλικού σε διάφορες μορφές.

⁹³ οι λαϊκές ταξινομίες για τα φυτά, τα ζώα και τα μουσικά όργανα και στυλ, και άλλα αντικείμενα, εκφράσεις για αριθμούς και μέτρα, αλλά και μορφολογικά παραδείγματα

⁹⁴ Austin, P. K. (2006). Data and language documentation. *Essentials of language documentation*, 87-112.

Μετά την καταγραφή των δεδομένων ακολουθεί η επεξεργασία των δεδομένων η οποία ενδεχομένως και να απαιτεί περισσότερο χρόνο και από την καταγραφή και αφού ολοκληρωθεί η διαδικασία της επεξεργασίας ακολουθεί η παρουσίαση και δημοσίευση του έργου. Η ψηφιακή τεχνολογία δίνει πολλές δυνατότητες και για τα τρία παραπάνω στάδια κυρίως όσον αφορά τον τρόπο αποθήκευσης , αρχειοθέτησης και παρουσίασης των δεδομένων.

Επιπλέον υπάρχει η δυνατότητα αποθήκευσης μεγάλης ποσότητας ψηφιακών δεδομένων και πολλές δυνατότητες επεξεργασίας και οργάνωσής τους σε ψηφιακές βάσεις δεδομένων⁹⁵ καθώς και η πιθανή οργάνωσή της σε ψηφιακές βιβλιοθήκες. Η ψηφιακή μορφή των δεδομένων κάνει ευκολότερη και την ανάκτηση. Πάρα πολύ σημαντικό είναι να κρατηθεί back up και να υπάρχουν αρκετά αντίτυπα καθώς η ύπαρξη πολλών αντιτύπων αυξάνει τις πιθανότητες διάσωσης των δεδομένων αυτών σε περίπτωση που κάποιο από τα αντίτυπα καταστραφεί⁹⁶ αυτό σημαίνει ότι υπάρχει η δυνατότητα ανάκτησης των δεδομένων αυτών και από ερευνητές στο μέλλον.

2.14.2.1. Η Αρχειοθέτηση

Τα αρχεία ήχου μπορούν να αποθηκευτούν και αρχειοθετηθούν σε διάφορους τύπους αρχείων όπως wav, ενώ τα αρχεία video σε mpeg4. Όσον αφορά τα αρχεία κειμένου ενώ η καταγραφή τους μπορεί να γίνει σε αρχεία Word, XLS, FMpro, Shoebox /Toolbox, η αρχειοθέτηση τους μπορεί να γίνει σε μορφή αρχείου xml. Η γλώσσα αυτή με τη χρήση των ετικετών (tags) δίνει τη δυνατότητα αποθήκευση μεγάλης ποσότητας δεδομένων. Επιπλέον για την παρουσίαση των δεδομένων χρησιμοποιείται η γλώσσα HTML (HyperText Markup Language), η οποία επίσης μπορεί να χρησιμοποιηθεί και για την αποθήκευση των δεδομένων⁹⁷. Κατά συνέπεια η ψηφιακή τεχνολογία διευκολύνει πολύ την γλωσσολογική τεκμηρίωση παρέχοντας της πολλές δυνατότητες.

⁹⁵ Woodbury, A. C. (2003). Defining documentary linguistics. *Language documentation and description*, 1(1), 35-51.

⁹⁶ Austin, P. K. (2006). Data and language documentation. *Essentials of language documentation*, 87-112. σελ.3.

⁹⁷ Bird, S., & Simons, G. (2003). Seven dimensions of portability for language documentation and description. *Language*, 557-582.

2.15. Η Εκμάθηση Γλώσσας

Η εκμάθηση της γλώσσας περιλαμβάνει τους τρόπους και τις διεργασίες μέσω των οποίων καταφέρνει ένας άνθρωπος να αποκτήσει τη μητρική του γλώσσα. Η μητρική γλώσσα είναι η πρώτη γλώσσα με την οποία έρχεται σε επαφή ένας άνθρωπος και υιοθετείται για την υπόλοιπη ζωή του ως η γλώσσα ομιλίας, σκέψης και έκφρασης των συναισθημάτων.

Η εκμάθηση της 1^{ης} γλώσσας εκτός από το νόημα των λέξεων και των πραγμάτων περιλαμβάνει και επικοινωνιακές δομές και κανόνες, όπως για παράδειγμα πότε πρέπει να σταματάει κάποιος τον λόγο και να συνεχίζει ο επόμενος, το ότι δεν πρέπει να μιλάμε όλοι μαζί και το ότι σε ένα διάλογο υπάρχουν ρόλοι.

Μέσα από διάφορα πειράματα που έχουν πραγματοποιηθεί, προκύπτει το συμπέρασμα ότι υπάρχει μια κρίσιμη περίοδος κατά την πρώιμη παιδική ηλικία κατά την οποία πρέπει να επιτευχθεί η κατάκτηση της μητρικής γλώσσας. Σε περίπτωση που δεν γίνει αυτό η γλωσσική ανάπτυξη είναι προβληματική⁹⁸. Αυτό αποδεικνύεται και από πολλά αληθινά περιστατικά άγριων παιδιών τα οποία βρέθηκαν στο δάσος σε μεγάλη ηλικία χωρίς να μπορούν να μιλήσουν. Οι προσπάθειες που έγιναν προκειμένου να μάθουν να μιλούν ήταν αποτυχημένες. Τα παιδιά αυτά εκτός από τη δυσκολία που υπάρχει να μάθουν να μιλούν είναι πολύ δύσκολο και να ενταχθούν στο κοινωνικό σύνολο. Από αυτό προκύπτει το συμπέρασμα ότι υπάρχει σχέση μεταξύ εκμάθησης μητρικής γλώσσας και κοινωνικοποίησης⁹⁹.

Πολύ μεγάλης σημασίας μπορούν να θεωρηθούν οι 12 πρώτοι μήνες ζωής ενός παιδιού κατά τους οποίους παράγει κάποιους ήχους και ειδικότερα προσπαθεί να επαναλάβει τους ήχους που ακούει.

2.15.1. Βασικές κατηγορίες της εκμάθησης της γλώσσας

Η εκμάθηση της μητρικής γλώσσας χωρίζεται σε δύο κατηγορίες την αντίληψη του λόγου και την παραγωγή λόγου.

⁹⁸ Πίτα, Ρ. (1998). Ψυχολογία της γλώσσας. Αθήνα: Ελληνικά Γράμματα. Σελ.107

⁹⁹ Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization. *Culture theory: Essays on mind, self, and emotion*, 276-320.

2.15.1.1. Η αντίληψη του λόγου (speech perception)

Η αντίληψη του λόγου περιλαμβάνει την ερμηνεία, κατανόηση και αναγνώριση των ήχων της γλώσσας από έναν άνθρωπο. Η διαδικασία της ομιλίας σ' έναν άνθρωπο ξεκινάει με τη λήψη του πρώτου ήχου που ακούει όταν γεννιέται και τη διαδικασία της ακρόασης.

Τα νεογέννητα παιδιά έχουν τη δυνατότητα να διακρίνουν ήχους και μπορούν να ξεχωρίσουν γλωσσικούς από μη γλωσσικούς ήχους που σημαίνει ότι αντιλαμβάνονται τον ήχο της ομιλίας. Επίσης μπορούν να ξεχωρίσουν μια αντρική φωνή από μια γυναικεία καθώς επίσης την φωνή της μητέρας του από τις υπόλοιπες φωνές. Τέλος σε πειράματα που πραγματοποιήθηκαν σε παιδιά ηλικίας 1 έως 4 ετών διαπιστώθηκε ότι υπήρχε η δυνατότητα να διακρίνουν τα σύμφωνα /b/ και /p/ ενηλίκων ατόμων¹⁰⁰.

Υπάρχουν δύο απόψεις όσον αφορά την ανάπτυξη του λόγου. Η πρώτη είναι η ορθολογιστική άποψη σύμφωνα με την οποία η ικανότητα της απόκτησης της γλώσσας είναι έμφυτη, δηλαδή προϋπάρχει κατά τη γέννηση του ανθρώπου (Πλάτωνας, Descartes, Chomsky). Η δεύτερη είναι η άποψη σύμφωνα με την οποία υπάρχει συσχέτιση γνώσης – εμπειρίας και άρα η γνώση αποκτάται με την εμπειρία (Locke, Hume).

2.15.1.2. Η παραγωγή λόγου (speech production)

Είναι η διαδικασία με την οποία οι σκέψεις μας μετατρέπονται σε λόγο μέσα από την επιλογή του κατάλληλου λεξιλογίου, γραμματικής και συντακτικού.

Η κατάκτηση της μητρικής γλώσσας απαιτεί την εκμάθηση του λεξιλογίου της γλώσσας, τη δυνατότητα σωστής σύνταξης και κατανόησης της σημασίας κάθε λέξης¹⁰¹. Επειδή η μητρική γλώσσα είναι η πρώτη γλώσσα του ανθρώπου είναι και αυτή με την οποία σκέφτεται να εκφράζεται. Οποιαδήποτε δεύτερη ή τρίτη γλώσσα είναι δύσκολο να αφομοιωθεί στον ίδιο βαθμό καθώς η εκμάθηση της μητρικής

¹⁰⁰ Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303-306.

¹⁰¹ Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, 22(1), 1-38.

γλώσσας πραγματοποιείται στην πρώιμη παιδική ηλικία κατά την οποία ο άνθρωπος αναπτύσσεται νοητικά και σωματικά.

2.15.2. Επίπεδα γλωσσικής ανάπτυξης

Η γλωσσική ανάπτυξη ενός παιδιού θα πρέπει να ολοκληρωθεί σε τρία επίπεδα : φωνολογικό, συντακτικό και σημασιολογικό.

2.15.2.1. Η Κατάκτηση της φωνολογίας

Τα φωνητικά όργανα ενός νεογέννητου παιδιού όπως ο λάρυγγας, ο φάρυγγας και η γλώσσα δεν έχουν το ίδιο μέγεθος με εκείνα των ενηλίκων γι' αυτό και οι ήχοι που παράγει ένα παιδί είναι διαφορετικοί. Οι αλλαγές ξεκινούν στην ηλικία των 4 μηνών και ολοκληρώνονται όταν το παιδί γίνει 3 χρονών όπου αλλάζει θέση ο λάρυγγας.

Ο τρόπος επικοινωνίας των παιδιών αρχικά είναι το κλάμα που δηλώνει πόνο, πείνα, αδιαθεσία. Από τεσσάρων μηνών το παιδί ξεκινάει να αρθρώνει συλλαβές και να κάνει φωνητικό παιχνίδι. Όταν πάει έξι μηνών και μέχρι την ηλικία των 12 μηνών ξεκινάει το βάβισμα δηλαδή η παραγωγή συνεχόμενων συλλαβών. Αρχικά οι συλλαβές είναι ίδιες και τις επαναλαμβάνει το μικρό παιδί και κατόπιν το βάβισμα περιέχει διαφορετικές συλλαβές. Όταν το παιδί φθάσει στην ηλικία του ενός έτους ξεκινάει να αρθρώνει τις πρώτες του λέξεις, αλλά συνεχίζει να αρθρώνει μπερδεμένες συλλαβές, ενώ κατανοεί ονόματα αντικειμένων. Από την ηλικία αυτή το λεξιλόγιο του παιδιού συνεχώς αυξάνεται και έως την ηλικία των 3 ετών μπορεί να φθάνει και τις 1000 λέξεις.

2.15.2.2. Η Κατάκτηση της σύνταξης

Η κατάκτηση της σύνταξης γίνεται σταδιακά. Από την ηλικία των 12 έως 18 μηνών το παιδί παράγει μικρές προτάσεις με λίγες λέξεις οι οποίες είναι κυρίως ουσιαστικά και πολλές φορές είναι κατανοητές μόνο από τους οικείους του παιδιού. Μέχρι το τρίτο έτος της ηλικίας ένα παιδί μπορεί να χρησιμοποιεί προτάσεις οι οποίες

περιέχουν τουλάχιστον τρεις λέξεις και μέχρι και το 5 και έκτο έτος της ηλικίας η ανάπτυξη των συντακτικών κανόνων στο λόγο ενός παιδιού είναι ραγδαία.

2.15.2.3. Η Κατάκτηση της σημασιολογίας

Υπάρχει η αντίληψη ότι πρώτα ένα παιδί καταλαβαίνει το νόημα μιας λέξης και μετά τη χρησιμοποιεί. Βέβαια αρχικά η σημασία που αποδίδει ένα παιδί σε μια λέξη μπορεί να μην είναι η ίδια με αυτήν που αποδίδουν οι ενήλικοι γι' αυτό το λόγο η σημασιολογική ανάπτυξη ενός ανθρώπου θεωρείται ολοκληρωμένη στην ηλικία των 8. Παρόλα αυτά ο άνθρωπος συνεχίζει να εμπλουτίζει το λεξιλόγιο του και αργότερα με νέες λέξεις και σημασίες. Όσο πιο πλούσιο είναι το λεξιλόγιο ενός παιδιού τόσο πλουσιότερος είναι και ο λόγος του.

2.15.3. Θεωρητικές προσεγγίσεις

Παρόλο που έχουν γίνει πάρα πολλές έρευνες είναι πάρα πολύ δύσκολο να εντοπιστούν οι μηχανισμοί που δίνουν τη δυνατότητα στα παιδιά να ξεχωρίζουν λέξεις, να αποκτήσουν γραμματικές και συντακτικές δεξιότητες ώστε να κατανοήσουν και να παράγουν γλώσσα.

Κάποιες από τις επιστημονικές εξηγήσεις που έχουν δοθεί είναι οι ακόλουθες:

- Κατά τη θεωρία των συμπεριφοριστών και του Skinner το εξωτερικό περιβάλλον και η αλληλεπίδραση του παιδιού με τα ερεθίσματα του περιβάλλοντος το βοηθούν να μαθαίνει λέξεις με τη συμβολή της θετικής και της αρνητικής επιβράβευσης.
- Άλλη εξήγηση έδωσε ο Noam Chomsky ο οποίος υποστήριξε ότι τα παιδιά διαθέτουν ένα σύστημα γλωσσικής γνώσης το οποίο τους επιτρέπει να μπορούν να αναλύουν διάφορα γλωσσικά ερεθίσματα του περιβάλλοντος τους. Δηλαδή διαθέτουν όλους τους απαραίτητους μηχανισμούς και σε συνδυασμό με τα κατάλληλα ερεθίσματα, η γλώσσα προκύπτει την κατάλληλη στιγμή. Επίσης το γεγονός ότι μπορούν και μαθαίνουν γρήγορα

τη μητρική τους γλώσσα σημαίνει ότι υπάρχουν οι βιολογικές καταβολές για την ανάλυση της γλώσσας του περιβάλλοντος τους.

- Επίσης πολλοί υποστηρίζουν ότι άλλος ένας παράγοντας που βοηθά τα παιδιά να αποκτήσουν γλωσσικές γνώσεις είναι η μίμηση και η προσπάθεια τους να μιμηθούν τους ενήλικες. Αυτή όμως η θεωρία από μόνη της δεν επαρκεί ειδικά όσον αφορά την ανάπτυξη της σύνταξης.
- Τέλος άλλη μια εξήγηση είναι μέσω συσχετισμών μέσω των φωνητικών ομοιοτήτων σε λέξεις κάτι το οποίο πάντως δεν επαρκεί προκειμένου να εξηγήσουμε εξ ολοκλήρου την κατάκτηση της γλώσσας.
- Η γλωσσική ανάπτυξη σχετίζεται με την κοινωνική αλληλεπίδραση και την επικοινωνιακή αλληλεπίδραση σύμφωνα με τον Bruner.

2.15.3.1. Οι απόψεις του Jean Piaget

Ξεχωριστή μνεία πρέπει να γίνει στον γνωστό φιλόσοφο και ψυχολόγο Jean Piaget, σύμφωνα με τον οποίο η ανάπτυξη συντελείται μετά από τέσσερα συγκεκριμένα στάδια όπου κάθε ένα προϋποθέτει το προηγούμενο. Αυτά είναι:

- Το αισθητηριοκινητικό το οποίο διαρκεί από 0 έως 2 ετών και περιλαμβάνει τις αισθητηριακές και κινητικές λειτουργίες. Μια βασική ιδιότητα που κατακτάται την περίοδο αυτή είναι η αίσθηση ότι τα αντικείμενα δεν χάνονται όταν απομακρύνονται από το οπτικό μας πεδίο.
- Το επόμενο στάδιο είναι το προλογικό στάδιο το οποίο διαρκεί από 2 έως 6 με 7 ετών. Τη περίοδο αυτή τα παιδιά αποκτούν μια εγωκεντρική στάση και δεν υιοθετούν εναλλακτικές στάσεις
- Το επόμενο στάδιο είναι από 6 με 7 ετών έως 12 και ονομάζεται στάδιο της λογικής σκέψης. Τα παιδιά μπορούν να υιοθετούν εναλλακτικές απόψεις
- Το στάδια της τυπικής σκέψης το οποίο λαμβάνει χώρα κατά την εφηβεία και ο άνθρωπος αυτή την περίοδο μπορεί να σκεφτεί λογικά.

Σύμφωνα με τον Piaget υπάρχουν δύο βασικοί μηχανισμοί : της αφομοίωσης και της προσαρμογής. Η αφομοίωση αφορά τον τρόπο με τον οποίο προσλαμβάνονται

οι γνώσεις και η προσαρμογή ο τρόπος με τον οποίο τροποποιούνται. Επίσης ο Piaget αναγνώρισε τρία είδη πρωίμου εγωκεντρικού λόγου των μικρών παιδιών.

- Η επανάληψη ή ηχολαλία όπου τα παιδιά επαναλαμβάνουν δικούς τους ήχους ή ήχους άλλων.
- Μονολόγους όπου τα παιδιά μιλούν μόνα τους εκφράζοντας τις απόψεις τους.
- Τους ομαδικούς μονολόγους όπου ομάδες παιδιών φαίνεται ότι συμμετέχουν σε συζήτηση.

2.15.4. Το Σύστημα Ανταλλαγής Δεδομένων Παιδικής γλώσσας

Το 1984 δημιουργήθηκε από τον Brian MacWhinney και την Catherine Snow το Σύστημα Ανταλλαγής Δεδομένων Παιδικής Γλώσσας» CHILDES (Child Language Data Exchange System) με σκοπό τη συλλογή και ανάλυση των γλωσσικών δεδομένων του αυθόρμητου παιδικού λόγου τα οποία συλλέγονται σε πραγματικές φυσικές καταστάσεις¹⁰². Το CHILDES αποτελεί κεντρικό αποθετήριο δεδομένων παιδικής γλώσσας, αναλύσεις συνομιλιών και δεδομένα από παιδιά.¹⁰³

Περιλαμβάνει πλήθος δεδομένων σε περισσότερες από 20 γλώσσες, τα οποία περιέχουν γλωσσικά δεδομένα από παιδιά, από άτομα που βρίσκονται σε κατάσταση αφασίας αλλά και από ενήλικες¹⁰⁴. Επιπλέον προκειμένου να διευκολυνθεί η ανταλλαγή δεδομένων και να αυτοματοποιηθεί η διαδικασία ανάλυσης των δεδομένων έχουν αναπτυχθεί διάφορα εργαλεία όπως το CHAP για αντιγραφή και κωδικοποίηση των δεδομένων και το CLAN για την ανάλυση των δεδομένων.

Χρησιμοποιώντας τη βάση δεδομένων CHILDES, ένας ερευνητής μπορεί να έχει πρόσβαση σε δεδομένα από μια σειρά ερευνητικών προγραμμάτων. Το περισσότερο υλικό αφορά φωνητικά δεδομένα σε αγγλική γλώσσα και είναι οργανωμένα σε καταλόγους βάσει του ονόματος του ερευνητή που πραγματοποίησε την έρευνα.

¹⁰² MacWhinney, B. (1996). The CHILDES system. *American Journal of Speech-Language Pathology*, 5(1), 5-14

¹⁰³ MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of child language*, 12(2), 271-295.

¹⁰⁴ MacWhinney, B. (1996). The CHILDES system. *American Journal of Speech-Language Pathology*, 5(1), 5-14.

Στη βάση δεδομένων υπάρχουν και καταγραφές που αφορούν και άλλες γλώσσες εκτός από τα αγγλικά . Υπάρχουν καταγραφές στα Γερμανικά, σε αφρικανικές γλώσσες, στα Εβραϊκά, στα Ουγγρικά κ.α. . Όλες οι εγγραφές αυτές έχουν περάσει από ειδικά τεστ εγκυρότητας και είναι κωδικοποιημένες με τον κατάλληλο τρόπο. Μάλιστα τα δεδομένα αυτά είναι ιδιαίτερα χρήσιμα για τους γλωσσολόγους οι οποίοι χρειάζεται να αναλύουν δεδομένα και σε άλλες γλώσσες εκτός από τα αγγλικά. Επιπλέον υπάρχουν καταγραφές από ήχους και ομιλία δίγλωσσων μικρών παιδιών ηλικιών 1 έως 7 ετών .

Η πρόσβαση στα δεδομένα της βάσης γίνεται από τον ιστότοπο <http://childes.psy.cmu.edu>. για τους ερευνητές που δεν μπορούν να επισκευτούν το αποθετήριο και είναι ελεύθερη υπό την προϋπόθεση ότι θα προηγηθεί η απαραίτητη εγγραφή μέλους.

2.16. Ιστορική Γλωσσολογία

Είναι γεγονός ότι οι γλώσσες με την πάροδο του χρόνου μεταβάλλονται και αλλάζουν κάποια στοιχεία τους, λόγω κοινωνικών καταστάσεων, της μετανάστευσης, της εισαγωγής ξενόφερτων λέξεων κ.τ.λ. Αυτό είναι ορατό σε πολλές γλώσσες όπως για παράδειγμα η σύγχρονη Αγγλική Γλώσσα που έχει σημαντικές διαφορές με την Αγγλική Γλώσσα του 15^{ου} αιώνα, η Ελληνική Γλώσσα η οποία προέρχεται από την Αρχαία Ελληνική η οποία όμως με το πέρασμα των χρόνων έχει υποστεί φωνολογικές και μορφολογικές αλλαγές, ενώ αντίστοιχες αλλαγές υπάρχουν και μεταξύ της παλαιάς και νέας Γαλλικής και Γερμανικής Γλώσσας.

Η Ιστορική Γλωσσολογία εξετάζει αυτές τις μεταβολές, καθώς επίσης τη σχέση που μπορεί να έχουν κάποιες γλώσσες μεταξύ. Επίσης η ιστορική γλωσσολογία εξετάζει και την ετυμολογία των λέξεων δηλαδή την ιστορία και την προέλευση τους. Ένας από τους σκοπούς της Ιστορικής Γλωσσολογίας είναι ο εντοπισμός συγγένειας και κοινών χαρακτηριστικών μεταξύ γλωσσών. Με αυτό τον τρόπο δημιουργούνται ομάδες γλωσσών που ανήκουν σε μια πρωτογλώσσα. Ένα άλλο αντικείμενο της Ιστορικής Γλωσσολογίας είναι η ανακατασκευή μιας παλαιότερης εκδοχής μιας γλώσσας.

Οι αλλαγές που μπορεί να υποστεί μια γλώσσα με την πάροδο του χρόνου μπορεί να είναι μικρές ή μεγαλύτερες ανάλογα τη γλώσσα ή τη χρονική περίοδο. Μπορεί να οφείλονται σε κάποιο ιστορικό γεγονός, σε γεωγραφικές μετακινήσεις πληθυσμών, στην επικοινωνία και τη γειτνίαση μεταξύ λαών και σε πολιτιστικές και πολιτισμικές επιδράσεις.

Ο σημαντικός γλωσσολόγος Ferdinand de Saussure έκανε διάκριση της γλωσσολογίας σε σύγχρονη και διαχρονική . Η διαχρονική εξετάζει τις αλλαγές σε κάποια γλώσσα σε διαδοχικές περιόδους. Αντίθετα η σύγχρονη εξετάζει μια γλώσσα σε μια συγκεκριμένη χρονική περίοδο η οποία συνήθως είναι το παρόν. Η διαχρονική μελέτη των γλωσσών είναι αντικείμενο της ιστορικής γλωσσολογίας αντίθετα η συγχρονική μελέτη αφορά τους υπόλοιπους γλωσσολογικούς κλάδους.

2.16.1. Μέθοδοι Ιστορικής Γλωσσολογίας

Υπάρχουν δύο βασικοί τρόποι έρευνας της ιστορικής γλωσσολογίας οι οποίοι είναι εξίσου σημαντικοί. Πρόκειται για τη συγκριτική μέθοδο και τη μέθοδο της εσωτερικής ανασυγκρότησης .

2.16.1.1. Η Ιστορικοσυγκριτική Γλωσσολογία

Η συγκριτική μέθοδος έχει ως σκοπό όπως το φανερώνει και το όνομα τη σύγκριση μεταξύ δύο ή περισσότερων γλωσσών προκειμένου να αποδειχθεί ότι υπάρχει κάποια συσχέτιση, συγγένεια και ομοιότητα μεταξύ τους και κατά συνέπεια ότι προέρχονται από κάποια προγενέστερη πρωτότυπη γλώσσα.

Θεμελιώθηκε στις αρχές του 19^{ου} αιώνα από τον Franz Bopp και αμφισβητήθηκαν για πρώτη φορά τα πρωτεία των δύο κλασικών γλωσσών: της Αρχαίας Ελληνικής και της Λατινικής Γλώσσας και αντίθετα τα πρωτεία δόθηκαν στη Σανσκριτική. Σύμφωνα με την ιστορικοσυγκριτική μέθοδο υπάρχουν μια σειρά από τεχνικές οι οποίες μπορούν να χρησιμοποιηθούν για να εξαχθούν τα χαρακτηριστικά μιας παλαιότερης γλώσσας.

Η ιστορικοσυγκριτική γλωσσολογία επηρεάστηκε από τη θεωρία του Saussure και του δομισμού καθώς και από τη θεωρία του Τσόμσκι και την Καθολική Γραμματική. Πλέον έχει επικρατήσει ο όρος ιστορική γλωσσολογία και το ενδιαφέρον επικεντρώνεται σε αλλαγές και μεταβολές ενός συστήματος ή μια κατάστασης μιας γλώσσας.

Βασική στην ιστορικοσυγκριτική μέθοδο είναι η σύγκριση μεταξύ των φωνημάτων αντίστοιχων λέξεων μεταξύ γλωσσών.

2.16.1.2. Η μέθοδος της Εσωτερικής Ανασυγκρότησης

Η μέθοδος της εσωτερικής ανασυγκρότησης αφορά την αποκατάσταση και κατασκευή μιας παλαιότερης μορφής μιας γλώσσας ή η κατασκευή μιας πρωτογλώσσας βάσει γλωσσών που ανήκουν στην ίδια ομάδα. Η μέθοδος της εσωτερικής ανασυγκρότησης συγκρίνει κυρίως τις εσωτερικές αλλαγές σε μια γλώσσα και όχι σε σχέση με μια άλλη γλώσσα ή μια ομάδα γλωσσών.

2.16.2. Βασικές Αρχές της ιστορικής γλωσσολογίας

- ✚ Μια από τις βασικές αρχές της ιστορικής γλωσσολογίας είναι η συνέχεια. Συνέχεια υπάρχει σε μια γλώσσα όταν έχει μένει ίδιος ο κορμός της και τα βασικά δομικά της στοιχεία ανεξάρτητα των αλλαγών που μπορεί να πραγματοποιηθούν με το πέρασμα των χρόνων.
- ✚ Άλλη ιδιότητα είναι η αρχή της συμβατικότητας των γλωσσικών σημείων, δηλαδή της διαφορετικότητας των λέξεων μιας γλώσσας από τις αντίστοιχες σε σημασία λέξεις μιας άλλης γλώσσας. Αρκετές φορές συναντάμε ομοιότητες σε λέξεις γλωσσών που δεν ανήκουν στην ίδια οικογένεια. Αυτό ενδεχομένως και να οφείλεται και σε δανεισμό λέξεων από μια γλώσσα σε μια άλλη γλώσσα. Σε περίπτωση που πρόκειται για λέξεις γλωσσών με κοινή καταγωγή η ομοιότητα πιθανόν να οφείλεται στη συγγένεια αυτή. Υπάρχουν

περιπτώσεις όπου η ομοιότητα υφίσταται μόνο στη μορφή και όχι στη σημασία για παράδειγμα η λέξη «ja» στη γερμανική γλώσσα σημαίνει ναι ενώ η λέξη «για» στην ελληνική γλώσσα έχει εντελώς διαφορετική σημασία καθώς είναι πρόθεση που δηλώνει αιτία, σκοπό κτλ.

🌈 Τέλος άλλη ιδιότητα είναι η αρχή της σταθερότητας/ κανονικότητας των φωνητικών μεταβολών. Κάθε γλώσσα έχει δικούς της φωνητικούς κανόνες όσον αφορά τα φωνήματα της. Αν μια άρθρωση διατηρείται σε μια λέξη μιας γλώσσας τότε διατηρείται ίδια σε όλες τις λέξεις μιας γλώσσας.

2.16.3. Χρονολόγηση μεταβολών

Μια άλλη σημαντική εργασία που γίνεται στη ιστορική γλωσσολογία είναι η χρονολόγηση των γλωσσικών μεταβολών καθώς μια λέξη μπορεί να μεταβληθεί περισσότερες από μια φορές με το πέρασμα των χρόνων. Η ακριβής χρονολόγηση είναι δύσκολη γι' αυτό το λόγο χρησιμοποιείται η σχετική χρονολόγηση η οποία ορίζει τη σειρά με την οποία πραγματοποιήθηκαν οι αλλαγές. Στη χρονολόγηση των αλλαγών σε μια γλώσσα μπορεί να συμβάλει η χρονολόγηση παλαιών πρωτότυπων επιγραφών. Αυτό μπορεί να γίνει και με τη βοήθεια σύγχρονων τεχνικών μεθόδων. Μια τέτοια μέθοδος είναι η χρονολόγηση άνθρακα. Η χρονολόγηση μπορεί να βασίζεται και στα συμφραζόμενα τα οποία μπορεί να αφορούν ιστορικά γεγονότα και πρόσωπα μιας συγκεκριμένης περιόδου

Επιπλέον ανάλογα με τις αλλαγές που υφίσταται μια γλώσσα στο λεξιλόγιο της με το πέρασμα του χρόνου, οι γλωσσολόγοι χωρίζουν την ιστορία της σε διάφορα στάδια. Οι περίοδοι στις οποίες χωρίζεται η ελληνική γλώσσα είναι: η Πρωτοελληνική , η Μυκηναϊκή ,η Κλασσική Ελληνική, η Ελληνιστική, η Μεσαιωνική Ελληνική και η Νέα Ελληνική Γλώσσα.

2.16.4. Η Ετυμολογία των λέξεων

Ετυμολογία είναι η μελέτη της ιστορίας και προέλευσης μιας λέξης. Μελετά επίσης τις αλλαγές μιας λέξης με την πάροδο του χρόνου . Στην ετυμολογία χρησιμοποιείται η συγκριτική μέθοδος . Σκοπός εκτός άλλων είναι η δόμηση του λεξικού

παλαιότερων εκδοχών μιας γλώσσας αλλά και της πρωτογλώσσας στην οποία ανήκει. Παράδειγμα η ετυμολογία της λέξης ετυμολογία είναι: < *ἔτυμον*, “αληθινή έννοια” + -λογία, “μελέτη της ”, από την λέξη *λόγος*>.

2.16.5. Σχέση με την ανάκτηση της πληροφορίας

Είναι γεγονός ότι το διαδίκτυο και οι νέες τεχνολογίες έχουν φέρει επανάσταση και έχουν μπει στην καθημερινότητά μας, στον εργασιακό χώρο, στην εκπαίδευση και στην ψυχαγωγία. Το ζητούμενο είναι το κατά πόσο μπορεί η πληροφορική και το διαδίκτυο να προκαλέσει αλλαγές στη μορφή κάποιας σύγχρονης ομιλούμενης γλώσσας.

Η ορολογία που χρησιμοποιείται στις νέες τεχνολογίες είναι αγγλική αλλά παρόλα αυτά υπάρχουν και αντίστοιχες λέξεις σε άλλες γλώσσες. Στην Ελληνική Γλώσσα για παράδειγμα υπάρχουν οι λέξεις: διαδίκτυο, ηλεκτρονικό μήνυμα, ηλεκτρονική διεύθυνση, πολυμέσα, ενώ οι αντίστοιχοι αγγλικοί – διεθνείς όροι είναι internet, email, url, multimedia. Σε μεγάλο βαθμό η χρήση των ελληνικών λέξεων αγνοείται ενώ υπάρχουν λέξεις για τις οποίες μπορεί ακόμη να μην υπάρχει αντίστοιχη μετάφραση και να χρησιμοποιείται αποκλειστικά η αγγλική λέξη. Έτσι λοιπόν νέες λέξεις όπως web, selfie, i screenshot, κ.α. προστίθενται στην καθημερινότητά μας.

Το φαινόμενο αυτό ενδεχομένως να μεταφράζεται ως γλωσσική επιρροή της αγγλικής γλώσσας σε άλλες χώρες του κόσμου. Το διαδίκτυο συμβάλει στην παγκοσμιοποίηση, την γρήγορη και φθηνή επικοινωνία μεταξύ των πολιτών διαφόρων χωρών και γι’ αυτό το λόγο υπάρχει η ανάγκη ύπαρξης μιας παγκόσμιας γλώσσας η οποία είναι η αγγλική. Σε πολλές περιπτώσεις γλώσσες κρατών παραγκωνίζονται από την αγγλική γλώσσα στην ηλεκτρονική επικοινωνία. Ως χαρακτηριστικό παράδειγμα αναφέρονται οι ασιατικές γλώσσες των οποίων η χρήση στη διαδικτυακή επικοινωνία δε θεωρείται τόσο διαδεδομένη όσο θα έπρεπε να είναι..

2.16.5.1. Το Φαινόμενο των greeklish

Παρόλο που τα θετικά στοιχεία από το διαδίκτυο και τις νέες τεχνολογίες είναι πολλά για τη γνώση και την ανάπτυξη, ενδεχομένως να μην ισχύει το ίδιο για κάποιες γλώσσες που αλλάζουν με την εισαγωγή νέων λέξεων, αλλά και για κάποια αλφάβητα μη λατινικά τα οποία πολλές φορές αγνοούνται. Μέσα σ' αυτό το πλαίσιο εμφανίζεται και το φαινόμενο των greeklish το οποίο αφορά τη χρήση του λατινικού αλφάβητου για την ηλεκτρονική επικοινωνία σε ελληνική γλώσσα.

Υπάρχουν τρεις απόψεις οι οποίες διατυπώνονται ως εξής :

- Σύμφωνα με την πρώτη τα greeklish αποτελούν πολύ σοβαρή απειλή για την ελληνική γλώσσα, δεδομένου ότι χρησιμοποιούνται και από τους νέους και μπορεί να επηρεάσει αρνητικά και τον προφορικό λόγο, αλλά και το ελληνικό αλφάβητο.¹⁰⁵
- Σύμφωνα με τη δεύτερη άποψη το φαινόμενο αυτό είναι μια μόδα η οποία θα περάσει και περισσότερο οφείλεται στην τεχνολογική πρόοδο.
- Τέλος σύμφωνα με την τρίτη άποψη το φαινόμενο είναι αποτέλεσμα της παγκοσμιοποίησης και των αρνητικών συνεπειών της.¹⁰⁶

2.17. Γενική Γλωσσολογία

Η γενική γλωσσολογία συγκρίνει διάφορες γλώσσες και εξετάζει τις δομές τους προκειμένου να εντοπίσει ομοιότητες και κοινά χαρακτηριστικά γνωρίσματα αλλά και τα σημεία στα οποία διαφοροποιούνται.

Στη Γενική Γλωσσολογία είναι απαραίτητη η μελέτη διαφόρων γλωσσών και η αναζήτηση κοινών χαρακτηριστικών, ομοιοτήτων και διαφοροποιήσεων σε σημασιολογικό, πραγματολογικό, μορφολογικό και φωνολογικό επίπεδο. Η μελέτη και σύγκριση διαφορετικών γλωσσών δεν είναι κάτι πολύ εύκολο και απλό επειδή οι

¹⁰⁵ Gavrilidou, M., Rehm, G., & Uszkoreit, H. (2012). *The Greek language in the digital age*. Springer.

¹⁰⁶ Koutsogiannis, D., & Mitsikopoulou, B. (2003). Greeklish and Greekness: Trends and discourses of “glocalness”. *Journal of computer-mediated communication*, 9(1), JCMC918.

γλώσσες που ομιλούνται σήμερα παγκοσμίως είναι πάρα πολλές, αυτό έχει ως συνέπεια ότι μια συνολική σύγκριση γλωσσικών στοιχείων καθώς και συντακτικών και γραμματικών δομών είναι κάτι το οποίο απαιτεί χρόνο.

Επιπλέον στη Γενική Γλωσσολογία πολύ σημαντική είναι η δημιουργία γενικών θεωρητικών μοντέλων της δομής, των ιδιοτήτων, των βασικών λειτουργιών και των μηχανισμών που δημιουργούν διαφοροποίηση των γλωσσών μεταξύ τους καθώς και αλλαγών σε μια γλώσσα κατά τη διάρκεια της ιστορίας της.

2.17.1. Η ομιλία

Η **ομιλία** είναι ένα ανθρώπινο εργαλείο μετάδοσης μηνυμάτων, έκφρασης συναισθημάτων, επικοινωνίας το οποίο χρησιμοποιεί τη γλώσσα. Κάποια από τα βασικά χαρακτηριστικά της ομιλίας είναι:

- **Η Άρθρωση** η οποία χωρίζεται σε μικρότερες και μεγαλύτερες μονάδες οι οποίες μπορεί να έχουν σημασιολογικό περιεχόμενο, μπορεί όμως και όχι. Οι μονάδες που διαθέτουν σημασιολογικό περιεχόμενο ονομάζονται μονάδες του επιπέδου πρώτης άρθρωσης και αποτελούνται από λέξεις και μορφήματα. Οι μονάδες της δεύτερης άρθρωσης δε διαθέτουν σημασιολογικό περιεχόμενο και είναι απλά φωνήματα. Εν συνεχεία συνδυασμοί λέξεων δηλαδή μονάδων επιπέδου πρώτης άρθρωσης μπορούν να εκφράσουν συναισθήματα, να γίνουν μέσω επικοινωνίας κτλ. Το γεγονός ότι ο άνθρωπος διαθέτει τα παραπάνω δύο είδη αρθρώσεων, τον διαφοροποιεί από τα υπόλοιπα όντα της φύσης.
- **Η ασυνέχεια** η οποία υπάρχει υπό την έννοια ότι η ακόμη και σε λέξεις που μοιάζουν υπάρχει τουλάχιστον μια διαφορά σε φωνήματα ή ακόμη και στον τονισμό, που αποτελεί διαφοροποιητικό παράγοντα. Π.χ. οι λέξεις κορμί και κορμό.
- **Το αυθαίρετο του γλωσσικού σημείου.** Σύμφωνα με τον Ferdinand Saussure τα γλωσσικά σημεία όπως έχει αναφερθεί και παραπάνω είναι ένας συνδυασμός σημαίνοντος και σημαινόμενου. Σύμφωνα με τον Saussure και τα δύο είναι ψυχολογικές έννοιες και η σχέση τους είναι αυθαίρετη. Από τη συνένωση αυτών των δύο προκύπτει ένα γλωσσικό σημείο. Διαφορετικά

σημαίνοντα μπορεί να δηλώνουν το ίδιο ή περίπου το ίδιο σημαίνόμενο σε διάφορες γλώσσες ή ακόμη και στην ίδια γλώσσα. Π.χ. /rimenas/ και /voskos/. Τα μέλη μιας κοινότητας έχουν τα ίδια σημαίνοντα.

- **Η παραγωγικότητα** δηλαδή η δυνατότητα του ανθρώπου να παράγει και να κατανοεί μεγάλο αριθμό εκφωνημάτων σε ποικιλία και θεματολογία.
- **Η πολλαπλότητα εκφοράς** για την οποία είναι απαραίτητη είναι η ύπαρξη των απαραίτητων μηχανισμών κωδικοποίησης και αποκωδικοποίησης,
- **Ο φωνητικός χαρακτήρας** ο οποίος διαφοροποιείται από γλώσσα σε γλώσσα
- **Η γραμμικότητα** καθώς μια λέξη χρειάζεται ένα χρονικό διάστημα για να αρθρωθεί και ο χρόνος μπορεί να μετρηθεί σε μια διάσταση.

2.17.2. Η Γλώσσα

Γλώσσα είναι ο κοινωνικά αποδεκτός κώδικας ή ένα συμβατικό σύστημα αναπαράστασης εννοιών με την χρησιμοποίηση αυθαίρετων συμβόλων και κανόνων που διέπουν το συνδυασμό αυτών των συμβόλων. Ο όρος γλώσσα κατά κύριο λόγο χρησιμοποιεί το λόγο ως μέσω επικοινωνίας και ανθρώπινης συμπεριφοράς. Επίσης η γλώσσα εξετάζεται και ως γραμματικό σύστημα που βρίσκεται πίσω από το λόγο.

Σύμφωνα με τον **Ferdinand Saussure** αντικείμενο της γλωσσολογίας πρέπει να είναι η γλώσσα και όχι η ομιλία γι' αυτό στο έργο του κάνει διάκριση μεταξύ τους. Με τη γλώσσα επικοινωνούμε με ήχους και φθόγγους που αντιστοιχούν σε νοήματα και σημασίες όταν συνδέονται οι φθόγγοι μεταξύ τους και δημιουργούν γλωσσικά σημεία. Για τη γλώσσα είναι απαραίτητη η ύπαρξη ενός πομπού, ενός δέκτη, του μηνύματος, της κωδικοποίησης του μηνύματος, της θεματολογίας και του καναλιού επικοινωνίας.

Κάποιες από τις ιδιότητες της γλώσσας οι οποίες ξεχωρίζουν τον άνθρωπο από τα υπόλοιπα όντα είναι ότι:

- Η δυνατότητα να μπορεί ένας άνθρωπος που γνωρίζει μια γλώσσα, να μαθαίνει και άλλες γλώσσες.
- Η δυνατότητα να χρησιμοποιεί κάποιος τη γλώσσα προκειμένου να υπεκφεύγει.

- Η δυνατότητα να εξηγούμε τι είναι γλώσσα ή να μιλάμε για τη γλώσσα, χρησιμοποιώντας τη γλώσσα.

Σύμφωνα με τους Vygotsky και Piaget η πρακτική ενασχόληση των μικρών παιδιών συνοδεύεται πάντα με λόγια. Ομοίως σύμφωνα με τον Έγκελς υπάρχει σχέση μεταξύ ομιλίας και εργασίας. Η σχέση αυτή ενδεχομένως είναι και η αιτία γένεσης της γλώσσας. Υπάρχουν και άλλες θεωρίες όπως αυτή η οποία ορίζει ότι ο άνθρωπος στην προσπάθεια του να μιμηθεί τους ήχους της φύσης και των ζώων εφεύρε την ομιλία. Η εμφάνιση της γλώσσας οφείλεται και σε περιβαλλοντικές αλλαγές, οι οποίες είχαν ως επακόλουθο μια σειρά μεταβολών στην ανατομία του ανθρώπου και ειδικά σε κάποια όργανα όπως ο φάρυγγας, ο λάρυγγας, αλλά και ανθρώπινος εγκέφαλος.

Η ομιλία συνεχώς εξελίσσεται και ως εκ τούτου οι διάφορες γλώσσες αλλάζουν και ανανεώνονται με το πέρασμα του χρόνου, ενώ υπάρχουν και γλώσσες που χάνονται ή συγχωνεύονται με άλλες πιο δυνατές γλώσσες. Κάποιοι από τους λόγους εξέλιξης μιας γλώσσας είναι ο γλωσσικός δανεισμός και οι αλληλεπιδράσεις μεταξύ διαφορετικών γλωσσών λόγω ιστορικών συγκυριών και περιβαλλοντικών παραγόντων.

2.17.3. Η δομή της γλώσσας

Η γλώσσα αποτελείται από ένα σύνολο στοιχείων όπως τα φωνημάτα, τα μορφήματα, τις λέξεις και τις μεταξύ τους σχέσεις. Υπάρχουν επίσης δομικοί, γραμματικοί και συντακτικοί κανόνες οι οποίοι βοηθούν τον γλωσσολόγο να αναλύσει και να χαρακτηρίσει γλωσσικές προτάσεις αν είναι γραμματικές ή μη γραμματικές, αν υπάρχουν προβλήματα στη μορφολογία, στη σύνταξη και τη φωνολογία π.χ. η πρόταση «το παιδί τραγουδούν» δεν είναι σωστή γραμματικά και μορφολογικά.

2.17.4. Οι ομοιότητες μεταξύ διαφόρων γλωσσών

Σήμερα οι γλώσσες που ανήκουν στην ίδια οικογένεια γλωσσών όπως για παράδειγμα οι γλώσσες που ανήκουν στην Ινδοευρωπαϊκή Οικογένεια οι οποίες έχουν πολλές ομοιότητες σε όλα τα επίπεδα της γλωσσολογικής ανάλυσης όπως μορφολογικό, σημασιολογικό, φωνολογικό. Π.χ. η λέξη επτά στα ελληνικά είναι *septem* στα λατινικά, *seven* στα αγγλικά, *sieben* στα γερμανικά κτλπ. Οι ομοιότητες αυτές ονομάζονται **συστηματικές** και δεν οφείλονται σε σύμπτωση ή κάποιο δανεισμό. Από τις συστηματικές ομοιότητες προκύπτουν αντιστοιχίες χαρακτήρων. Π.χ. από συστηματικές ομοιότητες μεταξύ της ελληνικής γλώσσας και της αγγλικής προκύπτει ότι το ελληνικό /p/ αντιστοιχεί στο αγγλικό /f/ (**πατήρ**-**father**, **πτερόν**-**feather**, **πόδα**-**foot**).

Ομοιότητες μεταξύ γλωσσών μπορεί να οφείλονται και σε δανεισμό. Χαρακτηριστικό παράδειγμα το ότι την περίοδο του μεσαίωνα αρκετές σλάβικες λέξεις δανείστηκαν στην ελληνική γλώσσα. Χαρακτηριστικό παράδειγμα είναι η λέξη πέστροφα – όπου στη σέρβικη γλώσσα είναι *păstrāv*. Αρκετές φορές οι λέξεις που δανείζονται αλλάζουν στην καθομιλουμένη γλώσσα. Π.χ. η γαλλική λέξη *tracteur* η οποία έχει δανειστεί στην ελληνική γλώσσα (τρακτέρ), αλλά στην καθομιλουμένη τρακτέρι.

2.17.5. Η γλώσσα και η ομιλία ως μέσω επικοινωνίας στο διαδίκτυο

Η χρήση του λόγου στο διαδίκτυο είναι πολύ είναι πολύ μεγάλης σημασίας. Ο Παγκόσμιος αλλά και ο Σημασιολογικός Ιστός βασίζονται στη γλώσσα, το λόγο και τη γλωσσική ποικιλότητα. Κάποιες από τις χρήσεις της γλώσσας στο διαδίκτυο είναι ως μέσω έκφρασης, για να επικοινωνούμε, ως μέσω συνεργασίας, για να ενημερωνόμαστε, για εκπαιδευτικούς λόγους.

Επίσης η γλωσσική ποικιλότητα είναι απαραίτητη στο διαδίκτυο. Υπηρεσίες όπως η αναζήτηση στο διαδίκτυο και η αυτόματη μετάφραση υποστηρίζουν πολλές ξένες γλώσσες πλην της αγγλικής. Αρκετοί οργανισμοί της Ευρωπαϊκής Ένωσης δαπανούν ετησίως περίπου 1 δις για διατήρηση της πολυγλωσσίας μεταφράζοντας κείμενα και

για τη διερμηνεία της προφορικής επικοινωνίας.¹⁰⁷ Η διατήρηση της πολυγλωσσίας συμβάλλει και στη διάσωση το γλωσσών, αλλά και στη διευκόλυνση των χρηστών του διαδικτύου. Αρκεί να λάβει κάποιος υπόψη ότι βάσει μιας πρόσφατης έκθεσης της Ευρωπαϊκής Επιτροπής, το 57% του διαδικτύου χρησιμοποιούν υπηρεσίες του ηλεκτρονικού εμπορίου σε γλώσσες διαφορετικές από τη μητρική τους. Στην Ελλάδα γι' αυτό του λόγο κρίθηκε αναγκαία η ίδρυση του «Ινστιτούτου Επεξεργασίας Λόγου» το οποίο έχει βασικό σκοπό την ανάπτυξη και προώθηση της Γλωσσικής και της Πολιτιστικής Τεχνολογίας στην Ελλάδα, ώστε να είναι δυνατή η χρήση της Ελληνικής Γλώσσας σε διάφορες υπηρεσίες του διαδικτύου¹⁰⁸.

Επειδή η χρήση της γλώσσας και του λόγου είναι βασικής σημασίας στο διαδικτυακό κόσμο ο γνωστός γλωσσολόγος Ντέιβιντ Κρύσταλ εισήγαγε τον όρο της Διαδικτυακής Γλωσσολογίας¹⁰⁹. Ο κλάδος αυτός μελετά τα νέα γλωσσικά και εκφραστικά στυλ και γλωσσικές δομές που εμφανίζονται εξαιτίας του διαδικτύου αλλά και τη χρήση τηλεφώνων νέας γενιάς και άλλων σύγχρονων τεχνολογικών συσκευών¹¹⁰. Η μελέτη του Ντέιβιντ Κρύσταλ γίνεται με χρήση της συγκριτικής γλωσσολογικής μεθόδου και μέσα από αυτή τη σύγκριση προκύπτει η ύπαρξη νέου λεξιλογίου ειδικά στους νέους, το οποίο περιλαμβάνει λέξεις όπως : txt'ing, blogging κτλ. Τέλος αξιοσημείωτο είναι ότι κάποιες νέες συντομογραφίες όπως 'LOL', 'FYI' και 'OMG' από το χώρο του διαδικτύου συμπεριέλαβε το επίσημο λεξικό της Οξφόρδης του 2011 (Oxford English Dictionary)¹¹¹.

2.18. Δικανική Γλωσσολογία

Υπάρχει ξεχωριστός κλάδος της γλωσσολογίας ο οποίος εξειδικεύεται σε κείμενα νομικών υποθέσεων και ονομάζεται δικανική γλωσσολογία. Ο ίδιος υπό-κλάδος ονομάζεται αλλιώς και νομική, εγκληματολογική ή ιατροδικαστική γλωσσολογία.

¹⁰⁷ Gavrilidou, M., Rehm, G., & Uszkoreit, H. (2012). *The Greek language in the digital age*. Springer.

¹⁰⁸ <http://www.ilsp.gr/>

¹⁰⁹ https://el.wikipedia.org/wiki/Διαδικτυακή_γλωσσολογία.

¹¹⁰ <http://www.davidcrystal.com/books-and-articles>

¹¹¹ <https://www.translatemedia.com/translation-blog/internet-accelerating-language-development/>

Κάποιες από τις βασικές αρμοδιότητες της είναι : η μελέτη της γλώσσας των νομικών κειμένων, της γλώσσα που χρησιμοποιείται από δικηγόρους και δικαστές στις δίκες και ανακρίσεις και τέλος τη μελέτη της γλώσσας ως αποδεικτικό στοιχείο.

Ακόμη περιλαμβάνει τη μελέτη και ανάπτυξη της επαγγελματικής, νομικής μετάφρασης, την απόδειξη της πατρότητας ανώνυμων απειλητικών μηνυμάτων, επιστολών και της ανίχνευσης περιπτώσεων λογοκλοπής¹¹². Επιπλέον περιλαμβάνει τη μελέτη και ανάλυση σε γραπτά μηνύματα κινητού τηλεφώνου και ηλεκτρονικού ταχυδρομείου, τις περιπτώσεις λογοκλοπών σε συγγραφικά έργα όπως μυθιστορήματα, φοιτητικές εργασίες και δημοσιογραφικά άρθρα για το νόημα λέξεων και εκφράσεων.

Πολύ σημαντική είναι μελέτη ανώνυμων μηνυμάτων από εξειδικευμένους γλωσσολόγους, η οποία μπορεί να αποκαλύψει κάποια στοιχεία για την ταυτότητα του συντάκτη. Στοιχεία που λαμβάνονται υπόψη και μπορούν να οδηγήσουν στην απόδειξη της πατρότητας ενός εγγράφου ανώνυμου συντάκτη είναι: Ο γραφικός χαρακτήρας, το ύφος και η γλώσσα που χρησιμοποιείται, τα πιθανά ορθογραφικά λάθη που επαναλαμβάνονται σ' ένα κείμενο, τα σημεία στίξης και συγκεκριμένες λέξεις ή εκφράσεις που επαναλαμβάνονται. Σε περίπτωση όπου υπάρχουν ύποπτα άτομα για τη σύνταξη μηνυμάτων, ένα στοιχείο που μπορεί να βοηθήσει είναι η σύγκριση του ύφους και της γλώσσας του κειμένου με τον προφορικό λόγο του ύποπτου ατόμου. Επίσης οι δικανικοί γλωσσολόγοι εκφράζουν τις απόψεις τους για μηνύματα, ηχογραφημένες ομιλίες και ερμηνεύουν το ύφος και τον τρόπο γραφής του μηνύματος.

Η σημασία της δικανικής γλωσσολογίας είναι πολύ μεγάλη και μάλιστα σε περιπτώσεις όπου δικανικοί γλωσσολόγοι καλούνται να καταθέσουν σε δίκες, επηρεάζουν την κρίση του δικαστηρίου για την έκβαση της υπόθεσης.

¹¹² Γεωργίου, Γ. Β. (2011). Γλώσσα και ισχύς σε μια δίκη φόνου στην Κύπρο: η χρήση της Διαλέκτου και η εναλλαγή κώδικα. *Selected papers on theoretical and applied linguistics*, 19, 491-498.

2.18.1. Η γλώσσα των νομικών κειμένων

Η γλώσσα με την οποία γράφονται τα νομικά κείμενα έχει τους δικούς της κανόνες και αρχές. Για παράδειγμα υπάρχει συγκεκριμένη σύνταξη και μορφολογία. Κάποιοι από τους κανόνες που ισχύουν για τα νομικά κείμενα είναι εξής: Τα ρήματα γράφονται σε οριστική έγκλιση και όχι στην προστακτική και μάλιστα ο κανόνας αυτός δεν ισχύει μόνο για την ελληνική γλώσσα αλλά και για άλλες ξένες γλώσσες¹¹³. Επίσης τα ρήματα στις κύριες προτάσεις είναι πάντα στον ενεστώτα ενώ στις δευτερεύουσες προτάσεις μπορεί να είναι και σε άλλο χρόνο και χρησιμοποιείται συνήθως η παθητική φωνή¹¹⁴.

Πολλά νομικά κείμενα είναι δύσκολα κατανοητά από έναν μη ειδήμονα. Αυτό που τα κάνει δυσνόητα είναι η ορολογία που περιέχουν και η διατύπωση με την οποία γράφονται. Πολλά νομικά κείμενα, συμβόλαια, ακόμη και κείμενα ετικετών προϊόντων περιέχουν τόσο δυσνόητους και διαφορούμενους όρους ώστε πολλοί πολίτες να πέφτουν θύματα και να εξαπατούνται. Κάποιες φορές τα γράμματα σε συμβόλαια, σε εγγυήσεις συσκευών και σε ετικέτες είναι πολύ μικρά και είναι δύσκολο να διαβαστούν και οι νομικοί όροι που περιέχουν είναι κατανοητοί από τους νομικούς γλωσσολόγους.

2.18.2. Η γλώσσα των νομικών διαδικασιών

Επίσης στην νομική γλωσσολογία ανήκει και η γλώσσα που χρησιμοποιείται στις δικαστικές διαδικασίες, στις ανακρίσεις καθώς και σε περιπτώσεις όπου ανακρίνονται ή εξετάζονται από το δικαστήριο ειδικές κατηγορίες ατόμων όπως ανήλικοι και άτομα που υποφέρουν από ψυχικά νοσήματα.

Στα δικαστήρια πραγματοποιούνται διάλογοι μεταξύ δικαστών, μαρτύρων και κατηγορουμένων. Στους διαλόγους αυτούς οι δικαστές και δικηγόροι έχουν την απόλυτη εξουσία να πραγματοποιούν τις ερωτήσεις που θέλουν προκειμένου να επιτύχουν τον πλήρη έλεγχο του μάρτυρα. Κάποιες φορές οι ερωτήσεις είναι

¹¹³ Stolyarova, A. (2014). Πραγματολογικοί παράγοντες στη χρήση των γραμματικών κατηγοριών του ρήματος στα νεοελληνικά νομοθετικά κείμενα. *Стиль і переклад*, (1), 193-206.

¹¹⁴ Stolyarova, A. (2014). Πραγματολογικοί παράγοντες στη χρήση των γραμματικών κατηγοριών του ρήματος στα νεοελληνικά νομοθετικά κείμενα. *Стиль і переклад*, (1), 193-206.

διοφορούμενες ώστε να ο μάρτυρας ή ο κατηγορούμενος να αναγκαστεί να πέσει σε αντιφάσεις. Επίσης συνήθης τακτική είναι η επιλογή των κατάλληλων ερωτήσεων ώστε να λιγοστεύουν οι πιθανές απαντήσεις από την πλευρά του μάρτυρα.

Μεγάλη σημασία για τους ακροατές της δίκης έχει το ύφος, η γλώσσα, η κανονική ένταση της φωνής και η κανονική χωρίς διακοπές ροή του λόγου από τους μάρτυρες. Στοιχεία όπως αυτά μπορούν να κερδίσουν την εύνοια και την υποστήριξη στην κρίση των ακροατών. Αντίθετα λιγότερο πιστικοί είναι όσοι ο λόγος τους περιέχει αοριστολογίες, δισταγμό, αβεβαιότητα, χρήση των λέξεων «κύριε», «κυρία», ενισχυτικά επιρρήματα όπως «κατηγορηματικά όχι», «σίγουρα». Μεγάλη σημασία παίζει και ο χρόνος που χρησιμοποιεί κάποιος για να απαντήσει σε κάποια ερώτηση καθώς οι πιο περιεκτικές και συντομότερες απαντήσεις είναι πιο πιστικές, σε αντίθεση με τις πολύ μεγάλες διηγήσεις οι οποίες μπορεί να φαίνονται ως κατασκευασμένες και σκηνοθετημένες¹¹⁵.

2.18.3. Η λογοκλοπή

Λογοκλοπή σύμφωνα με το Γ. Μπαμπινιώτη «είναι η ιδιοποίηση με αθέμιτο τρόπο ξένης πνευματικής εργασίας»¹¹⁶. Κυρίως τη συναντάμε σε μυθιστορήματα ή φοιτητικές εργασίες και βασικό χαρακτηριστικό της είναι το γεγονός ότι αυτοί που τη διαπράττουν αποφεύγουν να αναφέρουν το όνομα του συγγραφέα ή συντάκτη του πρωτότυπου κειμένου. Χαρακτηριστική περίπτωση αποτελεί το μυθιστόρημα «Da Vinci Code» για το οποίο είχε κατηγορηθεί ο συγγραφέας του Dawn Brown για λογοκλοπή από το μυθιστόρημα «*The Holy Blood and the Holy Grail*»¹¹⁷.

¹¹⁵ Γεωργίου, Γ. Β. (2011). Γλώσσα και ισχύς σε μια δίκη φόνου στην Κύπρο: η χρήση της Διαλέκτου και η εναλλαγή κώδικα. *Selected papers on theoretical and applied linguistics*, 19, 491-498.

¹¹⁶ Λεξικό της Νέας Ελληνικής Γλώσσας

¹¹⁷ Coulthard, M., Johnson, A., & Wright, D. (2016). *An introduction to forensic linguistics: Language in evidence*. Routledge. Σελ. 3

2.18.4. Σύνδεση με ανάκτηση πληροφορίας

Σχετική με τη νομική γλωσσολογία είναι η ηλεκτρονική ταυτοποίηση εγγράφων καθώς και η ηλεκτρονική αναγνώριση λογοκλοπών. Η ηλεκτρονική ταυτοποίηση εγγράφων περιλαμβάνει την ταυτοποίηση χειρόγραφης και μηχανογραφημένης γραφής, την ταυτοποίηση υπογραφών, την ύπαρξη πιθανών νοθεύσεων και αλλοιώσεων σε κείμενα.

Βέβαια τέτοιες τεχνικές εκτός από κατάλληλο λογισμικό απαιτούν και κατάλληλο υλικό όπως μηχανήματα ψηφιακής απεικόνισης, υπέρυθρες και υπεριώδεις πηγές φωτός. Απαραίτητες είναι οι συσκευές ηλεκτροστατικής ανίχνευσης (Electrostatic Detection Devices) . Οι συσκευές αυτές επιτυγχάνουν την εμφάνιση πληροφοριών που αφορούν κάποιες δραστηριότητες του συντάκτη ενός κειμένου.

Στην περίπτωση ηλεκτρονικών μηνυμάτων για την εξέταση τους απαιτούνται οι εξής προϋποθέσεις: κατά την ανάλυσή τους να διατηρείται η ακεραιότητά, η αυθεντικότητα και η εγκυρότητα των εξαγόμενων δεδομένων. Σκοπός τέτοιων τεχνικών είναι ο εντοπισμός κρυμμένων πληροφοριών στο υπό εξέταση κείμενο και η σύγκρισή του με άλλα έγγραφα προκειμένου να διαπιστωθεί αν πρόκειται για τον ίδιο συγγραφέα¹¹⁸.

Όσον αφορά την ηλεκτρονική αναγνώριση περιπτώσεων λογοκλοπής υπάρχουν πάρα πολλές προτεινόμενες μέθοδοι και λογισμικά στο διαδίκτυο. Οι προσεγγίσεις που ακολουθούνται κυρίως είναι δύο: η πρώτη αφορά την χρήση μηχανών αναζήτησης καθώς οι περισσότερες λογοκλοπές γίνονται μέσω διαδικτύου και μηχανών αναζήτησης. Η δεύτερη προσέγγιση αφορά την απευθείας σύγκριση κειμένων για τυχών ομοιότητες.

Μια από αυτές περιγράφεται από τους Soleman, S., & Fujii, A¹¹⁹ η οποία αφορά εξέταση για τυχόν λογοκλοπές με τη βοήθεια μηχανών αναζήτησης. Η μέθοδος αυτή προτείνει την εξέταση των προτάσεων με παραπομπές αλλά και των προτάσεων δίχως παραπομπές. Δέχεται ως είσοδο το υπό εξέταση έγγραφο καθώς και έγγραφα τα οποία εξετάζονται για τυχόν ομοιότητα με το πηγαίο έγγραφο. Για κάθε ένα έγγραφο

¹¹⁸ Agarwal, A., Gupta, M., Gupta, S., & Gupta, S. C. (2011). Systematic digital forensic investigation model. *International Journal of Computer Science and Security (IJCSS)*, 5(1), 118-131.

¹¹⁹ Soleman, S., & Fujii, A. (2017, September). Plagiarism Detection Based on Citing Sentences. In *International Conference on Theory and Practice of Digital Libraries* (pp. 485-497). Springer, Cham.

που συγκρίνεται με το πηγαίο έγγραφο δίνεται ένα ποσοστό ομοιότητας το οποίο δεν πρέπει να ξεπερνάει μια συγκεκριμένη τιμή. Εκτός από λογισμικά υπάρχουν και ιστότοποι στο διαδίκτυο που μας βοηθούν να εξετάσουμε αν ένα έγγραφο είναι προϊόν λογοκλοπής όπως ο ελεγκτής λογοκλοπής της quetext¹²⁰, της smallseotools¹²¹ κ.α. .

2.19. Ανάλυση λόγου - discourse analysis

Ο λόγος είναι εργαλείο έκφρασης της ανθρώπινης νόησης και αποτελεί πρωταρχικό μέσο αντίδρασης και βασικό χαρακτηριστικό του κοινωνικού μας συστήματος. Η κατηγορία αυτή της γλωσσολογίας είναι πολύ βασική και περιλαμβάνει προσεγγίσεις για την ανάλυση του λόγου, είτε είναι γραπτός είτε προφορικός. Δηλαδή εκτός από γραπτά κείμενα εξετάζονται συζητήσεις και ομιλίες. Σκοπός της δεν είναι να αναλύσει τη δομή του κειμένου αλλά τις κοινωνικές, ανθρωπολογικές και ψυχολογικές προεκτάσεις του. Στην ανάλυση του λόγου εξετάζεται τι περιέχει ένα κείμενο αλλά και τι δεν περιέχει και απουσιάζει από αυτό.

Κεντρικό ρόλο στην ανάλυση του λόγου έχει παίξει ο φιλόσοφος Michel Foucault . Σύμφωνα με τον Foucault οι λόγοι είναι σύνολα αποφάνσεων όπως διατυπώσεων μιας άποψης, μιας ιδέας, μιας γνώμης. Τα σύνολα αυτά ρυθμίζονται από κανόνες και σχετίζονται με την υποκειμενικότητά μας.

Σύμφωνα με τους (Wetherell, M., Taylor, S., & Yates, S. J. (2001)) μεγάλο ενδιαφέρον παρουσιάζουν για τους αναλυτές του λόγου διάφορα κείμενα όπως : πολιτικοί λόγοι, συζητήσεις σε chatrooms , συνεντεύξεις ασθενών κ.α.¹²² Ο λόγος και οι εκφράσεις που εμπεριέχει διαφέρουν ανάλογα το επάγγελμα, την ιδιότητα και την κοινωνική ομάδα στην οποία ανήκει κάποιος. Για παράδειγμα διαφορετικές εκφράσεις χρησιμοποιεί ένας γιατρός σ' έναν ασθενή, ένας πολιτικός απευθυνόμενος στο λαό, ένας εκπαιδευτικός σε μια σχολική τάξη. Από το λόγο ενός ατόμου

¹²⁰ <https://www.quetext.com/>

¹²¹ <https://smallseotools.com/plagiarism-checker/>

¹²² Wetherell, M., Taylor, S., & Yates, S. J. (2001). *Discourse theory and practice: A reader*. Sage.

μπορούμε να αντλήσουμε πληροφορίες όπως το επίπεδο μόρφωσης του, τα συναισθήματα του, το επάγγελμα του, την ηλικία του και την ιδεολογία του.

Ένα συγκεκριμένο γεγονός μπορεί να έχει διαφορετικό νόημα για έναν άνθρωπο ο οποίος μπορεί να το βλέπει υπό διαφορετική σκοπιά. Αυτό το γεγονός μεταφράζεται και με διαφορετικές δράσεις του συγκεκριμένου ανθρώπου και κατά συνέπεια οδηγεί στο συμπέρασμα ότι η γλώσσα δομεί ολόκληρη τη κοινωνία.¹²³ Αυτό έρχεται σε συμφωνία με τη θεωρία του Ferdinand de Saussure σύμφωνα με την οποία η σχέση μεταξύ σημαίνοντος και σημαινόμενου είναι αυθαίρετη και είναι αποτέλεσμα κοινωνικών συμβάσεων καθώς η γλώσσα δεν καθορίζεται από την αλήθεια στην οποία αναφέρεται.

Κάποια πολύ σημαντικά στοιχεία στην ανάλυση του λόγου είναι :το πλαίσιο στο οποίο τοποθετείται ο λόγος όπως π.χ. είδηση, αφήγηση. Οι δείκτες του λόγου οι οποίοι είναι μικρές συγκεκριμένες λέξεις ή επιφωνήματα τα οποία σπάνε την ομιλία μας σε μέρη και δείχνουν τη σχέση μεταξύ των τμημάτων. Όταν συγκρούονται δύο ή περισσότεροι λόγοι έχουμε ανταγωνισμό.

2.19.1. Κριτική ανάλυση του λόγου

Η κριτική ανάλυση του λόγου εξετάζει το λόγο μέσα από τις κοινωνικές διαστάσεις του και τον τρόπο με τον οποίο συνδέονται η γλώσσα, η κοινωνία και η ιδεολογία¹²⁴. Στην κριτική ανάλυση λόγου το κέντρο βάρους τοποθετείται στο νόημα το οποίο προκύπτει. Γι' αυτό το λόγο δεν πρέπει οι φράσεις να εξετάζονται μεμονωμένα αλλά σε συνδυασμό με τις υπόλοιπες προτάσεις προκειμένου να υπάρχει ολοκληρωμένο νόημα. Επίσης μεγάλης σημασίας είναι και το κοινωνικό, πολιτισμικό και ιστορικό πλαίσιο μέσα στο οποίο τοποθετείται ένας λόγος. Συνήθως η κριτική ανάλυση του

¹²³ Phillips, L., & Jorgensen, M. W. (2009). Ανάλυση Λόγου. Θεωρία και Μέθοδος. Α. Κιουπκιολής, Αθήνα: Παπαζήσης.

¹²⁴ Στάμου, Α.Γ. (2014). Η κριτική ανάλυση του: Μελετώντας τον ιδεολογικό ρόλο της γλώσσας. Στο Ανάλυση λόγου: Θεωρία και εφαρμογές, Μ. Γεωργαλίδου, Μ. Σηφianού, Β. Τσάκωνα (επιμ.). Νήσος: Αθήνα, 149-187.

λόγου επικεντρώνεται σε κείμενα που προέρχονται από τα ΜΜΕ, την πολιτική, την εκπαίδευση και άλλα δημόσια έγγραφα¹²⁵.

Σύμφωνα με τον Fairclough ο λόγος περιλαμβάνει τρεις λειτουργίες την ταυτιστική, τη σχεσιακή και την ιδεοληπτική¹²⁶. Η ταυτιστική λειτουργία αντιστοιχεί στην ταυτότητα/ταυτότητες που αποδίδονται σε κάποιο πρόσωπο μέσα από το κείμενο, η σχεσιακή με τις πιθανές κοινωνικές σχέσεις οι οποίες προκύπτουν μεταξύ των πρωταγωνιστών του κειμένου και η ιδεοληπτική με τη γνώσεις και τα νοήματα.

Η ανάλυση του λόγου γίνεται σε περιγραφικό, επεξηγηματικό και ερμηνευτικό επίπεδο. Το περιγραφικό επίπεδο αφορά τη χρήση λεξιλογίου, γραμματικής και κειμενικής δομής, το ερμηνευτικό επίπεδο τη σύνδεση του κειμένου με το γλωσσικό και εξωγλωσσικό του περιβάλλοντος, το επεξηγηματικό επίπεδο σχετίζει το κείμενο με το ευρύτερο κοινωνικοπολιτικό πλαίσιο. Σύμφωνα με τον Fairclough η ανάλυση των γλωσσικών χαρακτηριστικών ενός κειμένου εξαρτάται από την ανάλυση της ρηματικής πρακτικής.

Σημαντική στην κριτική ανάλυση του λόγου είναι και η θεωρία του γλωσσολόγου Michael Alexander Halliday στο έργο του οποίου δίνεται έμφαση στο ρόλο των ρημάτων. Έτσι λοιπόν διακρίνει τρεις διαδικασίες οι οποίες είναι :

- **Οι υλικές διαδικασίες** στις οποίες συναντάμε ρήματα που σχετίζονται με δράσεις. Το υποκείμενο είναι ο δράστης (actor) ο οποίος κάνει κάτι ή υποβάλλεται σε μια διαδικασία, ενώ ο δέκτης της διαδικασίας είναι ο στόχος (goal).
- **Οι νοητικές διαδικασίες** στις οποίες περιλαμβάνονται ρήματα που αφορούν συναισθήματα και την αντίληψη. Το υποκείμενο είναι «ο αισθανόμενος» ο οποίος πραγματοποιεί μια διαδικασία το περιεχόμενο της διαδικασίας ονομάζεται «φαινόμενο».

¹²⁵ Στάμου, Α.Γ. (2014). Η κριτική ανάλυση του: Μελετώντας τον ιδεολογικό ρόλο της γλώσσας. Στο Ανάλυση λόγου: Θεωρία και εφαρμογές, Μ. Γεωργαλίδου, Μ. Σηφianού, Β. Τσάκωνα (επιμ.). Νήσος: Αθήνα, 149-187.

¹²⁶ Phillips, L., & Jorgensen, M. W. (2009). Ανάλυση Λόγου. Θεωρία και Μέθοδος. Α. Κιουπκιολής, Αθήνα: Παπαζήσης.

- **Οι συσχετικές** στις οποίες τα ρήματα αποδίδουν κάποια ιδιότητα στο υποκείμενο. Π.χ. Ο Γιώργος είναι μουσικός.¹²⁷

Σημαντικό ρόλο παίζει και το ύφος του κειμένου όπως διαμορφώνεται από τη χρήση ρημάτων, εκφράσεων, επιθέτων και γλωσσικών δομών. Το ύφος εξαρτάται ανάλογα το μήνυμα που θέλει να περάσει ο συντάκτης του λόγου, το γενικότερο εξωγλωσσικό πλαίσιο και τους αποδέκτες του. Το ύφος μπορεί να είναι επίσημο, αυστηρό, προφορικοποιημένο. Το επίσημο ύφος μπορεί να γίνεται με χρήση γ' προσώπου, παθητικών μετοχών και του κατάλληλου λεξιλογίου. Το προφορικοποιημένο ύφος μπορεί να περιλαμβάνει χρήση α' ενικού προσώπου, ελλειπτικών προτάσεων, καθημερινών λέξεων και συναισθηματικής εμπλοκής.

Τέλος υπάρχει και μια σχετικά πρόσφατη προσέγγιση η οποία είναι η ψυχολογική προσέγγιση της Ανάλυσης του Λόγου. Σε αυτήν την περίπτωση η συζήτηση είναι ο τρόπος μέσω του οποίου οι άνθρωποι εκφράζουν συναισθήματα, επιθυμίες, επινοούν ιστορίες. Η Λογοψυχολογία μελετάει περισσότερο τα κίνητρα τα οποία ωθούν τους ανθρώπους σε συμπεριφορές, πράξεις και θεωρεί ότι υπάρχουν νοητικές δομές που καθορίζουν τις πράξεις μας.

2.19.2. Ανάλυση συναισθημάτων (sentimental analysis)

Η ύπαρξη συναισθημάτων σε ένα κείμενο μπορεί να φανερώνεται σε λέξεις και εκφράσεις οι οποίες υπάρχουν διάσπαρτες στον παραγόμενο λόγο¹²⁸. Κάποιες από αυτές τις λέξεις σχετίζονται με θετικά συναισθήματα και κάποιες με αρνητικά. Σχετικός είναι ο όρος της «συναισθηματικής νοημοσύνης/ ευφυΐας» (EQ) η οποία αφορά τον εντοπισμό των συναισθημάτων στο λόγο. Τα συναισθήματα αυτά μπορεί να μαρτυρούν και τη συναισθηματική κατάσταση στην οποία βρίσκεται ένας άνθρωπος ή το βαθμό ικανοποίησης του από μια κατάσταση.

¹²⁷ Στάμου, Α.Γ. (2014). Η κριτική ανάλυση του: Μελετώντας τον ιδεολογικό ρόλο της γλώσσας. Στο Ανάλυση λόγου: Θεωρία και εφαρμογές, Μ. Γεωργαλίδου, Μ. Σηφianού, Β. Τσάκωνα (επιμ.). Νήσος: Αθήνα, 149-187.

¹²⁸ Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

Σκοπός της ανάλυσης συναισθημάτων (sentimental analysis / opinion analysis) είναι η μελέτη και ανάλυση των απόψεων, των συναισθημάτων, των στάσεων και αντιλήψεων¹²⁹. Η αυτόματη εξαγωγή των συναισθημάτων από γραπτό λόγο μέσα από το διαδίκτυο και ιδίως από τα κοινωνικά μέσα δικτύωσης όπως facebook και twitter έχει λάβει μεγάλες διαστάσεις καθώς μπορεί να έχει διάφορες εφαρμογές όπως στο marketing και τις στρατηγικές αποφάσεις μιας εταιρείας, στη λήψη πολιτικών αποφάσεων, στη γνώμη σχετικά για έναν πολιτικό, ένα προϊόν ή μια υπηρεσία.

Για να πραγματοποιηθεί κάτι τέτοιο χρειάζονται οι κατάλληλοι αλγόριθμοι οι οποίοι κάποιες φορές χρησιμοποιούν κατάλληλα λεξικά συναισθημάτων και συμβάλλουν στην ταξινόμηση των δεδομένων σε κατηγορίες όπως: δυσαρεστημένος, θυμωμένος, ικανοποιημένος, αισιόδοξος. Τα λεξικά αυτά στην ουσία είναι λίστες λέξεων και εκφράσεων οι οποίες αφορούν θετική, αρνητική ή ουδέτερη άποψη¹³⁰ ή συγκεκριμένο συναίσθημα. Επίσης χρησιμοποιούνται οι κατάλληλες τεχνικές επεξεργασίας φυσικής γλώσσας όπως λημματοποίηση και διάφορες τεχνικές όπως η τεχνική bag of words στην οποία κάθε πληροφορία που σχετίζεται με τη σειρά ή τη δομή των λέξεων στο έγγραφο απορρίπτεται και υπολογίζονται στην επεξεργασία των δεδομένων μόνο συγκεκριμένες λέξεις και εκφράσεις.

Υπάρχουν διάφορες τέτοιου είδους εφαρμογές στο διαδίκτυο, σε διάφορες γλώσσες οι οποίες συνήθως απευθύνονται σε εταιρείες. Σε πολλές από αυτές υπάρχει η δυνατότητα επεξεργασίας μεγάλου όγκου δεδομένων και ο χρήστης μπορεί να επιλέξει το κοινωνικό δίκτυο από το οποίο επιθυμεί να γίνει η εξόρυξη των δεδομένων.

2.20. Γλωσσολογία και λογοτεχνία

Ένα λογοτεχνικό κείμενο διαφέρει από ένα άλλο κείμενο καθώς πρέπει να προκαλεί συναισθήματα όπως συγκίνηση, φόβο, χαρά, λύπη και να προσπαθεί να μεταφέρει στον αναγνώστη εικόνες. Γι' αυτό το λόγο στη γλώσσα γραφής των λογοτεχνικών

¹²⁹ Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

¹³⁰ Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*.

κειμένων συναντάμε γλωσσολογικές έννοιες όπως τα κοσμητικά επίθετα, την παρήχηση, την ομοιοκαταληξία, τις μεταφορές, τις μετωνυμίες κ.α. Επίσης το λεξιλόγιο είναι πλούσιο και πολλές φορές περιέχει λέξεις που δε συνηθίζονται στην καθομιλουμένη γλώσσα.

Επίσης χρησιμοποιούνται συγκεκριμένες δομές στην αφήγηση. Οι Γάλλοι δομιστές θεωρούν ότι οι δομές αυτές περιέχουν συστατικά όπως οι δράστες (acteurs), οι ρόλοι δράσεως (actantial roles) και κάποιες άλλες λειτουργίες βασικές στην αφήγηση (actantial model).¹³¹

Η συμβολή της γλωσσολογίας στη λογοτεχνία είναι να διερευνήσει πως ένα λογοτεχνικό έργο γίνεται αντιληπτό από τον άνθρωπο, ποια συναισθήματα του προκαλεί και σε ποιες δομές, εκφράσεις και γραμματικά φαινόμενα αντιστοιχεί. Επίσης η γλωσσολογία εξετάζει και τις εκπαιδευτικές δυνατότητες που έχει ένα λογοτεχνικό κείμενο, τους μηχανισμούς της γλώσσας και τους νόμους και κανόνες που διέπουν ένα λογοτεχνικό έργο.

Οι παράγοντες που είναι απαραίτητοι για κάθε επικοινωνία είναι ο πομπός, ο δέκτης, η πληροφορία που περιέχει η επικοινωνία, το μήνυμα, η γλώσσα ή η κωδικοποίηση του μηνύματος που είναι κοινή για τον πομπό και το δέκτη, το κανάλι επικοινωνίας. Με κάθε ένα από τους παράγοντες συνδέεται μια βασική λειτουργία. Τα συστατικά στοιχεία του κάθε μηνύματος ονομάζονται σημεία. Τα σημεία μπορεί να είναι γλωσσικά, μουσικά, αριθμοί, εικόνες κτλ.

Στη λογοτεχνία τα σημεία αποτελούν οι λέξεις και τα ονόματα που περιέχει το κάθε κείμενο. Κάθε σημείο έχει ένα σημαίνον το οποίο αντιστοιχεί σε μια έννοια ή μια ιδέα δηλαδή ένα σημαινόμενο. Κάθε σημείο μπορεί να είναι μοναδικό ή να υπάρχουν και άλλα ομώνυμα σημεία με την ίδια ακριβώς σημασία.

¹³¹ Μπαμπινιώτης, Γ. (1991). *Γλωσσολογία και Λογοτεχνία από την Τεχνική στην Τέχνη του Λόγου*. Αθήνα.

2.20.1. Λογοτεχνικό Ύφος

Το λογοτεχνικό ύφος είναι ο ιδιαίτερος τρόπος με τον οποίο χρησιμοποιούνται όλα τα μέσα που διαθέτει η γλώσσα προκειμένου ο συγγραφέας του κείμενου να περάσει τα μηνύματα και τις εικόνες που επιθυμεί στο κοινό και να προσδώσει στο κείμενο ένα ιδιαίτερο χαρακτήρα. Το λογοτεχνικό ύφος είναι από τα σημαντικότερα στοιχεία με τα οποία ασχολείται η γλωσσολογία με τη λογοτεχνία, σε σημείο η γλωσσολογική ανάλυση λογοτεχνικών κειμένων πολλές φορές να ταυτίζεται με την υφολογία.

Σύμφωνα με τον Γ. Μπαμπινιώτη υπάρχει σχέση μεταξύ γλώσσας, λόγου, ομιλίας και ύφους. Η ομιλία περιλαμβάνει το ατομικό ύφος το οποίο σχετίζεται με την κοινή γλώσσα και το λογοτεχνικό ύφος. Επίσης λόγος γίνεται για τον αυτοματισμό που αφορά τη χρήση της γλώσσας έτσι ώστε να προσελκύει την προσοχή και να δίνει την αίσθηση του ασυνήθιστου και μη αυτοματοποιημένου.¹³²

Κάποια από τα είδη του ύφους που μπορεί να έχει ένα λογοτεχνικό έργο είναι: λιτό, απλό, ειρωνικό, αυστηρό, επίσημο, χιουμοριστικό, λαϊκό, σεμνό, προσβλητικό ή αγενές. Ανάλογα το ύφος που θέλει να επιλέξει ο συγγραφέας επιλέγει το κατάλληλο λεξιλόγιο, τις κατάλληλες εκφράσεις, γραμματικά φαινόμενα, συντακτικές δομές. Για παράδειγμα στο ποίημα *Ιθάκη* του Κ. Καβάφη ο ποιητής σε ηλικία 47 ετών θέλει να δώσει συμβουλές χρησιμοποιεί διδακτικό και παραινετικό ύφος και γι' αυτό κάνει χρήση υποτακτικής και προστακτικής έγκλισης, ενώ χρησιμοποιεί το δεύτερο ενικό πρόσωπο. Επίσης κάνει χρήσεις συμβολισμών όπως «οι Κύκλωπες και οι Λαιστρυγόνες».

Ο λογοτέχνης επιθυμεί να περάσει ένα μήνυμα ή μια εικόνα. Αφού αποφασίσει τη σύνταξη ή τη συντακτική δομή που θέλει να ακολουθήσει, για κάθε λέξη της φράσης που θέλει να χρησιμοποιήσει βρίσκει εναλλακτικές. Από κάθε εναλλακτική επιλέγει αυτήν η οποία ταιριάζει καλύτερα στο ύφος που θέλει να προσδώσει.

Πολλές φορές σε ποιήματα και μυθιστορήματα ή διηγήματα υπάρχουν ασυνήθιστα δομικά και σημασιολογικά σχήματα που φαντάζουν ξένα από την πραγματικότητα.¹³³ Το φαινόμενο αυτό είναι έντονο στον υπερεαλισμό, αλλά και σε λαϊκά και δημοτικά ποιήματα. Για παράδειγμα μια χαρακτηριστική περίπτωση στους

¹³² Μπαμπινιώτης, Γ. (1991). *Γλωσσολογία και Λογοτεχνία από την Τεχνική στην Τέχνη του Λόγου*. Αθήνα. Σελ. 104 - 105.

¹³³ Μπαμπινιώτης, Γ. (1991). *Γλωσσολογία και Λογοτεχνία από την Τεχνική στην Τέχνη του Λόγου*. Αθήνα. Σελ. 111.

στίχους ενός ποντιακού ποιήματος είναι : «Μοιρολογούνε τα βουνά κλαίνε οι άκρες των ποταμών».

2.20.2. Σημασιολογική δομή στην ποίηση

Η σημασιολογική δομή περιλαμβάνει όλες τις γλωσσικές δομές μέσω των οποίων ο λογοτέχνης προσπαθεί να μεταδώσει το μήνυμά του. Περιλαμβάνει μορφήματα, λέξεις και φράσεις.¹³⁴ Η σημασιολογική δομή ενός ποιητικού κειμένου αποτελείται από δύο βασικά λεξικά στοιχεία:

- Τις επιλογές οι οποίες είναι οι συνειδητές και επαναλαμβανόμενες λέξεις και φράσεις που χρησιμοποιούνται συνήθως.
- Τις αποκλίσεις οι οποίες είναι πρωτότυπες λέξεις ή εκφράσεις που επινοεί ο λογοτέχνης και σε κάποιες περιπτώσεις προσδίδει μια μοναδικότητα στο έργο του που το κάνει να ξεχωρίζει από όλα τα υπόλοιπα. Τέτοιες λέξεις και εκφράσεις εμπλουτίζουν το λεξιλόγιο της γλώσσας, μεταφέρονται στο λαό και μπορεί ακόμη και να υιοθετηθούν και να χρησιμοποιούνται και στην καθημερινότητα.

2.20.3. Απόδοση πατρότητας σε κείμενα (Author Recognition)

Κάθε λογοτέχνης διαθέτει το δικό του προσωπικό ύφος. Ενδεχομένως να υπάρχουν επιρροές από άλλα λογοτεχνικά έργα, αλλά σε κάθε περίπτωση ο δημιουργός επιχειρεί να δώσει το δικό του προσωπικό στίγμα. Το προσωπικό ύφος ενός λογοτέχνη μπορεί να περιλαμβάνει μεταξύ άλλων τη χρήση των ίδιων δομών και κάποιες πρωτότυπες λέξεις ή εκφράσεις. Επίσης κάθε δημιουργός έχει τα δικά του βιώματα, τη δική του προσωπική ζωή και εμπειρίες καθώς επίσης συγκεκριμένες επιρροές. Όλα αυτά πολλές φορές αποτυπώνονται στο λογοτεχνικό του έργο κάνοντας το να ξεχωρίζει από τα υπόλοιπα. Από όλα αυτά προκύπτει το συμπέρασμα

¹³⁴ Μπαμπινιώτης, Γ. (1991). *Γλωσσολογία και Λογοτεχνία από την Τεχνική στην Τέχνη του Λόγου*. Αθήνα. Σελ. 173.

ότι ο τρόπος γραφής μπορεί να συνδέεται με την ταυτότητα του δημιουργού ή με κάποια άλλα βασικά στοιχεία όπως το φύλο, η ηλικία , τις επιρροές του και τα ψυχολογικά χαρακτηριστικά. Επιπλέον η γλώσσα γραφής μπορεί να προσδιορίζει την περίοδο στην οποία γράφτηκε το έργο.

Υπάρχουν περιπτώσεις κειμένων, βιβλίων, ποιημάτων για τα οποία δε σώζεται το όνομα του δημιουργού. Πολλά τέτοια κείμενα ανακαλύπτονται σε βιβλιοθήκες, αρχεία. Η γλωσσολογία μπορεί να συμβάλει στην προσπάθεια απόδοσης συγγραφέα σ' ένα λογοτεχνικό κείμενο με τη βοήθεια της ανάκτησης της πληροφορίας, της επεξεργασίας φυσικής γλώσσας και της μηχανικής μάθησης.

Μέσω της μηχανικής μάθησης υπάρχει η δυνατότητα δημιουργίας αλγορίθμων οι οποίοι μπορούν να εκπαιδεύονται με δεδομένα και μπορούν να κάνουν προβλέψεις. Μια τέτοια πρόβλεψη μπορεί να είναι η ταυτότητα του δημιουργού ή κάποια προσωπικά στοιχεία του. Μεγάλη είναι και η συμβολή του διαδικτύου από το οποίο μπορεί να ανακτηθεί μεγάλος όγκος πληροφορίας.

Κάποιοι από του τύπους εφαρμογών απόδοσης πατρότητας είναι οι εξής:

- Μέσω μιας λίστας συγγραφέων με δείγματα γραφής του έργου τους και παράλληλα διαθέτουμε και ένα κείμενο του οποίου δε γνωρίζουμε το δημιουργό, αλλά γνωρίζουμε ότι είναι κάποιος από τη λίστα. Σε αυτή την περίπτωση προσπαθούμε να εντοπίσουμε ποιος συγγραφέας από τη λίστα είναι ο δημιουργός του έργου.
- Μέσω λίστας συγγραφέων με δείγματα γραφής του έργου τους αλλά χωρίς να γνωρίζουμε αν κάποιος από αυτούς είναι ο συγγραφέας του έργου. Σε αυτή την περίπτωση ενδέχεται ο συγγραφέας του έργου να είναι κάποιος άλλος ο οποίος δεν περιέχεται στη λίστα.

Η πρώτη μέθοδος έχει σημειώσει μεγάλη πρόοδο και είναι ασφαλέστερη, καθώς υπάρχει περιορισμένος αριθμός διαθέσιμων επιλογών από τη λίστα. Μπορεί να αναπτυχθεί λοιπόν αλγόριθμος που μετατρέπει τα υφομετρικά στοιχεία του έργου σε δείκτες και να τους συγκρίνει με τους αντίστοιχους των έργων των συγγραφέων που βρίσκονται στη λίστα. Στην συνέχεια σε κάθε συγγραφέα αποδίδεται ένα σκορ και οι συγγραφείς με το υψηλότερο έχουν τη μεγαλύτερη πιθανότητα πατρότητας του έργου.

Απλά προγράμματα υπάρχουν στο ελεύθερο στο διαδίκτυο. Ένα απλό παράδειγμα περιέχει η εφαρμογή στην ιστοσελίδα της aicbt¹³⁵ η οποία είναι γραμμένη σε γλώσσα python και δίνει τη δυνατότητα να εισάγουμε χαρακτηριστικά κομμάτια έργου συγγραφέων και στη συνέχεια να εισάγουμε και ένα άλλο τμήμα έργου ενός άγνωστου συγγραφέα. Το λογισμικό μπορεί να μας εντοπίσει αν το έργο του άγνωστου συγγραφέα ανήκει σε κάποιο από αυτά τα οποία εισάγαμε για να εκπαιδεύσουμε το σύστημα.

2.21. Η Υπολογιστική Γλωσσολογία

Η υπολογιστική γλωσσολογία αποτελεί την τομή μεταξύ της πληροφορικής και της γλωσσολογίας. Σκοπός της υπολογιστικής γλωσσολογίας είναι η ανάπτυξη κώδικα ο οποίος επιτρέπει τη μοντελοποίηση της φυσικής γλώσσας από τον υπολογιστή προκειμένου να μπορεί να πραγματοποιήσει υπολογισμούς. Η υπολογιστική γλωσσολογία αποτελεί πολύ σημαντικό κλάδο της γλωσσολογίας γι' αυτό υπάρχει η Παγκόσμια Ένωση Υπολογιστικών Γλωσσολόγων (ACL) καθώς και η Ευρωπαϊκή Ένωση Υπολογιστικών Γλωσσολόγων (ACL)¹³⁶.

Βασική εφαρμογή της υπολογιστικής γλωσσολογίας είναι η μηχανική μετάφραση. Η Υπολογιστική Γλωσσολογία περιλαμβάνει και εφαρμογές που αφορούν τον προφορικό λόγο όπως αναγνώριση φωνής, παραγωγή ομιλίας αλλά και εφαρμογές που αφορούν το γραπτό λόγο όπως: ο διαχωρισμός των λέξεων, ο γραμματικός έλεγχος, η εξόρυξη δεδομένων από ένα κείμενο, η δημιουργία υπολογιστικών συστημάτων με διαλογική ικανότητα¹³⁷, η αυτόματη διόρθωση κειμένου την οποία διαθέτουν κάποιες εφαρμογές όπως το MS Office Word¹³⁸.

¹³⁵ <http://www.aicbt.com/authorship-attribution/online-software/>

¹³⁶ Tantos, A., Markantonatou, S., Anastassiadis Symeonidis, A., Kyriakopoulou, P., Τάντος, Α., Μαρκαντωνάτου, Σ., ... & Κυριακοπούλου, Π. (2015). Υπολογιστική γλωσσολογία.

¹³⁷ <https://plato.stanford.edu/entries/computational-linguistics/>

¹³⁸ Tantos, A., Markantonatou, S., Anastassiadis Symeonidis, A., Kyriakopoulou, P., Τάντος, Α., Μαρκαντωνάτου, Σ., ... & Κυριακοπούλου, Π. (2015). Υπολογιστική γλωσσολογία.

2.21.1. Τα επίπεδα της υπολογιστικής γλωσσολογίας

Για την ανάπτυξη υπολογιστικών γλωσσικών εφαρμογών η υπολογιστική γλωσσολογία προϋποθέτει τη σύμπραξη και άλλων γλωσσολογικών υποκλάδων και ιδίως της φωνολογίας, της μορφολογίας, της σύνταξης και της σημασιολογίας. Επίσης απαραίτητη είναι η χρήση λεξικών και κάποιων μαθηματικών εργαλείων. Έτσι λοιπόν διαμορφώνονται οι παρακάτω κατηγορίες της υπολογιστικής γλωσσολογίας:

- **Η Υπολογιστική φωνολογία** η οποία περιλαμβάνει δομές δεδομένων και αλγορίθμους. Παλαιότερα γίνονταν χειροκίνητα με χαρτί και μολύβι. Τα φωνολογικά δεδομένα στην υπολογιστική γλωσσολογία αποτελούνται από κείμενα, λίστες λέξεων και παραδείγματα που αποτελούνται από τυποποιήσεις λέξεων και φράσεων τα οποία προέρχονται από ηχητικά δεδομένα. Μέσα από τη διαδικασία της ψηφιοποίησης είναι δυνατή η υπολογιστική επεξεργασία τους.¹³⁹
- **Η Υπολογιστική Μορφολογία** της οποίας ο σκοπός είναι η είσοδος μιας σειράς φωνημάτων και η παραγωγή εξόδου η οποία θα περιέχει τη λέξη ή τα φωνήματα της εισόδου αναλυμένα. Αν για παράδειγμα η είσοδος είναι η λέξη «houses» η πιθανή έξοδος μπορεί να ήταν της μορφής : house+N+PL ή house+s. Χρήσιμα εργαλεία για την υπολογιστική μορφολογία είναι λεξικά καθώς μπορεί να απαιτούνται για την μορφολογική ανάλυση συγκεκριμένων περιπτώσεων λέξεων όπως για παράδειγμα τα ρήματα Αγγλική Γλώσσα που αλλάζουν εντελώς μορφή κατά το σχηματισμό του αόριστου και του παρακείμενου π.χ. take-took-taken.
- **Η Αυτόματη συντακτική ανάλυση** (parsing) η οποία έχει περιγραφεί σε προηγούμενο κεφάλαιο. Η αυτόματη συντακτική ανάλυση δέχεται ως είσοδο κείμενο φυσικής γλώσσας και επιστρέφει τη συντακτική δομή του.
- **Η Υπολογιστική σημασιολογία** κατά την οποία πραγματοποιείται προσπάθεια αναπαράστασης του νοήματος μιας φυσικής γλώσσας σε μια σημασιολογική αναπαράσταση. Σκοπός δεν είναι μόνο η αναπαράσταση της κάθε λέξης αλλά του νοήματος της κάθε φράσης και της σχέσης μεταξύ των μερών μιας πρότασης. Η διαφοροποίηση με την πραγματολογική ανάλυση

¹³⁹ Bird, S. (1994). Introduction to computational phonology. *Computational Linguistics*, 20(3), 1-8.

είναι ότι δε λαμβάνονται καθόλου υπόψη τα συμφραζόμενα για την εξαγωγή του νοήματος μιας φράσης.

Υπάρχουν διάφοροι τρόποι από τους οποίους μπορεί να πραγματοποιηθεί αυτό. Ο πιο συνήθης τρόπος είναι η προτασιακή λογική και η κατηγορηματική λογική. Με αυτόν τον τρόπο τα μέρη από τα οποία αποτελείται μια πρόταση περιγράφονται μέσω αντικειμένων και ιδιοτήτων. Μέσω της κατηγορηματικής λογικής διαφαίνονται οι σχέσεις που υπάρχουν μεταξύ των λέξεων και των μορφημάτων μιας πρότασης οι οποίες μπορεί να είναι αληθείς ή ψευδείς .

Με τη χρήση της κατηγορηματικής λογικής οι παρακάτω προτάσεις θα γραφθούν ως εξής:

- Όσοι πεινάν τρώνε.: $\forall x \text{πεινάν} \Rightarrow \text{τρώνε}$
- Κάποιος που δεν αγαπά κανέναν. $\exists x \forall y \neg \text{αγαπά}(x,y)$
- Η ζωή είναι ωραία : $\text{ωραία}(\text{ζωή})$

2.21.2. Βασικές μαθηματικές έννοιες

Υπάρχουν κάποιες μαθηματικές έννοιες που είναι πολύ χρήσιμες για την ανάπτυξη των τυπικών γλωσσών. Πολύ μεγάλης σημασίας είναι η θεωρία των πιθανοτήτων. Σε περιπτώσεις ασαφειών ή αμφισημιών οι πιθανότητες μπορεί να λύσουν το πρόβλημα καθώς δεδομένου ενός συνόλου επιλογών μπορεί να επιλεγεί η πιο πιθανή και αυτή να είναι και η σωστή.

Επίσης πολύ χρήσιμη είναι και η θεωρία των συνόλων για τον ορισμό μοντέλων αναπαράστασης φυσικής γλώσσας. Ένα σύνολο είναι κάθε συλλογή σαφώς διακριτών και καλώς καθορισμένων αντικειμένων που προέρχονται από τον χώρο της εμπειρίας (αντικείμενα συγκεκριμένα) ή των διανοημάτων (αντικείμενα αφηρημένα), τα οποία θεωρούνται ως μια ολότητα ¹⁴⁰.

Επιπλέον χρήσιμη είναι και η θεωρία της λογικής και ειδικά ο προτασιακός λογισμός και ο κατηγορικός λογισμός. Ο προτασιακός λογισμός αποτελείται από προτάσεις. Δύο ή περισσότερες προτάσεις μπορούν να συσχετιστούν μέσω των πράξεων της σύζευξης ($\wedge$) που σε μια πρόταση μεταφράζεται με το «και», της διάζευξης ($\vee$) που

¹⁴⁰ <https://el.wikipedia.org/wiki/Σύνολο>

σε μια πρόταση μεταφράζεται με το «ή», και της συνεπαγωγής ($\rightarrow$) και της ισοδυναμίας ($\leftrightarrow$).

2.21.3. Σύνδεση με την ανάκτηση της πληροφορίας

Η υπολογιστική γλωσσολογία είναι ένα μεγάλο πεδίο της γλωσσολογίας το οποίο περιλαμβάνει ένα μεγάλο εύρος εφαρμογών που σχετίζεται με ανάκτηση φωνής και ανάκτηση κειμένου. Είδη έχει γίνει αναφορά σε πολλές τέτοιες εφαρμογές σε προηγούμενα κεφάλαια όπως η αναγνώριση φωνής, η σύνθεση φωνής, η σημασιολογική αναζήτηση, η αυτόματη μηχανική μετάφραση κειμένου, η αυτόματη παραγωγή περιλήψεων, η υπολογιστική λεξικολογία και η ανάκτηση πληροφοριών από κείμενο.

Ενδεικτικά χρήσιμο είναι να αναφερθούν κάποιες επιπλέον πολύ σημαντικές εφαρμογές της υπολογιστικής γλωσσολογίας. Μία τέτοια εφαρμογή είναι ο αυτόματος ορθογραφικός έλεγχος τον οποίο χρησιμοποιούν πολλές εφαρμογές ηλεκτρονικών υπολογιστών ή λογισμικών Ios και Android προκειμένου να διορθώνονται αυτόματα ορθογραφικά και συντακτικά λάθη. Επιπλέον υπάρχουν και εφαρμογές οι οποίες αποκλειστικά πραγματοποιούν γραμματικό και ορθογραφικό έλεγχο ενώ και μηχανές αναζήτησης όπως το google chrome υποστηρίζουν ορθογραφικό έλεγχο μέσω των κατάλληλων επεκτάσεων.

Μια άλλη χρήσιμη εφαρμογή της υπολογιστικής γλωσσολογίας είναι η χρήση φίλτρων αναζήτησης και αντικατάστασης τα οποία κάνουν την αναζήτηση πιο εξειδικευμένη αποκλείοντας κάποια μη επιθυμητά αποτελέσματα. Τα φίλτρα αυτά χρησιμοποιούν κανονικές εκφράσεις¹⁴¹. Εκτός από τις web μηχανές αναζήτησης τέτοια φίλτρα υπάρχουν και σε μηχανές αναζήτησης ιστοτόπων τα οποία είναι εξαιρετικά ισχυρά και αποτελεσματικά.

¹⁴¹ <https://support.google.com/analytics/answer/1034834?hl=el>

2.22. Γνωσιακή επιστήμη (cognitive science)

Η γνωσιακή επιστήμη εξετάζει τη λειτουργική συμπεριφορά του ανθρώπινου νου και τα νοητικά μοντέλα που διαθέτει καθώς επίσης και τους μηχανισμούς που σχετίζονται με την υποστήριξη των γνωσιακών λειτουργιών . Βασικό ζήτημα της Γνωσιακής Γλωσσολογίας είναι ο τρόπος με τον οποίο αποθηκεύεται η πληροφορία στο νου¹⁴².

Στις αρχές του 19^{ου} αιώνα οι φυσιολόγοι Luigi Galvani , Joannes Muller, Herman van Helmholtz, Gustav Flechner υποστήριζαν ότι μεταβίβαση πληροφοριών μέσω των νευρώνων του ανθρώπινου εγκεφάλου, έχει τη μορφή ηλεκτρικού σήματος¹⁴³. Σύμφωνα με τη γνωσιακή επιστήμη από το εξωτερικό περιβάλλον μεταφέρονται πληροφορίες στο ανθρώπινο γνωστικό σύστημα από το οποίο στη συνέχεια επεξεργάζονται. Ως έξοδο του συστήματος αυτού θα μπορούσαμε να ορίσουμε την ανθρώπινη συμπεριφορά που διαμορφώνεται από την επεξεργασία των ερεθισμάτων του περιβάλλοντος.

Ο Allan Pavo το 1941 διατύπωσε τη θεωρία της διπλής κωδικοποίησης σύμφωνα με την οποία στον ανθρώπινο εγκέφαλο υπάρχουν δύο εσωτερικά συστήματα για την αποθήκευση πληροφοριών που εισέρχονται μέσα από ερεθίσματα του περιβάλλοντος. Τα συστήματα αυτά είναι:

- η προφορική πληροφορία που προέρχεται από την αφήγηση και τον προφορικό λόγο.
- Η πληροφορία που προέρχεται μέσω της όρασης μέσα από εικόνες και το γραπτό λόγο.

Η πληροφορία αυτή αποθηκεύεται και όταν ανακαλείται από τη μνήμη ενεργοποιούνται και τα δύο συστήματα. Βέβαια οι πληροφορίες που δέχεται ένας άνθρωπος είναι πάρα πολλές και είναι αδύνατο να διατηρούνται στην ανθρώπινη μνήμη όλες. Ο ψυχολόγος Goerge Miller υποστήριξε ότι μπορούμε να ανακαλούμε στη μνήμη μας περιορισμένο αριθμό και μάλιστα το ίδιο υποστήριξε και ο Σίγμουντ Φρόυντ, σύμφωνα με τον οποίο κάποιες πληροφορίες δεν είναι δυνατό να διατηρούνται στη μνήμη αλλά παρόλα αυτά επηρεάζουν υποσυνείδητα την ανθρώπινη συμπεριφορά. Με άλλα λόγια ο άνθρωπος νους αποτελεί ένα πολύπλοκο

¹⁴² Πανάρετου,Ε.(2011). *Θέματα Γνωσιακής Γλωσσολογίας*.Αθήνα :Εκδόσεις Παπαζήση.

¹⁴³ Βοσνιάδου, Σ. (2004). *Γνωσιακή Επιστήμη: Η Νέα Επιστήμη του Νου*. Εκδόσεις Gutenberg, Αθήνα.

υπολογιστικό γνωσιακό σύστημα, ενώ η ανθρώπινη σκέψη αντιστοιχεί στον τρόπο επεξεργασίας πληροφοριών και δεδομένων.

Στη γνωσιακή γλωσσολογία το ενδιαφέρον επικεντρώνεται στη γνώση που αποκτάται μέσω της γλώσσας, σε αντίθεση με τη γενετική γραμματική που αφορά τη γνώση της γλώσσας.¹⁴⁴ Βασικό ερώτημα είναι το κατά πόσο η γλώσσα την οποία γνωρίζουμε επηρεάζει τον τρόπο που σκεφτόμαστε. Σίγουρα η γλώσσα δεν επηρεάζει τις οπτικές εικόνες που βλέπουμε, όμως παρόλα αυτά για κάποιες εικόνες δεν υπάρχουν λόγια για να τις περιγράψουμε προφορικά .¹⁴⁵

2.22.1. Δυισμός

Σύμφωνα με τη θεωρία του δυισμού το σύμπαν διέπεται από φυσικές και νοητικές ουσίες. Στις φυσικές ουσίες αντιστοιχεί η ύλη και στις νοητικές ο νους και η σκέψη. Οι δύο αυτές ουσίες παρόλο που είναι ανεξάρτητες μεταξύ τους κάποιες φορές αναπτύσσουν σχέσεις. Για παράδειγμα ένα υλικό αντικείμενο μπορεί να διεγείρει τα αισθητηριακά όργανα και να προκαλεί ένα συναίσθημα, μια νοητική κατάσταση η οποία ενδεχομένως θα προκαλέσει μια κίνηση ή μια πράξη. Ο νους παρόλο που είναι διαφορετικός από την ύλη και το σώμα, σύμφωνα με τη θεωρία του δυισμού συνδέεται προσωρινά με το σώμα μέσω μιας υπόφυσης στον ανθρώπινο εγκέφαλο.

Σύμφωνα με τον Noam Chomsky στον ανθρώπινο νου προϋπάρχουν κάποιοι γλωσσικοί κανόνες οι οποίοι χρησιμοποιούνται ως μέσω της γραπτής ή προφορικής έκφρασης. Ανάλογα με το τι θέλει να πει κάποιος, χρησιμοποιεί την κατάλληλη δομή. Έτσι με αυτόν τον τρόπο ο λόγος αποτελεί εικόνα του ανθρώπινου νου και της σκέψης.

Χρησιμοποιώντας την κατάλληλη δομή μπορούμε να χρησιμοποιήσουμε μια λέξη με διαφορετικό τρόπο. Μια λέξη δεν έχει απαραίτητα μια ερμηνεία, αλλά η χρήση της εξαρτάται από εικόνες οι οποίες βρίσκονται στον ανθρώπινο νου. Σχετικός είναι και ο όρος πολυσημία, ο οποίος αναφέρεται στις πολλές σημασίες που μπορεί να έχει ένας όρος .

¹⁴⁴ Geeraerts, D., & Cuyckens, H. (2007). Introducing cognitive linguistics. In *The Oxford handbook of cognitive linguistics*.

¹⁴⁵ <https://www.linguisticsociety.org/content/does-language-i-speak-influence-way-i-think>

Μπορεί επίσης να υπάρχει μια σημασία κυριολεκτική , ενώ ο ίδιος όρος μπορεί να χρησιμοποιηθεί με διαφορετικό τρόπο μεταφορικά. Στην περίπτωση αυτή έχουμε μετωνυμία όπου γίνεται χρήση μιας λέξης στη θέση μιας άλλης που θα έπρεπε να χρησιμοποιηθεί. Κάποια παραδείγματα μετωνυμίας είναι τα εξής: «Είμαι πτώμα στην κούραση», «ήπιε δύο ποτηράκια», «έφαγε δύο πιάτα», κ.τ.λ.

2.22.2. Σημασιολογικά πλαίσια

Μια ιδιότητα του ανθρώπινου νου είναι η κατηγοριοποίηση εννοιών. Οι έννοιες που εντάσσονται σε κάποιες κατηγορίες διαθέτουν κάποια συγκεκριμένα κοινά χαρακτηριστικά. Για να ανήκει μια έννοια σ' ένα σύνολο αρκεί να διαθέτει κάποια από τα χαρακτηριστικά γνωρίσματα της κατηγορίας. Υπάρχουν περιπτώσεις όπου κάποιες έννοιες διαθέτουν όλα τα χαρακτηριστικά της κατηγορίας και ονομάζονται πρωτότυπες.

Στην μνήμη του ανθρώπου υπάρχουν γνωσιακά σχήματα στα οποία οργανώνεται η γνώση. Τα σχήματα αυτά είναι δομές οι οποίες ενεργοποιούνται από την μνήμη και στις οποίες επιτρέπεται η συμπλήρωση επιπλέον στοιχείων. Οι διάφορες κατηγοριοποιημένες έννοιες οργανώνονται βάσει των γνωσιακών σχημάτων.

Τα γνωσιακά σχήματα τα διακρίνουμε σε σενάρια και πλαίσια. Τα σενάρια είναι συγκεκριμένες κατηγορίες γεγονότων που προσδιορίζουν μια κατάσταση . Κάθε σενάριο περιλαμβάνει σκηνικό, ρόλους, συνθήκες εισαγωγής, αποτελέσματα, αλληλοδιαδεχόμενες σκηνές. Τα γνωσιακά πλαίσια αντίθετα έχουν πιο στατική μορφή και οι έννοιες τους αλληλοσυνδέονται μεταξύ τους. Κάποιες από τις έννοιες που ανήκουν σ' ένα πλαίσιο μπορεί να αποτελούν επιμέρους κατηγορία.

Σύμφωνα με τον Fillmore τα νοητικά πλαίσια περιλαμβάνουν έννοιες και σχέσεις οι οποίες έχουν δομηθεί μέσω της εμπειρίας, ενώ ο Langacker χρησιμοποίησε τον όρο «πεδίο» το οποίο περιλαμβάνει νοητικές αναπαραστάσεις και εμπειρίες και άλλα είδη γνώσεων που χρειάζονται προκειμένου να μπορέσουμε να κατανοήσουμε μια έννοια ¹⁴⁶. Στην έννοια εντάσσονται η βάση και η εξέχουσα οντότητα. Η βάση αποτελεί ένα πεδίο , ενώ η εξέχουσα οντότητα είναι μια έννοια που εντάσσεται στη

¹⁴⁶ Πανάρετου,Ε.(2011). *Θέματα Γνωσιακής Γλωσσολογίας*.Αθήνα :Εκδόσεις Παπαζήση.

βάση. Για παράδειγμα μεταξύ της σχέσεις των όρων αυτοκίνητο και ρόδα, το αυτοκίνητο αποτελεί τη βάση και η ρόδα την εξέχουσα οντότητα.

2.22.3. Κατηγοριοποίηση κειμένου (Text Categorization)

Από τις χαρακτηριστικές ιδιότητες του ανθρώπινου νου είναι η κατηγοριοποίηση των εννοιών η οποία βοηθάει στη σύνδεση των γλωσσικών σημασιών. Κατηγοριοποίηση γενικώς είναι ο προσδιορισμός της κατηγορίας στην οποία υπάγεται μια έννοια.

Η κατηγοριοποίηση του κειμένου είναι μια διαδικασία κατά την οποία κείμενα φυσικής γλώσσας αντιστοιχίζονται σε κάποιες κατηγορίες από ένα προκαθορισμένο σύνολο. Η κατηγοριοποίηση ενός κειμένου μπορεί να γίνει με το χέρι, μπορεί να γίνει όμως και αυτόματα με την εφαρμογή τεχνικών μηχανικής μάθησης. Μέσω της μηχανικής μάθησης η διαδικασία είναι η εξής:

- Αρχικά γίνεται η εξαγωγή χαρακτηριστικών με τη βοήθεια διανυσματικών αναπαράστασεων.
- Στη συνέχεια το σύστημα εκπαιδεύεται με τα κατάλληλα δεδομένα
- Και μετά την εκπαίδευση το σύστημα είναι ικανό να πραγματοποιήσει τις κατάλληλες προβλέψεις.¹⁴⁷

Υπάρχουν δύο είδη αυτόματης κατηγοριοποίησης : η δυαδική όταν υπάρχουν δύο διαθέσιμες κατηγορίες και η multi-class όταν υπάρχει μεγάλος αριθμός κατηγοριών. Παράδειγμα δυαδικής κατηγοριοποίησης είναι η εύρεση μηνυμάτων spam.

Για την ταξινόμηση των κειμένων απαραίτητη είναι η χρήση των κατάλληλων αλγορίθμων όπως naïve bayes, svm¹⁴⁸, k-nearest neighbor κ.α. Η αυτόματη κατηγοριοποίηση κειμένου είναι μια πολύ σημαντική τεχνική που μπορεί να εφαρμοστεί σε ιστοσελίδες , σε διάφορες εφαρμογές, στην αυτόματη ευρετηρίαση συστημάτων ανάκτησης πληροφορίας, στο φιλτράρισμα εγγράφων στην εύρεση spam μηνυμάτων ηλεκτρονικού ταχυδρομείου και στις ιστοσελίδες που διαχειρίζονται

¹⁴⁷ <https://monkeylearn.com/text-classification/>

¹⁴⁸ Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137-142). Springer, Berlin, Heidelberg.

ερωτήσεις¹⁴⁹ όπως: yahoo answers , bing intelligent search, ενώ παλαιότερα υπήρχε και αντίστοιχη εφαρμογή της google, κ.α. .

2.23. Η Εφαρμοσμένη Γλωσσολογία

Η εφαρμοσμένη γλωσσολογία αναφέρεται σε πρακτικά ζητήματα που σχετίζονται με τη γλώσσα που χρησιμοποιεί μια χώρα ή ένα έθνος . Είναι διεπιστημονικός κλάδος και σχετίζεται με άλλες επιστήμες όπως η ανθρωπολογία, η ψυχολογία, η τεχνολογία, η πληροφορική, κ.α..

Ένας από τους βασικότερους τομείς της εφαρμοσμένης γλωσσολογίας είναι η εκμάθηση μιας δεύτερης γλώσσας και η εκμάθηση της τοπικής γλώσσας μιας χώρας σε μετανάστες προκειμένου να μπορέσουν να ενταχθούν με το υπόλοιπο κοινωνικό σύνολο. Επίσης μπορεί να αφορά τη γλωσσική πολιτική ή τη γλωσσική εκπαίδευση μιας χώρας. Τα τελευταία χρόνια μάλιστα υπάρχει η τάση να αναφέρεται ακόμη και στην εκμάθηση συγκεκριμένης ειδικής γλώσσας και ορολογίας που χρησιμοποιούνται σε διάφορους εξειδικευμένους τομείς όπως η ιατρική, η αεροπορία, το δίκαιο, η φυσική κ.α.¹⁵⁰ .

Για τους παραπάνω λόγους κάποιες χώρες διαθέτουν αρμόδιους οργανισμούς εφαρμοσμένης γλωσσολογίας όπου κάθε ένας από αυτούς μπορεί να διαθέτει διαφορετικό προσανατολισμό ή σκοπό. Σε κάποιες χώρες ασχολούνται με τη διαμόρφωση των κατάλληλων πολιτικών σε δίγλωσσες κοινωνίες π.χ. στον Καναδά που υπάρχουν δύο επίσημες γλώσσες τα Αγγλικά και τα Γαλλικά. Έτσι λοιπόν υπάρχει Αμερικανική Ένωση Εφαρμοσμένων Γλωσσών (AAAL), η Βρετανική Ένωση Εφαρμοσμένης Γλωσσολογίας (BAAL) καθώς επίσης και η Διεθνής Ένωση Εφαρμοσμένης Γλωσσολογίας.

¹⁴⁹ Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.

¹⁵⁰ <https://www.linguisticsociety.org/resource/applied-linguistics>

2.23.1. Εκμάθηση ξένης γλώσσας

Η εκμάθηση 2^{ης} γλώσσας είναι από τα βασικότερα ζητήματα που αφορούν την εφαρμοσμένη γλωσσολογία σήμερα για τον επιπλέον λόγο ότι σε πολλές χώρες και κοινωνίες υπάρχει μεγάλος αριθμός μεταναστών που δε γνωρίζει την τοπική γλώσσα.

Κατά πολλούς είναι δυσκολότερη η απόκτηση 2^{ης} γλώσσας σε ενήλικους απ' ότι σε μικρά παιδιά. Οι λόγοι είναι δύο : γιατί κατά τη διάρκεια της παιδικής ηλικίας αναπτύσσεται ο εγκέφαλος, αλλά και γιατί στα παιδιά μπορεί να δοθεί περισσότερος χρόνος για εξάσκηση. Βέβαια υπάρχουν και οι αντίθετες απόψεις σύμφωνα με τις οποίες η ευκολία στην εκμάθηση μιας ξένης γλώσσας σχετίζεται με την εξάσκηση, τον χρόνο έκθεσης στη γλώσσα και στο γεγονός ότι στα μεγαλύτερα άτομα υπάρχει περισσότερη ωριμότητα.

Γενικά για τη διδασκαλία μιας γλώσσας στα πλαίσια της εφαρμοσμένης γλωσσολογίας συγκρίνονται μέθοδοι διδασκαλίας της γλώσσας με βάση τη γλωσσολογική ανάλυση των επιπέδων δομής και τις θεωρίες κατάκτησης. Βασικό είναι να μπορέσει κάποιος να κατακτήσει βασικές γλωσσικές ιδιότητες όπως: η ακρόαση, η παραγωγή προφορικού λόγου, η κατανόηση και ο γραπτός λόγος.

Η ακρόαση αφορά τη δυνατότητα παρακολούθησης της ομιλίας. Βασική προϋπόθεση είναι η κατάκτηση του λεξιλογίου. Για την εκμάθηση μιας γλώσσας καλό είναι ο συνομιλητής να έχει καθαρό ήχο φωνής και να προσέχει την ταχύτητα που χρησιμοποιεί στην ομιλία του. Στην τελειότερη μορφή της πραγματοποιείται πλήρης κατανόηση μιας πρότασης. Διαφορετικά η κατανόηση μπορεί να γίνει από τα συμφραζόμενα μιας ολοκληρωμένης πρότασης ή από το συνολικό περιεχόμενο μιας ομιλίας.

Για την παραγωγή προφορικού λόγου ο ομιλητής έχει τη δυνατότητα επιλογής του λεξιλογίου και των μορφοσυντακτικών δομών. Ο προφορικός λόγος είναι πάντα πλουσιότερος όταν υπάρχει καλή γνώση του λεξιλογίου. Είναι αμεσότερος καθώς ο ομιλητής δεν έχει τη δυνατότητα του χρόνου να σκεφτεί και να προετοιμάσει το λόγο του. Βασική τεχνική που εφαρμόζεται κατά τη διδασκαλία είναι η μελέτη επιμέρους θεμάτων με ανάλυση και διδασκαλία του απαραίτητου λεξιλογίου και των κατάλληλων εκφράσεων, η διόρθωση των λαθών από τον εκπαιδευτικό και μερικές φορές ακόμη και η αυτοδιόρθωση κ.α.

2.23.2. Γλωσσικός Σχεδιασμός

Είναι η γλωσσική πολιτική που ακολουθείται από ένα κράτος ή ένα κοινωνικό σύνολο και μέσα από την οποία ορίζεται η επίσημη γλώσσα επικοινωνίας, τα γλωσσικά πρωτόκολλα επικοινωνίας ή η εθνική γλώσσα. Η εφαρμοσμένη γλωσσολογία έχει σκοπό να συμβάλει στον τρόπο διδασκαλίας εθνικών ή και μειονοτικών γλωσσών. Ο βασικός λόγος ύπαρξης της γλωσσικής πολιτικής είναι γιατί σε πολλά κράτη υπάρχουν διάφορες τοπικές διάλεκτοι ή γλώσσες οι οποίες πρέπει να λαμβάνονται υπόψη αλλά σε συνδυασμό πάντα με το γεγονός ότι θα πρέπει κάποιες από αυτές τις γλώσσες να είναι οι επίσημες του κράτους και να διδάσκονται στα σχολεία.

2.23.3. Σύνδεση με την ανάκτηση της πληροφορίας

Μέσω του web και των νέων τεχνολογιών δίνονται νέες εκπαιδευτικές δυνατότητες και γι' αυτό το λόγο οι τεχνολογίες της πληροφορίας και της επικοινωνίας λαμβάνονται υπόψη από τις εκπαιδευτικές πολιτικές πολλών κρατών. Ένα από τα πλεονεκτήματα που προκύπτουν είναι ότι δίνεται η δυνατότητα εκπαίδευσης από απόσταση και με αυτό τον τρόπο παρέχονται ίσες εκπαιδευτικές ευκαιρίες.

Χαρακτηριστικό είναι ότι μεγάλες εταιρείες όπως η Google και η Yahoo έχουν διαμορφώσει ειδικές εκπαιδευτικές πλατφόρμες όπως η «Google Suite for Education» και η «Bing in the Classroom». Η υπηρεσία «Google Suite for Education» περιλαμβάνει δωρεάν εργαλεία όπως αλληλογραφία, παρουσιάσεις, ιστοτόπους, έγγραφα, ημερολόγιο. Επιπλέον η υπάρχει και η πλατφόρμα «Google Classroom» με σκοπό τη διανομή και παράδοση εργασιών και ηλεκτρονικών σημειώσεων στους μαθητές σχολείων με ηλεκτρονικό τρόπο χωρίς να απαιτείται χαρτί.

Η υπηρεσία «Bing in the Classroom» παρέχει διάφορες δυνατότητες στους εκπαιδευτικούς όπως δωρεάν σχέδια μαθήματος, εκπαιδευτικά video, πλούσιο φωτογραφικό υλικό, ασφαλή αναζήτηση στο διαδίκτυο χωρίς διαφημιστικά.

Ένα άλλο βασικό ζήτημα της εφαρμοσμένης γλωσσολογίας είναι η εκμάθηση ξένων γλωσσών ειδικά σε μεγάλης ηλικίας άτομα λόγω της πολυπολιτισμικής κοινωνίας

πολλών χωρών λόγω της παρουσίας μεγάλου αριθμού μεταναστών. Ένα εργαλείο που παρέχεται δωρεάν από τον Παγκόσμιο ιστό είναι η αυτόματη μετάφραση κειμένων αλλά και ολόκληρων ιστοσελίδων, το οποίο εκτός από τις εκπαιδευτικές δυνατότητες που παρέχει, δίνει πρόσβαση σε υπηρεσίες και σε πληροφορίες σε άτομα που δε γνωρίζουν την επίσημη γλώσσα μιας χώρας.

2.24. Συστήματα Γραφής

Τα συστήματα γραφής αποτελούν οπτικό τρόπο αναπαράστασης λέξεων και με αυτόν τον τρόπο μπορούν να αποτυπωθούν κείμενα, τεκμήρια, λογοτεχνικά έργα ώστε να μπορούν να διαβαστούν και να αναλυθούν από μελλοντικές γενιές. Τα συστήματα γραφής αποτυπώνουν μαρτυρίες, ιστορίες, παραδόσεις.

Ουσιαστικά περισσότερες πληροφορίες γνωρίζουμε για αρχαίους πολιτισμούς που είχαν αναπτύξει συστήματα γραφής και ελάχιστα έως καθόλου πληροφορίες διαθέτουμε για λαούς που δεν είχαν αναπτύξει αλφάβητο και γραφή. Αρχαία δείγματα γραφής έχουν διασωθεί σε πάπυρους και σε μαρμάρινες πλάκες.

2.24.1. Είδη συστημάτων γραφής

Μια πρώτη ταξινόμηση των συστημάτων γραφής είναι σε «σημασιογραφικά ή ιδεογραφικά» και «φωνογραφικά ή φωνητικά ή λογογραφικά».^{151 152}

- Σημασιογραφικά είναι τα συστήματα όπου κάθε οπτικός χαρακτήρας τους αντιστοιχεί σε μια σημασία. Τέτοια παραδείγματα συναντάμε στα αρχαία συστήματα γραφής όπως τα ιερογλυφικά της Αρχαίας Αιγύπτου. Υπάρχουν και σήμερα τέτοια συστήματα με χαρακτηριστικό παράδειγμα την κινέζικη γραφή η οποία αποτελείται από ιδεογράμματα και λογογράμματα. Ένα

¹⁵¹ http://www.komvos.edu.gr/glwssa/odigos/thema_a12/thema_pdf.pdf

¹⁵² De Saussure, F. (1979). *Μαθήματα Γενικής Γλωσσολογίας*. Αθήνα: Εκδόσεις Παπαζήση.

τέτοιου είδους σύστημα γραφής διαθέτει πολύ μεγάλο αριθμό χαρακτήρων όπου ένας τέτοιος χαρακτήρας μπορεί να αντιστοιχεί ακόμη και σε ένα όνομα ή αντικείμενο. Το κινέζικο σύστημα γραφής για παράδειγμα διαθέτει περίπου 4000 σύμβολα. Παράδειγμα στην κινέζικη γλώσσα το σύμβολο «人» σημαίνει «ανθρώπινο», το σύμβολο «日» σημαίνει «ήλιος» και το σύμβολο «山» σημαίνει «βουνό».

- Η δεύτερη κατηγορία είναι φωνογραφικά συστήματα γραφής για τα οποία το κάθε σύμβολο αντιστοιχεί σε κάποιο φώνημα ή μόρφημα. Το κάθε φωνογραφικό σύστημα γραφής περιλαμβάνει μικρότερο αριθμό συμβόλων σε σχέση με τα σημασιογραφικά συστήματα γραφής. Στα συστήματα αυτά οι χαρακτήρες αντιστοιχούν σε σύμφωνα ή φωνήεντα και συλλαβές, ενώ υπάρχουν και σύμβολα για τα σημεία στίξης τα οποία είτε τα βάζουμε ανάμεσα στις λέξεις και μας διευκολύνουν στην ανάγνωση είτε λειτουργούν ως σύμβολα επιτονισμού ενός φωνήεντος. Επίσης κάποια φωνογραφικά συστήματα γραφής είναι αριστερόστροφα καθώς η γραφή ξεκινάει από αριστερά προς τα δεξιά και κάποια δεξιόστροφα στα οποία η γραφή ξεκινάει από τα δεξιά προς τα αριστερά. Παραδείγματα αριστερόστροφων αλφάβητων είναι το ελληνικό, το λατινικό, το κυριλλικό και το κοπτικό αλφάβητο, ενώ παραδείγματα δεξιόστροφων αλφάβητων είναι το αραβικό, το εβραϊκό, το ερμαϊκό και το φοινικικό.

Τα φωνογραφικά συστήματα μπορούμε να τα ταξινομήσουμε σε συλλαβικά ή συλλαβάρια¹⁵³ και τα αλφαβητικά συστήματα γραφής.

- Στα συλλαβικά συστήματα γραφής τα σύμβολα αναπαριστούν τις πιθανές συλλαβές της κάθε γλώσσας. Συνήθως τέτοια συστήματα έχουν περίπου 100 σύμβολα. Στην Αρχαία Ελλάδα τέτοια συστήματα γραφής ήταν η Γραμμική Β' και το Κυπριακό Συλλαβάριο¹⁵⁴.
- Στα αλφαβητικά συστήματα γραφής κάθε σύμβολο αντιστοιχεί σ' ένα φώνημα. Τα σύμβολα τέτοιων συστημάτων χωρίζονται σε σύμφωνα και φωνήεντα. Στα αλφαβητικά συστήματα γραφής τα σύμβολα είναι λιγότερα από εκείνα των συλλαβικών συστημάτων γραφής και είναι συνήθως από 20 έως 30. Υπάρχουν και κάποια αλφάβητα όπου κάποια σύμβολα μπορεί να μην αντιστοιχούν σε φώνημα αλλά σε μόρφημα. Για παράδειγμα στο κυριλλικό

¹⁵³ http://www.greek-language.gr/greekLang/studies/history/thema_04/24.html

¹⁵⁴ http://www.greek-language.gr/greekLang/studies/history/thema_04/24.html

αλφάβητο το σύμβολο «я» προφέρεται «για», το σύμβολο «ю» προφέρεται «γιου» και το σύμβολο «ѣ» προφέρεται «γιο».

Κάποια από τα αλφάβητα διαθέτουν μόνο σύμφωνα και τα φωνήεντα εννοούνται και δεν δηλώνονται. Παράδειγμα τέτοιων αλφάβητων είναι το εβραϊκό αλφάβητο, επίσης το αραβικό και το αραμαϊκό αλφάβητο διαθέτουν μόνο τρία μακρά φωνήεντα. Σε κάποια άλλα αλφάβητα εκτός από τα φωνήεντα και τα σύμφωνα υπάρχουν σύμβολα τα οποία χρησιμεύουν στο να προφέρουμε κατάλληλα το προηγούμενο γράμμα από το επόμενο φωνήεν. Χαρακτηριστικό παράδειγμα αποτελεί το κυριλλικό αλφάβητο στο οποίο το σύμβολο «Ѣ» ονομάζεται σκληρό σημείο και χρησιμεύει στο διαχωρισμό σκληρού συμφώνου από επόμενο φωνήεν και το σύμβολο «Ь» το οποίο ονομάζεται μαλακό σημείο και δηλώνει ότι το προηγούμενο σύμφωνο είναι ουρανικό (μαλακό).

2.24.2. Σημαντικά αλφάβητα

Τα πιο σημαντικά αλφάβητα είναι:

- το λατινικό το οποίο αρχικά χρησιμοποιήθηκε για τη λατινική γραφή και αργότερα διαδόθηκε και σε άλλους λαούς. Σήμερα χρησιμοποιείται στις περισσότερες χώρες της Ευρώπης και της Αμερικής. Σύμφωνα με τον Γ. Μπαμπινιώτη το λατινικό αλφάβητο είναι το Αρχαίο Ελληνικό αλφάβητο της Χαλκίδας που μεταφέρθηκε στην Ιταλία και χρησιμοποιήθηκε από τη λατινική γλώσσα. Το λατινικό αλφάβητο είναι το επίσημο της Ευρωπαϊκής Ένωσης.
- Το ελληνικό το οποίο είναι από τα παλαιότερα καθώς χρησιμοποιείται από τον 8^ο αιώνα π.Χ.
- Το κυριλλικό αλφάβητο το οποίο χρησιμοποιείται από τις σλαβικές γλώσσες όπως τη Βουλγαρική, τη Ρωσική, τη Σερβοκροάτικη και την Ουκρανική γλώσσα. Επινόηθηκε από τους Αγίους Κύριλλο και Μεθόδιο τον 9^ο αιώνα μ.Χ. .
- Το αραβικό αλφάβητο το οποίο χρησιμοποιείται στις αραβικές χώρες της Ασίας και της Αφρικής. Τα παλαιότερα σωζόμενα τεκμήρια στην αραβική γλώσσα είναι του 4^{ου} αιώνα μ.Χ.

2.24.3. Σύστημα γραφής Braille

Ξεχωριστή περίπτωση συστήματος γραφής αποτελεί το σύστημα Μπράιγ το οποίο επινοήθηκε από τον Louis Braille για την ανάγνωση τυφλών. Μάλιστα ο ίδιος ο L. Braille ήταν τυφλός και δημιούργησε τον κώδικα αυτό γιατί τα υπάρχοντα συστήματα τον δυσκόλευαν. Στο σύστημα αυτό εκτός από φωνήεντα και τα σημεία στίξης διαθέτει σύμβολα και για τους αριθμούς. Όλα τα σύμβολα αναπαριστώνται με ανάγλυφες κουκίδες οι οποίες σχηματίζονται με τη βοήθεια κατάλληλα διαμορφωμένων γραφομηχανών πάνω σε ειδικό τύπο χαρτιού.

2.24.4. Οπτική αναγνώριση χαρακτήρων

Η οπτική αναγνώριση χαρακτήρων είναι μια εφαρμογή η οποία χρησιμοποιείται για την αναγνώριση κειμένου το οποίο προέρχεται από φωτογραφίες, και σαρωμένα έγγραφα. Βασικός σκοπός είναι η χρησιμοποίηση και επεξεργασία του κειμένου αυτού χωρίς να απαιτείται η πληκτρολόγηση του ξανά από την αρχή.

Υπάρχουν δύο είδη οπτικής αναγνώρισης κειμένου. Η μια λειτουργεί με αντιστοίχιση με έτοιμα πρότυπα όπου το λογισμικό προσπαθεί να ταιριάξει τους χαρακτήρες της εικόνας ή του σαρωμένου εγγράφου με έτοιμα αποθηκευμένα πρότυπα που αντιστοιχούν σε χαρακτήρες. Η διαδικασία αυτή είναι λίγο χρονοβόρα καθώς απαιτεί πολλές επαναλήψεις.

Άλλος τρόπος είναι η εξαγωγή χαρακτήρων που ονομάζεται και **Ευφυής Αναγνώριση Χαρακτήρων** (ICR – Intelligent Character Recognition). Σε αυτή την περίπτωση η αναγνώριση γίνεται βάσει κριτηρίων και κανόνων όπως οι γωνίες που έχει ένας χαρακτήρας, ο αριθμός των πλευρών του, οι ενώσεις κτλ.. Η μέθοδος αυτή είναι πιο ευρέως χρησιμοποιημένη σήμερα από την αντιστοίχιση ετοιμών προτύπων και χρησιμοποιείται από τις εφαρμογές αναγνώρισης χαρακτήρων που διαθέτουν τα tablet και τα κινητά τηλέφωνα.

Επιπλέον υπάρχει και η αναγνώριση χειρόγραφων κειμένων η οποία όμως απαιτεί μια πιο πολύπλοκη διαδικασία η οποία συνδυάζει τις δύο παραπάνω μεθόδους. Πολλές φορές πάντως υπάρχουν δυσκολίες στην οπτική αναγνώριση κειμένου είτε αυτό είναι χειρόγραφο, είτε σκαναρισμένη φωτογραφία. Αυτό συμβαίνει όταν το

κείμενο δεν είναι πολύ καθαρό ή η φωτογραφία δεν έχει πολύ καλή ποιότητα και καθαρότητα.

Σχετικό λογισμικό διαθέτει η google το οποίο διαθέτει δυνατότητα αναγνώρισης χαρακτήρων για περισσότερες από 248 γλώσσες και δέχεται σαρωμένα έγγραφα ή έγγραφα σε μορφή pdf. Επίσης υπάρχουν πολλές δωρεάν διαδικτυακές εφαρμογές οι οποίες μετατρέπουν pdf έγγραφα σε επεξεργάσιμη μορφή.

2.25. Τυπολογία

Η τυπολογία είναι ο υποκλάδος της γλωσσολογίας ο οποίος συγκρίνει μελετά τις διάφορες γλώσσες και προκειμένου να εντοπιστούν ομοιότητες και διαφορές όσον αφορά τη δομή τους. Επίσης αντικείμενο της τυπολογίας είναι ο εντοπισμός ιδιαίτερων χαρακτηριστικών σε μια γλώσσα.

Εκ πρώτης όψεως η κάθε γλώσσα είναι διαφορετική από μια άλλη. Είναι για παράδειγμα δύσκολο να βρεις ομοιότητες της Γαλλικής Γλώσσας με την Αραβική. Παρόλα αυτά υπάρχουν κάποια γενικά χαρακτηριστικά τα οποία είναι κοινά για πολλές ή και για όλες τις γλώσσες του κόσμου. Οι ομοιότητες που μπορεί να υπάρχουν σε μια γλώσσα μπορεί να είναι στη γραμματική ή το συντακτικό . Για παράδειγμα μια πιθανή ομοιότητα είναι τα ουσιαστικά να είναι κλιτό μέρος του λόγου ή να υπάρχει αντιστοιχία μεταξύ των χρόνων κλίσης ενός ρήματος ή των εγκλίσεων.

2.25.1. Ομοιότητες βάσει της γλωσσικής οικογένειας

Οι λόγοι για τους οποίους δυο γλώσσες μπορεί να έχουν κοινά σημεία είναι διάφοροι. Το προφανές σίγουρα είναι ότι οι γλώσσες αυτές μπορεί να ανήκουν στην ίδια γλωσσική οικογένεια. Οι γλώσσες που ανήκουν στην ίδια οικογένεια έχουν κοινά

χαρακτηριστικά και κοινές ρίζες. Για κάθε μια από αυτές τις οικογένειες υπάρχει μια αρχική γλώσσα η οποία ονομάζεται πρωτογλώσσα και από την οποία έχουν προέλθει όλες οι υπόλοιπες. Οι σημαντικότερες οικογένειες γλωσσών που υπάρχουν σήμερα είναι οι: Νιγηροκογκολεζικές γλώσσες, Αυστρονησιακές γλώσσες, Τρανς-Νέα Γουινέα γλώσσες, Ινδοευρωπαϊκές γλώσσες, Σινοτιβητικές γλώσσες, Αφροασιατικές γλώσσες, Νειλοσαχαρικές γλώσσες, Πάμα-Νιουνγκάν γλώσσες, Οτομανγκουεϊκές γλώσσες, Αυστροασιατικές γλώσσες, Σέπικ-Ραμού γλώσσες, Τάι-Καντάι γλώσσες, Τουπικές γλώσσες, Δραβιδικές γλώσσες Μάγια.

Ομοιότητες σε γλώσσες μπορεί να οφείλονται και σε λόγους γεωγραφικούς και γι' αυτό το λόγο πολλές φορές ομοιότητες εμφανίζονται σε γλώσσες γειτονικών χωρών, λαών που βρίσκονται σε μακρόχρονη στενή επαφή ή χωρών που βρίσκονται σε μια ευρύτερη περιοχή. Ομοιότητες επίσης μπορεί να οφείλονται σε κοινωνικούς ή ιστορικούς λόγους.

2.25.2. Ομοιότητες βάσει της μορφολογίας

Ένα από τα βασικά πεδία σύγκρισης γλωσσών είναι οι ομοιότητες στη μορφολογία τους. Έτσι λοιπόν μπορούμε να ταξινομήσουμε τις διάφορες γλώσσες σε:

- **Συγκολλητικές γλώσσες** οι οποίες είναι οι γλώσσες όπου λέξεις δημιουργούνται από τη συγκόλληση μορφημάτων. Στις λέξεις αυτές η συγκόλληση χρησιμοποιείται εκτεταμένα και μπορεί να αντιστοιχεί σε γραμματική πληροφορία όπως χρόνος κλίσης, αριθμός, έγκλιση. Επίσης οι πληροφορίες που μπορεί να παρέχει ένα τέτοιο πρόσφυμα μπορεί να είναι λεξικές. Π.χ. στην τουρκική γλώσσα η λέξη «arabalarım» σημαίνει «τα αυτοκινητά». Αν τη λέξη την αναλύσουμε προκύπτει:

- η λέξη «araba» που σημαίνει «αυτοκίνητο»
- το μορφήμα «lar» το οποίο υποδηλώνει πληθυντικό αριθμό
- το μόρφημα «ım» το οποίο αντιστοιχεί στο κτητικό «μου»

Παραδείγματα συγκολλητικών γλωσσών είναι η Ιαπωνική, η Μαλαισιανή, η Τουρκική, η Ουγγρική, η Φιλανδική και η Εσθονική.

- **Απομονωτικές Γλώσσες** οι οποίες είναι οι γλώσσες στις οποίες οι λέξεις αποτελούνται από ένα μόνο μόρφημα. Στις γλώσσες αυτές οι γραμματικές

σχηματίζονται από πολλές λέξεις. Χαρακτηριστικά παραδείγματα τέτοιων γλωσσών είναι η Κινέζικη και η Βιετναμέζικη

- **Πολυσυνθετικές γλώσσες** στις οποίες μια ολόκληρη πρόταση αποτελείται από μία λέξη. Κάθε λέξη-πρόταση περιέχει πολλά μορφήματα. Χαρακτηριστικό παράδειγμα είναι η γλώσσα των Εσκιμώων και των Μαορί.
- **Οι Κλιτές γλώσσες** που διαθέτουν μεγάλο αριθμό μορφημάτων. Από συνδυασμούς μορφημάτων προκύπτουν μορφοσυντακτικές κατηγορίες. Στις γλώσσες αυτές οι λέξεις χωρίζονται σε δύο κατηγορίες τις κλιτές και τις άκλιτες. Οι κλιτές λέξεις αποτελούνται από τη ρίζα και τις καταλήξεις. Επίσης κάποιες κλιτές λέξεις μπορεί να διαθέτουν πρόθεμα ή ένθεμα. Π.χ. Η λέξη «άνομος» αναλύεται ως εξής ά (πρόθεμα)-νομ (ρίζα)- ος (κατάληξη). Κλιτές γλώσσες είναι η Ελληνική, οι λατινογενείς και οι σλαβικές γλώσσες.

2.25.3. Ομοιότητες βάσει καθολικών χαρακτηριστικών

Στη σύγχρονη γλωσσολογία μεγάλο ενδιαφέρον υπάρχει για την ύπαρξη ομοιοτήτων και διαφορών που βασίζονται σε πιο γενικά χαρακτηριστικά τα οποία ισχύουν σε πολλές γλώσσες. Για παράδειγμα πολλές γλώσσες διαθέτουν μέρη του λόγου όπως ρήματα και ουσιαστικά.

Ο λόγος της ύπαρξης κοινών καθολικών χαρακτηριστικών σε διάφορες είναι ένα ερώτημα στο οποίο επιδέχονται διάφορες απαντήσεις. Μια εξήγηση θα μπορούσε να είναι ότι προέρχονται από την ανθρώπινη σκέψη και την επεξεργασία γλωσσικών δεδομένων από τον ανθρώπινο εγκέφαλο. Επίσης τα καθολικά χαρακτηριστικά ενισχύουν την άποψη ότι όλες οι γλώσσες προέρχονται από μια πρωτογλώσσα η οποία με την πάροδο του χρόνου διαφοροποιήθηκε εξαιτίας κοινωνικών, γεωγραφικών και οικονομικών λόγων αλλά κάποια βασικά χαρακτηριστικά διατηρήθηκαν.

Τα καθολικά χαρακτηριστικά των γλωσσών διακρίνονται σε κατηγορίες:

- Υπάρχουν καθολικά χαρακτηριστικά που ισχύουν για όλες τις γλώσσες. Τα καθολικά αυτά χαρακτηριστικά ονομάζονται απόλυτα. Για παράδειγμα τέτοιο χαρακτηριστικό αποτελεί η διάκριση σε σύμφωνα και φωνήεντα, η οποία είναι

χρήσιμη σε όλες τις γλώσσες του κόσμου. Επίσης σε όλες τις γλώσσες υπάρχουν αντωνυμίες και διάκριση μεταξύ ουσιαστικών και ρημάτων.

- υπάρχουν κάποια άλλα που ισχύουν για μεγάλο αριθμό γλωσσών, αλλά όχι για όλες τις γλώσσες. Τα καθολικά αυτά χαρακτηριστικά ονομάζονται στατιστικά . Παράδειγμα καθολικού στατιστικού χαρακτηριστικού είναι ότι στην πλειοψηφία των γλωσσών το υποκείμενο προηγείται του αντικειμένου.
- τέλος υπάρχουν και κάποια άλλα για τα οποία η παρουσία τους συνεπάγεται την παρουσία ενός άλλου καθολικού χαρακτηριστικού .

2.25.4. Cross-lingual Information Retrieval

Το μεγαλύτερο μέρος των ιστοσελίδων και με ποσοστό μεγαλύτερο του 50% είναι γραμμένο στην αγγλική γλώσσα, ενώ υπάρχουν ιστοσελίδες οι οποίες μπορεί να περιέχουν κείμενα σε δύο ή και περισσότερες διαφορετικές γλώσσες. Η διαγλωσσική ανάκτηση πληροφορίας καθιστά δυνατή την ανάκτηση σχετικών πληροφοριών σε διαφορετική γλώσσα από αυτή του ερωτήματος. Αυτό είναι πολύ χρήσιμο ιδίως σε χώρες που δεν υπάρχει υψηλό επίπεδο γλωσσομάθειας και κατά συνέπεια δεν είναι δυνατή η πρόσβαση σε πληροφορίες και πολυμεσικά δεδομένα τα οποία είναι γραμμένα σε διαφορετική γλώσσα¹⁵⁵.

Υπάρχουν διάφορες τεχνικές για να πραγματοποιηθεί κάτι τέτοιο.

- Ένας τρόπος είναι η μετάφραση του ερωτήματος σε άλλες γλώσσες κάνοντας χρήση λεξικών και γραμματικών .
- Μια άλλη όχι και τόσο καλή τεχνική είναι η μετάφραση ολόκληρου της ιστοσελίδας από τη πηγαία γλώσσα στη γλώσσα που είναι γραμμένο το ερώτημα. Το βασικό μειονέκτημα της μεθόδου αυτής είναι ότι θα πρέπει να μεταφραστεί μεγάλος αριθμός ιστοσελίδων και άρα είναι μια μέθοδος που δε συμφέρει υπολογιστικά.

¹⁵⁵ Chin, J., Heymans, M., Kojoukhov, A., Lin, J., & Tan, H. (2014). *U.S. Patent No. 8,799,307*. Washington, DC: U.S. Patent and Trademark Office.

- Μια άλλη μέθοδος είναι η ταυτόχρονη μετάφραση και του ερωτήματος και των ιστοσελίδων σε μια κοινή γλώσσα. Όμως και αυτή η μέθοδος δεν είναι η πιο ιδανική.

Τα προβλήματα και οι δυσκολίες που υπάρχουν στη διαγλωσσική ανάκτηση της πληροφορίας είναι παρόμοια με αυτά που υπάρχουν στη μηχανική μετάφραση. Δυσκολίες όπως πολλαπλά νοήματα σε λέξεις, νοηματικές αμφισημίες, λεξικά με ελλειπή στοιχεία, ύπαρξη διαφορετικής μορφολογίας από γλώσσα σε γλώσσα όπως για παράδειγμα ανάμεσα σε κάποιες συγκολλητικές γλώσσες και κλιτές γλώσσες. Από τα παραπάνω στοιχεία προκύπτει ότι η διαγλωσσική ανάλυση ανάμεσα σε κάποιες γλώσσες που ανήκουν σε διαφορετική οικογένεια και έχουν διαφορετική φιλοσοφία, δεν είναι μια εύκολη υπόθεση.

Σχετική δυνατότητα αναζήτησης πληροφοριών σε διαφορετικές γλώσσες διαθέτετε παλαιότερα και η Google, αλλά πλέον την έχει καταργήσει.

2.26.Διερμηνεία και μετάφραση

Η διερμηνεία και η μετάφραση έχουν πολλά κοινά σημεία καθώς και στους δύο αυτούς επιστημονικούς κλάδους πρέπει να αποδοθεί το νόημα ενός κειμένου από μια γλώσσα σε μια άλλη. Παρόλα αυτά οι δύο αυτές επιστήμες δεν ταυτίζονται απόλυτα. Στη διερμηνεία το κείμενο είναι προφορικό και ο διερμηνέας πρέπει να επεξεργάζεται πληροφορίες με πολύ μεγάλη ταχύτητα. Στη μετάφραση το κείμενο είναι γραπτό και ο μεταφραστής έχει περισσότερο χρόνο στη διάθεση του επεξεργαστεί γλωσσικές πληροφορίες και δεδομένα προκειμένου το κείμενο που θα προκύψει να δίνει στον αναγνώστη την έννοια του πρωτότυπου.

Πολλοί υποστηρίζουν ότι η διερμηνεία και η μετάφραση είναι ξεχωριστοί κλάδοι παρόλα αυτά υπάρχει και η άποψη ότι αποτελούν υποκλάδο της γλωσσολογίας καθώς και στους δύο πρέπει να πραγματοποιηθεί γλωσσολογική ανάλυση.

2.26.1. Η Μετάφραση

Στη μετάφραση πρέπει να αποδίδεται όσο το δυνατόν πιο πιστά το νόημα και η μορφή του αρχικού κειμένου σε μια άλλη γλώσσα γι' αυτό θα μπορούσαμε να πούμε ότι το προς μετάφραση κείμενο δεν αποτελεί ενότητα γραμματική αλλά σημασιολογική.¹⁵⁶ Ο μεταφραστής επιπλέον θα πρέπει να λάβει υπόψη του εκτός από το νόημα ενός κειμένου και τις μορφοσυντακτικές δομές του, καθώς επίσης και τις διαφορές στους κανόνες της σύνταξης και της μορφολογίας της γλώσσας στην οποία πρόκειται να μεταφραστεί ένα κείμενο. Με άλλα λόγια απαιτείται γλωσσολογική ανάλυση του αρχικού κειμένου σε: φωνητικό, φωνολογικό, μορφολογικό, συντακτικό, σημασιολογικό και πραγματολογικό επίπεδο.

Στη περίπτωση μετάφρασης ενός κειμένου από τη μια γλώσσα σε μια άλλη, το σημαίνον είναι ένα οποιοδήποτε γλωσσικό στοιχείο του κειμένου και σημαινόμμενο η αντίστοιχη ερμηνεία που του αποδίδεται και το συνδέει με κάποιο αντικείμενο, έννοια ή κατάσταση του πραγματικού ή φανταστικού κόσμου.¹⁵⁷ Πολλές φορές στη μετάφραση από τη μια γλώσσα σε μια άλλη υπάρχουν περιπτώσεις όπου ένα σημαίνον μπορεί να αντιστοιχεί σε δύο ή περισσότερα σημαινόμενα της γλώσσας στην οποία πρέπει να πραγματοποιηθεί η μετάφραση. Παράδειγμα η λέξη «τραπεζίτης» μπορεί σημασιολογικά να αναφέρεται στον «ιδιοκτήτη τράπεζας» ή «σε δόντι στο πίσω μέρος του στόματος»¹⁵⁸. Με τη χρήση των κατάλληλων τεχνικών αλλά σε πολλές περιπτώσεις από τα συμφραζόμενα μπορεί να περιοριστεί το πρόβλημα αυτό. Επίσης υπάρχουν και οι περιπτώσεις συνωνυμιών όπου σε ένα σημαινόμμενο μπορεί να αντιστοιχούν παραπάνω από ένα σημαίνοντα. Σε αυτή την περίπτωση ο μεταφραστής πρέπει να επιλέξει την κατάλληλη λέξη λαμβάνοντας υπόψη το ύφος και το περιεχόμενο του υπολοίπου κειμένου.

Ζητούμενο κατά τη μετάφραση είναι η ύπαρξη σημασιολογικής και υφολογικής ισοδυναμίας. Αυτό σημαίνει ότι σ' ένα άρθρο ή ένα λογοτεχνικό κείμενο θα πρέπει να χρησιμοποιούνται μορφήματα λέξεις και εκφράσεις που να διατηρούν το συγκεκριμένο ύφος.

¹⁵⁶ Κεντρώτης, Γ. (1996). *Θεωρία και πράξη της Μετάφρασης*, Αθήνα: Δίαυλος. Σελ. 30.

¹⁵⁷ Κρίμπας, Π. (2017). *Εισαγωγή στη Θεωρία της Μετάφρασης*, Αθήνα: Γρήγορη. σελ. 41.

¹⁵⁸ <https://el.wiktionary.org/wiki/τραπεζίτης>.

Στην μετάφραση υπάρχουν περιπτώσεις όπου είναι δύσκολο να αποδίδεται το 100% του νοήματος του κειμένου. Βέβαια το ποσοστό του νοήματος το οποίο δεν αποδίδεται αποτελεί αμελητέα ποσότητα σε σχέση με το συνολικό κείμενο. Υπάρχουν επίσης περιπτώσεις όπου δεν υπάρχουν λέξεις για την απόδοση του νοήματος μιας λέξης σε μια γλώσσα. Χαρακτηριστικό παράδειγμα στη βιβλιογραφία αποτελεί η λέξη «χιόνι» η οποία δε μπορεί να αποδοθεί με τη χρήση κάποιας λέξης σε γλώσσες κάποιων λαών της Αφρικής. Στην μετάφραση ισχύει η «αρχή της συμβατικότητας» όπου σύμφωνα με τον Ferdinand de Saussure, η αντιστοιχία μεταξύ λέξεων και σημασιών αποτελούν άτυπες συμβάσεις μεταξύ ομάδων ανθρώπων.

2.26.2. Οι μεταφραστικές μονάδες

Τα βασικά επίπεδα της δομής του κειμένου είναι δύο: το λεκτικό επίπεδο και το δομικό επίπεδο. Το λεκτικό επίπεδο αφορά τις λέξεις τις φράσεις κάθε γλώσσας, ενώ το δομικό επίπεδο τους συντακτικούς, γραμματικούς και μορφολογικούς κανόνες που ισχύουν σε κάθε γλώσσα¹⁵⁹.

Μέσα σε αυτά τα πλαίσια είναι απαραίτητο να βρεθούν οι ελάχιστες μεταφραστικές μονάδες που συνθέτουν το κείμενο. Οι μεταφραστικές μονάδες είναι τμήματα του συνολικού κειμένου τα οποία διαθέτουν συνοχή και δεν πρέπει να μεταφράζονται ξεχωριστά.¹⁶⁰

Η κατάτμηση του κειμένου σε μεταφραστικές μονάδες είναι επιπλέον πολύ χρήσιμο εργαλείο της μηχανικής μετάφρασης και της γλωσσολογίας σωμάτων κειμένων. Οι μεταφραστικές μονάδες μπορούν να ορίζονται με διάφορα κριτήρια ανάλογα το είδος του κειμένου και των κριτηρίων που έχει θέσει ο γλωσσολόγος ή ο μεταφραστής. Κάποια από τα βασικά κριτήρια είναι τα εξής:

- Βάσει του λεξιλογίου
- Του επιπέδου συναρμογής, δηλαδή βάσει της μορφολογίας και της σύνταξης της γλώσσας πηγής και της γλώσσας προορισμού.

¹⁵⁹ Κρίμπας, Π. (2017). *Εισαγωγή στη Θεωρία της Μετάφρασης*, Αθήνα: Γρήγορη. σελ. 65.

¹⁶⁰ Τσίγκου, Μ. (2011). *Γλωσσολογία & Μετάφραση*. Αθήνα: Δίαυλος.

- Βάσει του νόηματος, δηλαδή η κάθε μεταφραστική μονάδα να έχει ένα ολοκληρωμένο νόημα.¹⁶¹

Επιπλέον η μεταφραστική μονάδα μπορεί να εξαρτηθεί βάσει κάποιων δομικών κριτηρίων τα οποία προσδίδουν αυτονομία κατά τη μετάφραση. Τέτοια δομικά σχήματα μπορεί να είναι:

- Μια γραμματική μονάδα
- Ένα ρηματικό ή ονοματικό κατηγορημα
- Ένα σύνταγμα. Π.χ. Στην όμορφη πλατεία.
- Μια συγκεκριμένη έκφραση στην οποία αποδίδεται συγκεκριμένο νόημα.

Επιπλέον υπάρχουν και ερευνητές όπου θεωρούν μεταφραστική μονάδα μια ολόκληρη πρόταση¹⁶² ή ένα ολόκληρο κείμενο¹⁶³.

2.26.3. Η Σύνδεση με ανάκτηση πληροφορίας

2.26.3.1. Η Μηχανική Μετάφραση

Η μηχανική μετάφραση αφορά την αυτόματη μετάφραση λέξεων και φράσεων από μια γλώσσα σε μια άλλη. Οι δυσκολίες είναι πολλές όπως για παράδειγμα η ύπαρξη πολλών σημασιών σε κάποιες λέξεις, η ύπαρξη αμφισημιών, η ύπαρξη μεταφορικών φράσεων, η διαφορετική μορφολογία και γραμματική ανάμεσα σε κάποιες γλώσσες. Παρόλα αυτά η ποιότητα και η ταχύτητα στην αυτόματη μετάφραση βελτιώνεται συνεχώς.

Τα τελευταία χρόνια υπάρχει μια στροφή των μεταφραστών προς τα σώματα κειμένου. Γι' αυτό το λόγο πολλά μεταφραστικά εργαλεία βασίζονται σε γλωσσικά δείγματα που προέρχονται από κείμενα της γλώσσας του πρωτότυπου κειμένου και της γλώσσας στην οποία θέλουμε να μεταφράσουμε το κείμενο. Κάποιες βασικές λειτουργίες που βασίζονται σε σώματα κειμένων μπορούν να ενσωματωθούν στα λογισμικά αυτά ώστε η μετάφραση να είναι όσο το δυνατόν περισσότερο ακριβής.

¹⁶¹ Τσίγκου, Μ. (2011). *Γλωσσολογία & Μετάφραση*. Αθήνα: Δίαυλος.

¹⁶² Τσίγκου, Μ. (2011). *Γλωσσολογία & Μετάφραση*. Αθήνα: Δίαυλος. Σελ 61

¹⁶³ Τσίγκου, Μ. (2011). *Γλωσσολογία & Μετάφραση*. Αθήνα: Δίαυλος. Σελ 63.

Βασική προϋπόθεση είναι στα σώματα κειμένων να υπάρχουν υπογλώσσες¹⁶⁴ δηλαδή ομάδες δεδομένων που μπορούν να συμβάλουν στη δημιουργία υπολογιστικών βοηθημάτων και υπολογιστικών λεξικών¹⁶⁵. Με τη βοήθειά των υπογλωσσών μπορούν να απομονώνονται κάποια βασικά στοιχεία μιας φυσικής γλώσσας τα οποία μπορούν να συμβάλουν στη μετάφραση μέσω της δημιουργίας συντακτικών και γραμματικών κανόνων που μπορούν να ενσωματωθούν σε μεταφραστικά λογισμικά.

Κάποιες από αυτές τις λειτουργίες είναι η δυνατότητα διερεύνησης του τρόπου χρήσης ενός συγκεκριμένου λήμματος σε μια γλώσσα και πιθανές λέξεις ή μορφήματα του το συνοδεύουν σε μεγάλη συχνότητα. Επίσης στη μετάφραση πολύ χρήσιμη είναι η χρήση λιστών που περιέχουν τις συχνότητες εμφάνισης μιας λέξης σε μεγάλο αριθμό κειμένων. Αυτό μπορεί να βοηθήσει στην εύρεση του κατάλληλου λήμματος για την απόδοση της σημασίας μιας λέξης από τη μια γλώσσα στην άλλη. Λογισμικά αυτόματης μετάφρασης υπάρχουν στο διαδίκτυο με δυνατότητα μετάφρασης σε πολλές γλώσσες και παρόλα τα προβλήματα οι δυνατότητες και η ακρίβεια βελτιώνονται συνεχώς. Η Google εκτός από την εφαρμογή google translator διαθέτει και δυνατότητα μετάφρασης του περιεχομένου μιας ολόκληρης ιστοσελίδας, ενώ η Bing κάνει χρήση του μεταφραστή της Microsoft.

2.26.3.2. Η Αυτοματοποιημένη Διερμηνεία

Υπάρχει και δυνατότητα αυτοματοποιημένης διερμηνείας αλλά είναι κάτι όχι τόσο απλό ακόμη καθώς είναι δυνατή με τη χρήση βίντεο διαμεσολάβησης.

Τα τελευταία χρόνια όμως έχουν κατασκευαστεί και συσκευές οι οποίες μπορούν να πραγματοποιήσουν διερμηνεία σε κάποιες γλώσσες ενώ η Google πολύ πρόσφατα κυκλοφόρησε εφαρμογή η οποία καθιστά δυνατή τη διερμηνεία φωνής με σχετική υποστήριξη σε πραγματικό χρόνο για 27 διαφορετικές γλώσσες. Η αναγνώριση της γλώσσας ομιλίας γίνεται αυτόματα, ενώ υπάρχει και η δυνατότητα μετάφρασης διαλόγου καθώς εμφανίζει μεταφρασμένο κείμενο το λόγου που αρθρώνει ο συνομιλητής.

¹⁶⁴ Υπογλώσσα είναι η γλώσσα που χρησιμοποιεί μια συγκεκριμένη ομάδα ομιλητών οι οποίοι μπορεί να ασχολούνται με ένα συγκεκριμένο θέμα ή να ασκούν συγκεκριμένο επάγγελμα.

¹⁶⁵ Σαριδάκης, Ι. (2010). *Σώματα Κειμένων και Μετάφραση*. Αθήνα: Παπαζήση. Σελ 33.

2.27. Σώματα κειμένων

Τα σώματα κειμένου αποτελούνται από συλλογές κειμένων σε ηλεκτρονική μορφή με σκοπό την γλωσσολογική τους ανάλυση. Τα τελευταία χρόνια η ανάγκη ύπαρξης γλωσσικών δεδομένων συγκεκριμένου τύπου καθώς και η ανάγκη ύπαρξης δεδομένων που προέρχονται από κείμενα έχει στρέψει την προσοχή της γλωσσολογίας προς τα σώματα κειμένου. Τα σώματα κειμένου εκτός άλλων περιέχουν και πρωτότυπα γλωσσικά δεδομένα τα οποία μιλήθηκαν σε φυσικές συνθήκες και όχι στα πλαίσια ενός πειράματος ή μιας έρευνας.

Η επεξεργασία και μελέτη των σωμάτων κειμένου μπορεί να δώσει απαντήσεις σε διάφορα ερωτήματα που αφορούν τη γλωσσολογία ή διάφορους υποκλάδους της όπως η μορφολογία η σύνταξη, η κοινωνιογλωσσολογία κ.α. Για παράδειγμα μπορούν να προκύψουν συμπεράσματα που αφορούν συνήθη ορθογραφικά λάθη μαθητών που μαθαίνουν δεύτερη γλώσσα, την αναζήτηση πλήθους συγκεκριμένων λέξεων ή γραμματικών φαινομένων και λέξεων που τα συνοδεύουν σε λογοτεχνικά κείμενα κ.α. Επίσης με τη χρήση των σωμάτων κειμένων μπορούμε να εξετάσουμε κάποιες συγκεκριμένες περιπτώσεις γλωσσικών δεδομένων και να δημιουργήσουμε κανόνες. Π.χ. τη χρήση ενός βοηθητικού ρήματος.

Σε σώματα κειμένων μπορεί να γίνει γλωσσολογική μελέτη. Για παράδειγμα μπορούμε να αναζητήσουμε όλα τα ουσιαστικά με κατάληξη –α. Επίσης από τα σώματα κειμένου και την ηλεκτρονική επεξεργασία τους μπορούν να προκύψουν πίνακες συχνοτήτων λέξεων , συχνοτήτων σημασιών των λέξεων και τις συσχετίσεων λέξεων με άλλες λέξεις¹⁶⁶.

2.27.1. Κριτήρια δημιουργίας σωμάτων κειμένου

Κάθε σώμα κειμένου πρέπει να πληρεί κάποιους συγκεκριμένους κανόνες και γι' αυτό η επιλογή ενός σώματος γίνεται βάσει κάποιων κριτηρίων. Από τα βασικότερα κριτήρια είναι η αντιπροσωπευτικότητα καθώς το σώμα κειμένου πρέπει να αποτελεί αντιπροσωπευτικό δείγμα της γλώσσας από την οποία προέρχεται.

¹⁶⁶ Φραντζή,Κ.(2012). *Εισαγωγή στην επεξεργασία σωμάτων κειμένων*. Αθήνα: Εκδοτικός Όμιλος Ίων.

Τέτοια κριτήρια είναι εσωτερικά και εξωτερικά. Στα εσωτερικά περιλαμβάνεται η σύνταξη, το λεξιλόγιο και οι γραμματικοί κανόνες. Τα εξωτερικά κριτήρια περιλαμβάνουν γλωσσικές εκφράσεις και σχετίζονται με το κοινωνικό περιβάλλον επικοινωνίας. Επίσης κριτήριο δημιουργίας ενός σώματος κειμένου μπορεί να είναι η θεματολογία του. Για παράδειγμα το περιεχόμενο του μπορεί να είναι φιλοσοφικό, λογοτεχνικό, θεολογικό, νομικό κτλ.

Σημαντικό επίσης είναι ένα σώμα κειμένου να είναι αξιόπιστο. Την αξιοπιστία μπορούν να την καθορίσουν τα μεταδεδομένα του και ιδίως αυτά τα οποία σχετίζονται με το κείμενο όπως: η ημερομηνία δημοσίευσης, η χρονική περίοδος σύνταξης, το όνομα του δημιουργού, το όνομα του εκδότη, το μέσω δημοσίευσης (βιβλίο, περιοδικό, εφημερίδα), ο τόπος δημοσίευσης, η θεματική του κατηγορία, πληροφορικά στοιχεία από άλλες εκδοχές του, ο αριθμός των λέξεων, ο αριθμός των προτάσεων, ο αριθμός των παραγράφων.

Τα σώματα κειμένου μπορούμε να τα εμπλουτίσουμε με γλωσσολογικές πληροφορίες που αφορούν τη γραμματική ή τη σύνταξη του κειμένου και σε αυτή την περίπτωση τα ονομάζουμε επισημειωμένα¹⁶⁷.

2.27.2. Οι Υπογλώσσες

Τα περισσότερα σώματα κειμένου περιλαμβάνουν ένα ή περισσότερα τμήματα δεδομένων κειμένου τα οποία είναι γραμμένα με βάση μια υπογλώσσα. Γενικά μια υπογλώσσα αποτελεί υποσύνολο της φυσικής γλώσσας με τους δικούς της γραμματικούς και συντακτικούς κανόνες και το δικό της λεξιλόγιο το οποίο μπορεί να διαθέτει εξειδικευμένη ορολογία και ειδικά μορφήματα.

Οι συντακτικοί κανόνες κάποιων υπογλωσσών καθώς και η ορολογία τους, πολλές φορές μπορεί να είναι άγνωστη στη γενική γλώσσα και μπορεί να μη χρησιμοποιείται σε άλλη υπογλώσσα. Παρόλο που η κάθε υπογλώσσα διαθέτει τους δικούς της κανόνες, ενδέχεται επιπλέον να χρησιμοποιεί μέρος ή και το σύνολο των κανόνων της φυσικής γλώσσας

¹⁶⁷ Φραντζή,Κ.(2012). *Εισαγωγή στην επεξεργασία σωμάτων κειμένων*. Αθήνα: Εκδοτικός Όμιλος Ίων.

Γενικώς μια υπογλώσσα διαθέτει.¹⁶⁸

- Οριοθετημένο θεματικό πεδίο.
- Λεξικολογικούς, συντακτικούς και σημασιολογικούς περιορισμούς
- Αποκλίνοντες γραμματικό – συντακτικούς κανόνες
- Κειμενική δομή
- Χρήση ειδικών συμβόλων

2.27.3. Είδη σωμάτων κειμένων

Τα σώματα κειμένου μπορούν να ταξινομηθούν σε πολλές κατηγορίες. Ανάλογα το είδος της πληροφορίας που περιέχουν υπάρχουν σώματα κειμένου ιστορικά, φιλοσοφικά, λογοτεχνικά κ.α. Ανάλογα με τις γλώσσες που περιλαμβάνει το κείμενο υπάρχουν μονόγλωσσα και πολύγλωσσα σώματα κειμένου.

Τα πολύγλωσσα μπορούν να χωριστούν σε μικτά και παράλληλα. Τα παράλληλα σώματα κειμένου περιέχουν ένα ή περισσότερα κείμενα σε μια γλώσσα καθώς και τις μεταφράσεις τους σε κάποια άλλη γλώσσα. Αυτό καθιστά τα παράλληλα σώματα κειμένου πολύ χρήσιμα για τη δημιουργία λογισμικού μηχανικής μετάφρασης.

Τα σώματα κειμένων επίσης μπορούμε να τα ταξινομήσουμε σε επισημειωμένα (annotated) και τα μη επισημειωμένα (un-annotated) . Τα επισημειωμένα είναι εμπλουτισμένα με γλωσσολογική πληροφορία σε αντίθεση με τα μη επισημειωμένα . Οι επισημειώσεις συνήθως αφορούν γραμματικό φαινόμενο ή συντακτικό τύπο .

2.27.4. Σύνδεση με την ανάκτηση της πληροφορίας

Τα σώματα κειμένου μπορούν να επεξεργαστούν και να σαρωθούν ηλεκτρονικά. Αυτό προσφέρει μεγαλύτερη ταχύτητα και αμεσότητα στην ανάκτησή τους καθώς επίσης γρήγορη ταχύτητα στην αναζήτηση ενός λήμματος, μιας φράσης κάτι το οποίο δε συμβαίνει στην χειρονακτική αναζήτηση .

¹⁶⁸ Σαριδάκης, Ι. (2010). *Σώματα Κειμένων και Μετάφραση*. Αθήνα: Παπαζήση.

2.27.4.1. Ανάλυση των σωμάτων κειμένου

Μια μέθοδος ανάλυσης κειμένου είναι αυτή των *n* γραμμάτων (*n*-gram) με την οποία πραγματοποιείται κατάτμηση του κειμένου σε ίσα τμήματα τα οποία μπορεί να αποτελούνται από χαρακτήρες, λέξεις, προτάσεις¹⁶⁹. Όπου *n* είναι ο αριθμός των χαρακτήρων που διαθέτει κάθε τέτοιο τμήμα. Όταν τα *n* τμήματα αποτελούνται από χαρακτήρες υπολογίζονται ως χαρακτήρες και τα σημεία στίξης όπως και τα κενά.

Π.χ. έστω ότι θέλω να αναλύσω την παρακάτω πρόταση: «He plays the piano». Η ανάλυση της σε 2 - gram είναι η εξής: [He], [_p], [la], [ys], [_t], [he], [_p], [ia], [no].

2.27.4.2. Ηλεκτρονικά σώματα κειμένου

Υπάρχουν στο διαδίκτυο προσβάσιμα πολλά σώματα κειμένου σε ηλεκτρονική μορφή. Αυτά τα σώματα κειμένου συνήθως συνοδεύονται από λογισμικό με το οποίο μπορούμε να τα επεξεργαστούμε και να ανακτήσουμε γλωσσική πληροφορία όπως για παράδειγμα συχνότητα εμφάνισης μιας λέξης.

Σημαντικά σώματα κειμένων είναι ο «Εθνικός Θησαυρός Ελληνικής Γλώσσας» ο οποίος περιέχει μέχρι στιγμής περισσότερες από 47.000.000 λέξεις και από τον οποίο μπορεί να γίνει αναζήτηση με χρήση λέξης, λήμματος ή μέρος του λόγου¹⁷⁰. Άλλο σημαντικό ελληνικό σώμα κειμένων είναι η «Πύλη της Ελληνικής Γλώσσας» του Κέντρου Ελληνικής Γλώσσας (ΚΕΓ). Σημαντικά ξενόγλωσσα σώματα κειμένων είναι το «Oxford English Corpus» το οποίο είναι από τα μεγαλύτερα σώματα κειμένων και περιέχει 2,1 δισεκατομμύρια λέξεις, χρησιμοποιείται στο λεξικό της Οξφόρδης και κάθε κείμενο του συνοδεύεται από μεταδεδομένα. Άλλο σημαντικό σώμα κειμένου είναι το American National Corpus το οποίο περιέχει 22 εκατομμύρια λέξεις.

¹⁶⁹ Αναστοπούλου, Ε. (2015). *Σύνθεση πληροφορίας στην πιστοποίηση γραφέα με ηλεκτρονικό κείμενο: μελέτη των n-grams* (Doctoral dissertation).

¹⁷⁰ http://hnc.ilsp.gr/index.php?current_page=main

2.27.4.3. Κανονικές εκφράσεις

Για την επεξεργασία της φυσικής γλώσσας είναι απαραίτητη η τυπική περιγραφή των δομικών της χαρακτηριστικών. Για την περιγραφή ακολουθιών χαρακτήρων πολύ χρήσιμες είναι οι κανονικές εκφράσεις οι οποίες αποτελούν έναν εναλλακτικό τρόπο περιγραφής μιας γλώσσας. Κανονική έκφραση ορισμένη σε ένα αλφάβητο Σ είναι μια λέξη του αλφαβήτου $\Sigma \cup \{ \cup, *, (,), \emptyset \}$. Με τη χρήση κανονικών εκφράσεων υπάρχει δυνατότητα αναζήτησης τμήματος κειμένου, χαρακτήρων και λέξεων με τη χρήση γλώσσας προγραμματισμού που υποστηρίζει κανονικές εκφράσεις. Γλώσσες προγραμματισμού όπως η Java, η C++, η C, η Python κ.α.

Μια κανονική έκφραση μπορεί να περιγραφεί και με τη χρήση ενός πεπερασμένου αυτόματου. Τα πεπερασμένα αυτόματα είναι απλές υπολογιστικές μηχανές οι οποίες διαθέτουν πεπερασμένο αριθμό καταστάσεων.

Ένα ντετερμινιστικό πεπερασμένο αυτόματο είναι μια πεντάδα $(Q, \Sigma, \delta, q_0, F)$ που αποτελείται από ένα πεπερασμένο σύνολο καταστάσεων Q , ένα αλφάβητο εισόδου Σ , μια συνάρτηση μετάβασης, μια αρχική κατάσταση q_0 και ένα σύνολο τελικών καταστάσεων F , υποσύνολο του Q ¹⁷¹.

2.28. Ανθρωπογλωσσολογία

Ο κλάδος αυτός της γλωσσολογίας αφορά τη σχέση του ανθρώπου και της γλώσσας του με τον πολιτισμό του¹⁷². Τον τρόπο με τον οποίο η γλώσσα επηρεάζει την κουλτούρα και τον πολιτισμό και διαμορφώνει κοινωνικές ομάδες. Στην ανθρωπογλωσσολογία επίκεντρο είναι ο άνθρωπος, οι μύθοι, οι παραδόσεις και οι συνήθειες κοινωνιών και η καθημερινότητά τους.

¹⁷¹ https://el.wikipedia.org/wiki/Ντετερμινιστικό_πεπερασμένο_αυτόματο

¹⁷² Τσιτσιπής, Λ.(2004). *Εισαγωγή στην Ανθρωπολογία της γλώσσας*. Αθήνα: Gutenberg.

Ένας από τους ρόλους των ανθρωπογλωσσολόγων είναι η καταγραφή γλωσσών που δεν έχουν αλφάβητο και είναι κυρίως προφορικές. Τέτοιου είδους γλώσσες πολλές φορές είναι υπό εξαφάνιση. Επιπλέον αντικείμενα ενδιαφέροντος στην ανθρωπογλωσσολογία είναι ο τρόπος με τον οποίο εκφράζεται σε μια γλώσσα το συναίσθημα και οι ιδεολογίες των μελών ενός κοινωνικού συνόλου.

2.28.1. Σχέση γλώσσας - κουλτούρας

Η συμπεριφορά και οι δραστηριότητες των μελών μιας γλωσσικής κοινωνίας σύμφωνα με πολλούς γλωσσολόγους σχετίζονται μεταξύ τους. Πολλοί γλωσσολόγοι όπως ο Γερμανός Von Humboldt υποστηρίζουν ότι η γλώσσα που μιλάνε τα μέλη μιας κοινότητας σχετίζεται με το πως βλέπουν τον κόσμο¹⁷³ και στο ίδιο συμπέρασμα καταλήγουν και άλλοι γλωσσολόγοι συγκρίνοντας διαφορετικές γλώσσες μεταξύ τους.

Ο Whorf συγκρίνοντας τη γλώσσα μιας Ινδιάνικης φυλής με διάφορες ευρωπαϊκές γλώσσες παρατήρησε ότι ενώ στις ευρωπαϊκές γλώσσες υπάρχει η διάκριση του χρόνου σε παρελθόν, παρόν και μέλλον, στη γλώσσα της φυλής Hopi ο χρόνος έχει μια συνέχεια και υπάρχει διαφορά όσον αφορά τη διάρκεια και την ένταση. Η διαφορετικότητα αυτή οφείλεται στις συνήθειες στην καθημερινότητα των διαφορετικών αυτών κοινωνιών και αποδεικνύει τη σχέση που έχει η γλώσσα με τοπικά, εθνολογικά και θρησκευτικά χαρακτηριστικά του κάθε λαού.

2.28.2. Σχέση γλώσσας γεωγραφίας

Επιπλέον πολύ σημαντικός είναι ο ρόλος του κλίματος, της γεωγραφίας ο οποίος καθορίζει το λεξιλόγιο μιας γλώσσας. Στη γλώσσα Inuit των Εσκιμώων η λέξη χιόνι μπορεί να αναπαρασταθεί με 4 διαφορετικούς τρόπους καθώς υπάρχει διαφορετική λέξη για το χιόνι που πέφτει, διαφορετική για το στεγνό χιόνι, για το λερωμένο χιόνι,

¹⁷³ Τσιτσιπής, Λ.(2004). *Εισαγωγή στην Ανθρωπολογία της γλώσσας*. Αθήνα: Gutenberg.

για το χιόνι στο έδαφος, επίσης υπάρχουν δεκατρείς διαφορετικές λέξεις στα φιλιπινέζικα που αναφέρονται στο ρύζι

Κάθε κοινότητα λοιπόν είναι ξεχωριστή και ενδεχομένως τα μέλη της σκέφτονται διαφορετικά από τα αντίστοιχα άλλων κοινοτήτων καθώς τα γεωγραφικά, τα θρησκευτικά και πολιτισμικά χαρακτηριστικά της κάθε κοινότητας δημιουργούν διαφορετικές ανάγκες.

Η γλώσσα ενός πολιτισμού μπορεί να περιέχει λέξεις οι οποίες μπορεί να μην υπάρχουν καθόλου στο λεξιλόγιο άλλων λαών – γλωσσών λόγω κουλτούρας και γεωγραφίας. Επίσης ενώ σε κάποιες γλώσσες που ανήκουν σε λαούς κάποιες έννοιες περιγράφονται με τη χρήση μιας λέξης σε κάποιους άλλους λαούς και γλώσσες μπορεί να περιγράφονται με δύο - τρείς λέξεις ή με μια πρόταση

2.28.3. Σχέση γλώσσας και σκέψης

Όπως προαναφέρθηκε σύμφωνα με τους Sapir Whorf η γλώσσα καθορίζει τη γνωστική λειτουργία και τη σκέψη. Αυτό σημαίνει ότι υπάρχουν διαφορές στη σκέψη και τις γνωστικές λειτουργίες σε ομάδες ατόμων που ομιλούν διαφορετικές γλώσσες. Κάθε γλώσσα επιβάλλει τη δική της κοσμοθεώρηση στους ομιλητές της

Οι λεξικολογικές διαφορές που υπάρχουν σε διάφορες γλώσσες δε σημαίνουν και διαφορά στην αντίληψη. Αντίθετα αντιστοιχούν σε διαφορές στις εμπειρίες και στην εξειδίκευση. Επίσης υπάρχουν και διαφορές στη γραμματική από γλώσσα σε γλώσσα. Για παράδειγμα στην κινέζικη γλώσσα δεν υπάρχει ο υποθετικός λόγος όπως σε άλλες γλώσσες και για υποθέσεις χρησιμοποιούνται πολύ σύνθετες δομές.

Επίσης η διαφορά στην γλώσσα μπορεί να σημαίνει και διαφορά στην απόκτηση διαφορετικών δεξιοτήτων. Παράδειγμα είναι τα μαθηματικά στην κινέζικη γλώσσα. Οι Κινέζοι στο αριθμητικό τους σύστημα πρέπει να γνωρίζουν τους αριθμούς από το 0 έως το 10. Αν θέλουν να αναφερθούν σ' ένα μεγαλύτερο αριθμό πρέπει να το κάνουν με σύνθετο τρόπο . Αν για παράδειγμα θέλουν να πουν τον αριθμό 14 λένε δέκα και τέσσερα. Σε αυτό το λόγο οφείλεται το γεγονός ότι οι Κινέζοι μπορούν με ευκολότερο τρόπο να κάνουν πράξεις με αριθμούς μεγαλύτερους του δέκα και κατά συνέπεια μπορούν να αποκτήσουν ευκολότερα μαθηματικές δεξιότητες.

2.28.4. Geographic information retrieval

Η Ανθρωπολογία ως επιστήμη μελετάει πολιτισμικά στοιχεία λαών τα οποία είναι άρρηκτα συνδεδεμένα με γεωγραφικά στοιχεία του τόπου στον οποίο ζουν .

Υπάρχει υποκλάδος της Ανάκτησης της Πληροφορίας ο οποίος ασχολείται με την ανάκτηση δεδομένων με γεωγραφική αναφορά. Σχετική είναι και η ύπαρξη γεωγραφικών πληροφοριακών συστημάτων Geographic Information System (GIS) που χρησιμοποιούνται για την εισαγωγή, ανάκτηση, διαχείριση, ανάλυση και απόδοση γεωγραφικών δεδομένων προκειμένου να επιτευχθεί η διαχείριση χρήσεων γης, φυσικών πόρων, περιβαλλοντολογικών στοιχείων κ.α.

Η Γεωγραφική Ανάκτηση της Πληροφορίας παρέχει πληροφορίες όπως πόσο δημοφιλής είναι ένα προϊόν σε μια περιοχή, σε ποιες χώρες παράγονται κάποια προϊόντα, στοιχεία που αφορούν είδη πανίδας και χλωρίδας μιας χώρας καθώς και άλλες κοινωνικού, οικονομικού και πολιτισμικού περιεχομένου πληροφορίες για κάποιο συγκεκριμένο μέρος του πλανήτη.

Απαραίτητα δομικά στοιχεία για τη δημιουργία ενός τέτοιου συστήματος είναι η ύπαρξη μιας βάσης δεδομένων με τα κατάλληλα στοιχεία, ένα ευρετήριο, ένα πρόγραμμα κατάταξης των αποτελεσμάτων και το κατάλληλο γραφικό περιβάλλον.¹⁷⁴

Σε ένα Γεωγραφικό Σύστημα Ανάκτησης Πληροφορίας συνήθως απαραίτητη είναι η διαδικασία του GeoTagging , δηλαδή της προσθήκη μεταδεδομένων που αφορούν γεωγραφικά στοιχεία όπως π.χ. τις γεωγραφικές συντεταγμένες και άλλα γεωγραφικά αναγνωριστικά στοιχεία¹⁷⁵ .

Κάποιες φορές στο σχεδιασμό μπορεί να προκύπτουν αμφισημίες λόγω πόλεων ή χωριών με ονομασία που παραπέμπει σε κάποιο όνομα ή πόλεων που έχουν το ίδιο όνομα με κάποιες άλλες πόλεις. Για παράδειγμα η Τουρκία (Turkey) στην Αγγλική γλώσσα σημαίνει γαλοπούλα, η Νότιος Αφρική είναι και χώρα αλλά ταυτόχρονα και

¹⁷⁴ Hariharan, R., Hore, B., Li, C., & Mehrotra, S. (2007, July). Processing spatial-keyword (SK) queries in geographic information retrieval (GIR) systems. In *19th International Conference on Scientific and Statistical Database Management (SSDBM 2007)* (pp. 16-16). IEEE.

¹⁷⁵ Luo, J., Joshi, D., Yu, J., & Gallagher, A. (2011). Geotagging in multimedia and computer vision—a survey. *Multimedia Tools and Applications*, 51(1), 187-211.

μια ευρύτερη περιοχή και πόλη η Newcastle υπάρχει στη Μεγάλη Βρετανία, αλλά και στη Νότια Αυστραλίας.¹⁷⁶

Κεφάλαιο 3^ο

Συμπεράσματα

Ο λόγος είτε γραπτός, είτε προφορικός αποτελεί βασικό προϊόν της ανάκτησης της πληροφορίας αφού το διαδίκτυο αποτελεί ένα τεράστιο χώρο συσσώρευσης δεδομένων κειμένου και οι ψηφιακές βιβλιοθήκες περιέχουν κυρίως δεδομένα κειμένου. Με βάση αυτά, η ανάπτυξη αποτελεσματικών υπολογιστικών συστημάτων ανάκτησης της πληροφορίας τα οποία θα μπορούν να ανιχνεύουν το σημασιολογικό περιεχόμενο εγγράφων στο διαδίκτυο και σε ψηφιακές βιβλιοθήκες, αποτελεί σημαντική πρόκληση. Ο ρόλος της επιστήμης της γλώσσας και τον υποκλάδων της αποτελεί πολύ σημαντικό παράγοντα για την ανάπτυξη τέτοιων εφαρμογών.

Η ανθρώπινη γλώσσα διαφέρει από τις υπόλοιπες γλώσσες λόγω των διάφορων λειτουργιών που πραγματοποιεί όπως : η επικοινωνία, ως μέσω συνεργασίας και διάδρασης μεταξύ κοινωνικών ομάδων. Σύμφωνα με τον Noam Chomsky «*γλώσσα είναι ένα σύνολο (πεπερασμένο ή μη) προτάσεων, κάθε μια από τις οποίες είναι πεπερασμένη σε έκταση και κατασκευασμένη από πεπερασμένο σύνολο στοιχείων*»¹⁷⁷.

Οι δομικές ιδιότητες της γλώσσας σύμφωνα με τον Chomsky είναι κάτι το αφηρημένο και εξειδικευμένο. Κάποιες από τις ιδιότητες της γλώσσας είναι η ευελιξία και η δυνατότητα τροποποίησης , η ανεξαρτησία από το ερέθισμα. Επίσης η γλώσσα διαθέτει βιολογική διάσταση και περιεχόμενο καθώς μας επηρεάζει ψυχοσωματικά υπό την έννοια ότι κάποιες λέξεις ή φράσεις που ακούμε μπορεί να μας δημιουργούν συναισθήματα όπως θυμός, ικανοποίηση, έκπληξη κτλ. . Η γλώσσα μας δίνει τη δυνατότητα να σκεφτόμαστε μέσω αυτής συνδυάζοντας πράγματα και έννοιες με συγκεκριμένες λέξεις.

¹⁷⁶ Amitay, E., Har'El, N., Sivan, R., & Soffer, A. (2004, July). Web-a-where: geotagging web content. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 273-280). ACM.

¹⁷⁷ Lyons, J. (1999). Εισαγωγή στη γλωσσολογία. Μ. Αραποπούλου, Α. Αρχάκης, Αθήνα : εκδ. Πατάκης.

Η γλωσσολογία είναι ένας επιστημονικός κλάδος με πολύ μεγάλη ιστορία αλλά παρόλα αυτά με την καθιέρωση της ως ξεχωριστό επιστημονικού πεδίου άρχισαν να αυτονομούνται και οι διάφοροι υποκλάδοι της. Εκτός από τους παραπάνω βασικούς υποκλάδους τα τελευταία χρόνια γίνεται λόγος και για «Αναπτυξιακή Γλωσσολογία», «Κλινική Γλωσσολογία», «Βιογλωσσολογία», «Γεωγραφική Γλωσσολογία», «Ανθρωπονομία», «Αρθρωτική Σύνθεση», «Αξίωμα της Κατηγορίας», «Βιοσημειωτική», «Κριτική Ανάλυση του Λόγου», «Εκπαιδευτική Γλωσσολογία» και η λίστα συνεχώς μεγαλώνει.

Για κάθε έναν από τους υποκλάδους, εξαιτίας της τεχνολογικής και πληροφοριακής προόδου καθώς και της μεγάλης πληθώρας εφαρμογών ανάκτησης της πληροφορίας, είναι δυνατό να βρεθεί συσχετισμός με την ανάκτηση της πληροφορίας. Βάσει όλων αυτών ανάγεται το συμπέρασμα ότι η ανάκτηση της πληροφορίας βασίζεται στη γλωσσολογία και δανείζεται πολλά στοιχεία από την θεωρία της.

Με την έλευση του σημασιολογικού ιστού αλλάζει πλέον η ανάκτηση της πληροφορίας και αυξάνεται ο αριθμός των προσφερόμενων υπηρεσιών ανάκτησης, με τη χρήση των ευφύων πρακτόρων που επικοινωνούν και ανταλλάσσουν πληροφορίες. Ο αριθμός των ιστοσελίδων πλέον έχει αυξηθεί δραματικά και μεγάλη πρόκληση πλέον αποτελεί η παροχή των κατάλληλων υπηρεσιών ανάκτησης.

Βιβλιογραφία

Paper

Amitay, E., Har'El, N., Sivan, R., & Soffer, A. (2004, July). Web-a-where: geotagging web content. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 273-280). ACM.

Baddeley, A. (1992). *Working memory*. *Science*, 255(5044), 556-559.

Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., & Su, Z. (2007, May). Optimizing web search using social annotations. In *Proceedings of the 16th international conference on World Wide Web* (pp. 501-510). ACM.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, 284(5), 28-37.

Bird, S. (1994). Introduction to computational phonology. *Computational Linguistics*, 20(3), 1-8.

Chin, J., Heymans, M., Kojoukhov, A., Lin, J., & Tan, H. (2014). *U.S. Patent No. 8,799,307*. Washington, DC: U.S. Patent and Trademark Office.

Duranti, A. (Ed.). (2009). *Linguistic anthropology: A reader* (Vol. 1). John Wiley & Sons.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303-306.

Geeraerts, D., & Cuyckens, H. (2007). Introducing cognitive linguistics. In *The Oxford handbook of cognitive linguistics*.

Hariharan, R., Hore, B., Li, C., & Mehrotra, S. (2007, July). Processing spatial-keyword (SK) queries in geographic information retrieval (GIR) systems. In *19th*

International Conference on Scientific and Statistical Database Management (SSDBM 2007) (pp. 16-16). IEEE.

Horowitz, D., & Kamvar, S. D. (2010, April). The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web* (pp. 431-440). ACM.

Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137-142). Springer, Berlin, Heidelberg.

Koutsogiannis, D., & Mitsikopoulou, B. (2003). Greeklish and Greekness: Trends and discourses of “glocalness”. *Journal of computer-mediated communication*, 9(1), JCMC918.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, 22(1), 1-38.

Luo, J., Joshi, D., Yu, J., & Gallagher, A. (2011). Geotagging in multimedia and computer vision—a survey. *Multimedia Tools and Applications*, 51(1), 187-211.

MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of child language*, 12(2), 271-295.

Miller, J. W. (1998). *U.S. Patent No. 5,805,911*. Washington, DC: U.S. Patent and Trademark Office.

Navigli, R. (2009). Word sense disambiguation: A survey. *ACM computing surveys (CSUR)*, 41(2), 10.

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization. *Culture theory: Essays on mind, self, and emotion*, 276-320.

Satow, M., & Widman, J. M. (2013). *U.S. Patent No. 8,521,743*. Washington, DC: U.S. Patent and Trademark Office.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*.

Wetherell, M., Taylor, S., & Yates, S. J. (2001). *Discourse theory and practice: A reader*. Sage.

Yarowsky, D. (1995). *Unsupervised word sense disambiguation rivaling supervised methods*. In *33rd annual meeting of the association for computational linguistics*.

Γεωργίου, Γ. Β. (2011). Γλώσσα και ισχύς σε μια δίκη φόνου στην Κύπρο: η χρήση της Διαλέκτου και η εναλλαγή κώδικα. *Selected papers on theoretical and applied linguistics*, 19, 491-498.

Ζαχάκου, Ζ. (2016). *Πραγματολογία και προβλήματα λόγου – τρόποι αντιμετώπισης*. *Πανελλήνιο Συνέδριο Επιστημών Εκπαίδευσης*, 2015(1), 482-491.

Κέρδη Jr, JH, Brown, NG, Mezey, Pδ, & Colombo, LM (2003). *US Patent No. 6,636,831*. Washington, DC: Γραφείο Διπλωμάτων Ευρεσιτεχνίας και Εμπορικών Σημάτων των ΗΠΑ.

Μπατσαλιά, Φ., Σελλά-Μάζη, Ε., & Μπαμπινιώτης, Γ. (1994). *Γλωσσολογική προσέγγιση στη θεωρία και τη διδακτική της μετάφραση: μορφο-συντακτικές, λεξιλογικές, σημασιολογικές, υφολογικές και πραγματολογικές μετατοπίσεις κατά τη διαδικασία της μετάφρασης...* Ιόνιο Πανεπιστήμιο.

Ευδόπουλος, Γ. Ι. (2016). Η γλωσσολογία σήμερα στόχοι, αντιλήψεις εφαρμογές και επαγγελματικές εφαρμογές. *Επιστημονική Επετηρίδα Τμήματος Φιλολογίας*, 35.

Στάμου, Α.Γ. (2014). Η κριτική ανάλυση του: Μελετώντας τον ιδεολογικό ρόλο της γλώσσας. Στο *Ανάλυση λόγου: Θεωρία και εφαρμογές*, Μ. Γεωργαλίδου, Μ. Σηφianού, Β. Τσάκωνα (επιμ.). Νήσος: Αθήνα, 149-187.

Βιβλία

Auroux, S.(2005). Η φιλοσοφία της γλώσσας. Αθήνα: Μεταίχμιο.

Austin, J. L. (1975). *How to do things with words*. Oxford university press.

Baeza -Yates, R. Ribeiro – Neto , B.(2015). *Ανάκτηση Πληροφορίας Οι έννοιες και η τεχνολογία πίσω από την αναζήτηση*. Θεσσαλονίκη: Εκδόσεις Τζιόλα.

Coulthard, M., Johnson, A., & Wright, D. (2016). *An introduction to forensic linguistics: Language in evidence*. Routledge.

De Saussure, F. (1979). *Μαθήματα Γενικής Γλωσσολογίας*. Αθήνα: Εκδόσεις Παπαζήση.

Gipfert, J., Himmelmann, N., & Mosel, U. (Eds.). (2006). *Essentials of language documentation* (Vol. 178). Walter de Gruyter.

Harley, T. (2008). *Η Ψυχολογία της Γλώσσας*. Θεσσαλονίκη: Univercity Studio Press.

Ladefoged, P. (2007). Εισαγωγή στη φωνητική. *Μ. Μπαλτατζάνη (μτφ.)*. Αθήνα: Εκδόσεις Πατάκης.

Lyons, J. (1999). Εισαγωγή στη γλωσσολογία. *Μ. Αραποπούλου, Α. Αρχάκης*, Αθήνα: Εκδόσεις Πατάκης.

Manning, C., Raghavan, P., Schutze, H. (2012). *Εισαγωγή στην ανάκτηση πληροφοριών*. Αθήνα: Κλειδάριθμος.

Phillips, L., & Jorgensen, M. W. (2009). Ανάλυση Λόγου. Θεωρία και Μέθοδος. *Α. Κιουπκιολής*, Αθήνα: Παπαζήσης.

Βοσνιάδου, Σ. (2004). *Γνωσιακή Επιστήμη: Η Νέα Επιστήμη του Νου*. Αθήνα: Gutenberg.

Κεντρώτης, Γ. (1996). *Θεωρία και πράξη της Μετάφρασης*. Αθήνα: Διάλογος.

Κρίμας, Π. (2017). *Εισαγωγή στη Θεωρία της Μετάφρασης*. Αθήνα: Γρήγορη.

Μπαμπινιώτης, Γ. (1985). *Εισαγωγή στη Σημασιολογία*. Αθήνα.

Μπαμπινιώτης, Γ. (1991). *Γλωσσολογία και Λογοτεχνία από την Τεχνική στην Τέχνη του Λόγου*. Αθήνα.

Μπέλλα, Σ.(2015). *Πραγματολογία, Από τη γλωσσική επικοινωνία στη γλωσσική διδασκαλία*. Αθήνα: Gutenberg.

Πανάρτου,Ε.(2011). *Θέματα Γνωσιακής Γλωσσολογίας*.Αθήνα :Εκδόσεις Παπαζήση.

Παυλίδου, Θ. (2008). *Επίπεδα γλωσσικής ανάλυσης*. Αθήνα: Ινστιτούτο Νεοελληνικών Σπουδών (Ιδρυμα Μανόλη Τριανταφυλλίδη).

Σαριδάκης, Ι. (2010). *Σώματα Κειμένων και Μετάφραση*. Αθήνα: Παπαζήση.

Στάμου, Α. (2014). Η κριτική ανάλυση λόγου: Μελετώντας τον ιδεολογικό ρόλο της γλώσσας. στο Μ., Γεωργαλίδου, Μ., Σηφιανού και Β., Τσάκωνα,(επιμ.), *Ανάλυση λόγου. Θεωρία και εφαρμογές, νήσος, Αθήνα*. Στάμου, Α., Γρίβα Ε., Δημητρίου, Π και Τσακιρίδου, Ε.

Τάντος, Α., Μαρκαντωνάτου, Σ., ... & Κυριακοπούλου, Π. (2015). *Υπολογιστική γλωσσολογία*.

Τσίγκου, Μ. (2011). *Γλωσσολογία & Μετάφραση*. Αθήνα: Δίαυλος.

Τσιτσιπής, Λ.(2004). *Εισαγωγή στην Ανθρωπολογία της γλώσσας*.Αθήνα: Gutenberg.

Φραντζή,Κ.(2012). *Εισαγωγή στην επεξεργασία σωμάτων κειμένων*. Αθήνα: Εκδοτικός Όμιλος Ίων.

Φιλιππάκη, Ε, W.(1992). *Εισαγωγή στη θεωρητική γλωσσολογία*. Αθήνα: Νεφέλη.

Ιστοσελίδες - Ηλεκτρονικά Βιβλία

Chapman, S. (Ed.). (2009). *Key ideas in linguistics and the philosophy of language*. Edinburgh University Press.

Gavrilidou, M., Rehm, G., & Uszkoreit, H. (2012). *The Greek language in the digital age*. Springer. Διαθέσιμο στη : <http://www.meta-net.eu/whitepapers/volumes/e-book/greek.pdf>

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167. Διαθέσιμο στο : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.244.9480&rep=rep1&type=pdf>.

Searle, J. R., & Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press. Διαθέσιμο στο: https://www.corwin.com/sites/default/files/upm-binaries/23507_hutchby~Vol_1_Ch_02.pdf

Steedman, M.(2014). *Computational Linguistics*. Ανακτήθηκε την : Ιανουάριο 12, 2019, Διαθέσιμο στην : <https://plato.stanford.edu/entries/computational-linguistics/>

Themistocleous, C. (2011). Introduction to Phonetics and Phonology [in Greek]. Open University of Cyprus. Διαθέσιμο στο:

https://www.researchgate.net/publication/258883078_Introduction_to_Phonetics_and_Phonology_in_Greek/link/0f31753348ff3e2bcd000000/download.

Μαγουλάς, Γ., Μαγουλά, Ε., 2017. *Θέματα Συγχρονικής και Διαχρονικής Γλωσσολογίας*. [ηλεκτρ. βιβλ.] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/6509>

Πούλος, Μ., 2015. *Σημασιολογική επεξεργασία της πληροφορίας*. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/2854>

Ρούσσου, Α. 2015. Γλώσσα και γραμματική. [Κεφάλαιο Συγγράμματος]. Στο Ρούσσου, Α. 2015. *Σύνταξη*. [ηλεκτρ. βιβλ.] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. κεφ 1. Διαθέσιμο στο: <http://hdl.handle.net/11419/582>

Διπλωματικές – Διατριβές:

Αναστοπούλου, Ε. (2015). *Σύνθεση πληροφορίας στην πιστοποίηση γραφέα με ηλεκτρονικό κείμενο: μελέτη των n-grams* (Doctoral dissertation).

Mikros, G., & Μικρός, Γ. (2015). Υπολογιστική υφολογία

Αναστοπούλου, Ε. (2015). *Σύνθεση πληροφορίας στην πιστοποίηση γραφέα με ηλεκτρονικό κείμενο: μελέτη των n-grams* (Doctoral dissertation).

Θεοδόση-Κόκκινου, Α. (2013). *Τεχνητά νευρωνικά δίκτυα και εφαρμογές στα συστήματα αυτόματου ελέγχου* (Doctoral dissertation).

Λιβιέρης, Ι. (2008). *Αποτίμηση μεθόδων εκπαίδευσης τεχνητών νευρωνικών δικτύων και εφαρμογές* (Doctoral dissertation).

Μπόρας, Ι. (2009). *Αυτόματος τεμαχισμός ψηφιακών σημάτων ομιλίας και εφαρμογή στη σύνθεση ομιλίας, αναγνώριση ομιλίας και αναγνώριση γλώσσας* (Doctoral dissertation).

Ιστότοποι:

AICBT(). *Online Authorship Attribution Tool*. Ανακτήθηκε : Ιούλιο 7, 2019, Διάθεσιμο στην : <http://www.aicbt.com/authorship-attribution/online-software/>

App Store(). iWordQ US. Ανακτήθηκε: Φεβρουάριο 8, 2019, Διαθέσιμο στην: <https://apps.apple.com/us/app/iwordq-us/id557929840>

Bana, Y. (2014). *Is the Internet Accelerating Language Development?*. Ανακτήθηκε την: Ιανουάριος 21, 2019. Διαθέσιμο στην: <https://www.translatemedia.com/translation-blog/internet-accelerating-language-development/>.

Biometric Solutions.(). Speaker Recognition. Ανακτήθηκε : Φεβρουάριο 14, 2019, Διαθέσιμο στην: <http://www.biometric-solutions.com/speaker-recognition.html>

Birner, B. (2019). *Does the language I speak influence the way I think?*. Ανακτήθηκε : Φεβρουάριο 15, 2019, Διαθέσιμο στην: <https://www.linguisticsociety.org/content/does-language-i-speak-influence-way-i-think>.

Copernic.(). Copernic Summarizer . Ανακτήθηκε: Ιούνιο 1, 2019. Διαθέσιμο στην : <https://www.copernic.com/en/products/summarizer/>

Crystal, D.(). Books and Articles. Ανακτήθηκε :Ιούλιο 27, 2019. Διαθέσιμο στην: <http://www.davidcrystal.com/books-and-articles>.

Dean, B. (2019). Google's 200 Ranking Factors: The Complete List 2019. Ανακτήθηκε: Απρίλιο 22, 2019. Διαθέσιμο στην : <https://backlinko.com/google-ranking-factors>.

EDC. (). What is word prediction. Ανακτήθηκε : Μάρτιο 1, 2019, Διαθέσιμο στην : http://www2.edc.org/ncip/library/wp/What_is.htm.

Quetext. ().Plagiarism Checker & Citation Assistant. Ανακτήθηκε: Ιούλιο 14, 2019. Διαθέσιμο στην : <https://www.quetext.com/>.

MonkeyLearn.(). *Text Classification*. Ανακτήθηκε: Ιούνιο 15, 2019, Διαθέσιμο στην: <https://monkeylearn.com/text-classification/>.

Princeton Univercity.(). Wordnet. Ανακτήθηκε την: Μάρτιο 11, 2019. Διαθέσιμο στην : <https://wordnet.princeton.edu/>.

Resoomer . Ανακτήθηκε : Ιούλιο 1, 2019. Διαθέσιμο στην: <https://resoomer.com/en/>

Sadaoki.F. (2010). *Speaker Recognition in Smart Environments*. Ανακτήθηκε : Μάιος 27,2019, Διαθέσιμο στην: <https://www.sciencedirect.com/topics/computer-science/speaker-recognition>.

Small Seo Tools.(). Plagiarism Checker. Ανακτήθηκε :Ιούλιο 7, 2019. Διαθέσιμο την: <https://smallseotools.com/plagiarism-checker/>

Spellbetterapp.(). *Spell Better for iPad - Literacy Support for Dyslexia*. Ανακτήθηκε : Αύγουστος 2, 2019, Διαθέσιμο στην: <https://www.spellbetterapp.com/>

Stanford Univercity.(). Computational Semantics. Ανακτήθηκε: Απρίλιο 25, 2019, Διαθέσιμο στην: <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1162/handouts/Computational-Semantics.pdf>.

Tucker, G ,R .(2019).*Applied Linguistics*. Ανακτήθηκε : Φεβρουάριο 26, 2019, Διαθέσιμο στην : <https://www.linguisticsociety.org/resource/applied-linguistics>.

-, (2015). The Phonological Loop. Ανακτήθηκε: Δεκέμβριο 19,2019, Διαθέσιμο στην : <https://workingmemoryrst.wordpress.com/the-phonological-loop>.

Διόφαντος.(). Φωνολογία, Ανακτήθηκε : Ιανουάριο 22, 2019, Διαθέσιμο στην : <http://ebooks.edu.gr/modules/ebook/show.php/DSGYM-A112/621/4005,17967/>

Θεοδωροπούλου, Μ. (). *Ψυχισμός και Γλώσσα*. Ανακτήθηκε : Μάρτιος 3, 2019, Διαθέσιμο στην : http://www.komvos.edu.gr/glwssa/odigos/thema_a5/a_5_thema.htm.

Ινστιτούτο Επεξεργασίας του Λόγου. (2019). *Εθνικός Θησαυρός Ελληνικής Γλώσσας*. Ανακτήθηκε: Απρίλιο 5, 2019, Διαθέσιμο στην : http://hnc.ilsp.gr/index.php?current_page=main.

Καθημερινή.(2014). Μια γλώσσα πεθαίνει κάθε δύο εβδομάδες λόγω ανάπτυξης. Ανακτήθηκε : Μάρτιο 26, 2019, Διαθέσιμο στην : <https://www.kathimerini.gr/782251/article/epikairothta/episthmh/mia-glwsa-pe8ainei-ka8e-dyo-evdomades-logw-anapty3hs>

Καραόλη, Μ.(2007). *Θέματα Ιστορίας της Ελληνικής Γλώσσας :Γλώσσα και γραφή*. Ανακτήθηκε : Μάρτιο 24, 2019, Διαθέσιμο στην : http://www.greek-language.gr/greekLang/studies/history/thema_04/24.html.

Κέντρο Ελληνικής Γλώσσας (2008). Κατανόηση Προφορικού Λόγου. Ανακτήθηκε: Ιανουάριο 5, 2019, Διαθέσιμο στην : http://www.greek-language.gr/greekLang/modern_greek/foreign/education/proposals/01a.html.

Πάντση, Ζ. (2012). Ιστορία της Γραφής. Ανακτήθηκε : Μάιο 23, 2019. Διαθέσιμο στην: <http://medusa.libver.gr/jspui/bitstream/123456789/3786/6/IstoriaGrafis.pdf>

Παπαναστασίου,Γ. (-). *Γλώσσα και Γραφή*. Ανακτήθηκε : Μάιο 20, 2019. Διαθέσιμο στην: http://www.komvos.edu.gr/glwsa/odigos/thema_a12/thema_pdf.pdf.

Yule, G.(1996). *ΠραγματολογίαΕισαγωγικά Σχόλια*. Ανακτήθηκε: Μάιο 29,2019, Διαθέσιμο στην: <http://www1.aegean.gr/social-anthropology/canakis/Courses%20offered/Pragmatics/Yule-chapter%201.htm>

-. *Φίλτρα αναζήτησης και αντικατάστασης*. Ανακτήθηκες την: Ιανουάριο 23, 2019. Διαθέσιμο στην : <https://support.google.com/analytics/answer/1034834?hl=el>