



**Πολυτεχνείο Κρήτης**

---

Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Τηλεπικοινωνιών

Διπλωματική Εργασία

# **Βέλτιστη Επιλογή Κειμένου για Σύστημα Σύνθεσης Φωνής**

**Από**

**Διονύσιο Κ. Παπαϊωάννου**  
(Α.Μ. 2000030022)

---

**Χανιά, Σεπτέμβριος 2006**

Διπλωματική Διατριβή

Σεπτέμβριος 12, 2006

# Βέλτιστη Επιλογή Κειμένου για Σύστημα Σύνθεσης Φωνής

**Διονύσιος Κ. Παπαϊωάννου**

Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών  
Πολυτεχνείο Κρήτης  
(Α.Μ. 2000030022)

**Επιβλέπων**

Καθ. Βασίλειος Διγαλάκης,  
Πολυτεχνείο Κρήτης

# Πολυτεχνείο Κρήτης

Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών

Οι υπογράφωντες πιστοποιούν ότι διάβασαν και υποδεικνύουν στην Πολυτεχνική Σχολή των Ηλεκτρονικών Μηχανικών και Υπολογιστών την αποδοχή της διπλωματικής διατριβής με τίτλο: «Βέλτιστη επιλογή κειμένου για σύστημα σύνθεσης φωνής» του Διονυσίου Κ. Παπαϊωάννου ως μέρος των προϋποθέσεων της απόκτησης του διπλώματος του Ηλεκτρονικού Μηχανικού και Μηχανικού Υπολογιστών.

Ημερομηνία: Σεπτέμβριος 2006

Επιβλέπων:

Καθ. Βασίλειος Διγαλάκης

Αναγνώστες:

Αν. Καθ. Αλέξανδρος Ποταμιάνος

Αν. Καθ. Αθανάσιος Λιάβας

# Πολυτεχνείο Κρήτης

---

Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών

Ημερομηνία: Σεπτέμβριος 2006

Συγγραφέας: **Διονύσιος Κ. Παπαϊωάννου**

Τίτλος: **Βέλτιστη Επιλογή Κειμένου για Σύστημα Σύνθεσης Φωνής**

Τμήμα: **Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών**

Δίπλωμα: **Δίπλωμα Μηχανικού** Μήνας: **Σεπτέμβριος** Χρονιά: **2006**

Το Πολυτεχνείο Κρήτης κατέχει την άδεια να μπορεί να διακινεί και να παρέχει αντίγραφα του παραπάνω τίτλου για μη εμπορικούς σκοπούς, όπου αυτό κρίνει φρόνιμο, έπειτα από αίτημα ιδιώτη ή πανεπιστημιακού ιδρύματος.

---

Υπογραφή του Συγγραφέα

Απαγορεύεται η αναπαραγωγή οποιαδήποτε τμήματος της διατριβής με οποιαδήποτε μέσο (φωτοτυπία, εκτύπωση, μικροφίλμ, αποθήκευση σε αρχείο πληροφοριών ή άλλη μηχανική ή ηλεκτρονική μέθοδο) χωρίς την έγγραφη άδεια του συγγραφέα.

*Στους γονείς μου Κωνσταντίνο και Ελένη...*

---

# Ευχαριστίες

Αυτή η πολύ ενδιαφέρουσα διπλωματική διατριβή ολοκληρώθηκε έπειτα από εκατοντάδες ώρες δημιουργικής εργασίας και πολλαπλών δοκιμών. Καθ' όλη τη διάρκεια της πραγμάτωσης της υπήρχε αμέριστη υποστήριξη από τον τομέα Τηλεπικοινωνιών του τμήματος των Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης κάτω από την επίβλεψη του Καθηγητή Βασίλειου Διαγιάκη. Χωρίς την συνεχή του ενθάρρυνση και εμπιστοσύνη σε μένα, θα ήταν αδύνατον να ολοκληρωθεί.

Θα επιθυμούσα να ευχαριστήσω τον Χρήστο Βοζνίδη, υποψήφιο διδάκτορα του τμήματος των Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών, για την βοήθεια και την αδιάκοπη υποστήριξη που μου προσέφερε, χωρίς την οποία η εργασία αυτή δεν θα είχε δημιουργηθεί. Ανεκτίμητης αξίας υπήρξε η βοήθεια του Ιωάννη Κλασινά, Μεταπτυχιακού φοιτητή του τμήματος των Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών πάνω σε αρκετά ζητήματα της εργασίας, καθώς με τις ιδέες του βοήθησε στην ολοκλήρωση αυτής.

Παράλληλα, θα ήθελα να ευχαριστήσω όλους τους φίλους μου για την πίστη τους στις ικανότητες μου και για τις αναζωογονητικές στιγμές που περάσαμε μαζί καθ' όλη τη διάρκεια της εργασίας μου. Τέλος, θα ήθελα να ευχαριστήσω τον πατέρα μου, Κωνσταντίνο Παπαϊωάννου, την μητέρα μου Ελένη, τον αδερφό μου Παναγιώτη και την μικρή μου αδερφή Μαριάννα για την ψυχολογική τους υποστήριξη, υπομονή και εμπιστοσύνη σε μένα. Είμαι περήφανος που έχω μια τέτοια οικογένεια η οποία μου

χάρισε πολύτιμες συμβουλές, έμπνευση και υποστήριξη όταν πραγματικά την είχα ανάγκη.

---

# Περίληψη

Η πρόοδος της επιστήμης και της έρευνας στον τομέα της Σύνθεσης Φωνής είναι τεράστια από τότε που δημιουργήθηκε το MITalk, στα μέσα της δεκαετίας του 80. Στις μέρες μας, η χρήση των σύγχρονων ηλεκτρονικών υπολογιστών και των εξελιγμένων αλγορίθμων που χρησιμοποιούνται στον τομέα της σύνθεσης φωνής μας επιτρέπει να συνθέσουμε φωνή τέτοιας ποιότητας που δύσκολα μπορεί κανείς να την ξεχωρίσει από τον ανθρώπινο λόγο.

Η βέλτιστη επιλογή κειμένων και συγκεκριμένα προτάσεων από ένα σύνολο αρχείων, οι οποίες θα χρησιμοποιηθούν για την δημιουργία της βάσης δεδομένων των τμημάτων φωνής (Speech Segment Database) για ένα σύστημα σύνθεσης φωνής είναι το θέμα αυτής της διπλωματικής διατριβής. Η εργασία αρχίζει με μία περιγραφή της δομής ενός συστήματος σύνθεσης φωνής, αναλύοντας διεξοδικά τα τμήματα που συνθέτουν το σύστημα αυτό. Ακολουθώντας, επικεντρωνόμαστε στο τμήμα αυτό που αφορά την βέλτιστη επιλογή κειμένων για την δημιουργία της βάσης δεδομένων των τμημάτων φωνής και θέλοντας να κατανοήσει ο αναγνώστης τον τρόπο που επιτυγχάνεται αυτή, περιγράφουμε την λειτουργία και την δομή των αρχείων λογισμικού (software modules) που μας βοηθούν να πετύχουμε τον στόχο αυτό. Στο σημείο αυτό γίνεται περιγραφή των διαφόρων αλγορίθμων που υλοποιήθηκαν για να πετύχουμε την βέλτιστη αυτή επιλογή. Τέλος, επιθυμώντας να δείξουμε την προσφορά αυτής της βέλτιστης επιλογής κειμένων στο αποτέλεσμα της σύνθεσης φωνής παραθέτουμε διάφορα στατιστικά δεδομένα και αποτελέσματα της εργασίας αυτής συγκρίνοντας τους αλγορίθμους που υλοποιήθηκαν για την επιλογή των προτάσεων.



---

# Περιεχόμενα

Ευχαριστίες .....	6
Περίληψη .....	8
Περιεχόμενα .....	9
<b>1. Εισαγωγή .....</b>	<b>11</b>
1.1 Σκοπός της Εργασίας αυτής .....	13
1.2 Περιγραμμά .....	14
<b>2. Συστήματα Σύνθεσης Φωνής.....</b>	<b>15</b>
2.1 Εισαγωγή στα Συστήματα Σύνθεσης Φωνής.....	15
2.2 Δομή ενός Συστήματος Σύνθεσης Φωνής.....	16
2.2.1 Υποσύστημα Επεξεργασίας Φυσικής Γλώσσας .....	17
2.2.2 Υποσύστημα Επεξεργασίας Φωνητικού Σήματος .....	22
2.2.2.1 Concatenative Σύνθεση Φωνής .....	23
<b>3. Βέλτιστη Επιλογή Κειμένων.....</b>	<b>27</b>
3.1 Προσέγγιση Προβλήματος.....	28
3.2 Επίλυση Προβλήματος.....	29

3.3 Σύστημα Βέλτιστης Επιλογής Προτάσεων .....	30
<b>4. Αποτελέσματα.....</b>	<b>46</b>
4.1 Αποτελέσματα κάθε Αλγορίθμου Επιλογής Προτάσεων .....	47
4.2 Σύγκριση των Αλγορίθμων Επιλογής Προτάσεων.....	55
<b>5. Τελικά συμπεράσματα.....</b>	<b>67</b>
5.1 Επίλογος.....	67
<b>Βιβλιογραφία .....</b>	<b>69</b>

---

# Εισαγωγή

Με τον όρο Σύνθεση Φωνής εννοούμε την αυτοματοποιημένη μετατροπή της φυσικής γλώσσας από τη μορφή κειμένου στη μορφή προφορικού λόγου. Ουσιαστικά η μηχανή προσπαθεί να μιμηθεί αυτό που ο άνθρωπος κάνει όταν διαβάζει δυνατά ένα γραπτό κείμενο. Υπάρχουν πάρα πολλές εφαρμογές όπου ένα αυτοματοποιημένο σύστημα σύνθεσης φωνής μπορεί να χρησιμοποιηθεί, όπως στην υποβοηθούμενη ομιλία ανθρώπων με γλωσσική αναπηρία, στην επικοινωνία ανθρώπου – υπολογιστή, στην τηλεφωνία ή σε περιπτώσεις όπου το οπτικό ανθρώπινο σύστημα είναι απασχολημένο.

Οι επιστήμονες προσπάθησαν να προσεγγίσουν το πρόβλημα της σύνθεσης φωνής, αλλά δυστυχώς, χωρίς ιδιαίτερη επιτυχία μέχρι τα μέσα της δεκαετίας του 80. Πιο συγκεκριμένα, τα αποτελέσματα στον τομέα της σύνθεσης φωνής ήταν κάτι περισσότερο από φτωχά. Υπήρχαν αρκετοί παράγοντες που περιόριζαν την έρευνα των επιστημόνων όπως η έλλειψη μεγάλων αποθηκευτικών μονάδων (στατικής και δυναμικής μνήμης) καθώς επίσης και η ανεπάρκεια επεξεργαστικής δύναμης και ισχύος των επεξεργαστών της εποχής αυτής, με αποτέλεσμα η πρόοδος και η έρευνα να παραμείνουν στάσιμες. Παράλληλα, οι επιστήμονες ως τότε, προσπαθούσαν να συνθέσουν προφορικό λόγο χρησιμοποιώντας βασικά σήματα και μοντελοποιώντας το ανθρώπινο σύστημα δημιουργίας φωνής, κάτι ιδιαίτερα δύσκολο και περίπλοκο αν σκεφτεί κανείς τα προβλήματα που πρέπει να διευθετηθούν λόγω της ύπαρξης πολλών διαφορετικών φωνών και σημάτων. Όταν αναλογιζόμαστε το τι κάνει ένας άνθρωπος όταν διαβάζει δυνατά, τότε μόνο μπορούμε να κατανοήσουμε πόσο περίπλοκη και

δύσκολη είναι η αυτοματοποιημένη σύνθεση φωνής. Έτσι, για να απαγγελθεί σωστά μία λέξη ή μία πρόταση, ο αναγνώστης δεν αρκεί μόνο να την «διαβάσει», αλλά θα πρέπει να χρησιμοποιήσει τον σωστό τόνο, το κατάλληλο ύφος, ενέργεια και ρυθμό της φωνής του, ώστε το αποτέλεσμα να είναι φυσικό.

Σημαντική εξέλιξη και πρόοδος στον τομέα της σύνθεσης φωνής έγινε όταν οι επιστήμονες συνειδητοποίησαν ότι εφόσον δεν είμαστε σε θέση να αναπαράγουμε επιτυχώς τον ανθρώπινο τρόπο σύνθεσης φωνής είναι πιο εύκολο να προσπαθήσουμε να τον αντιγράψουμε. Έτσι, αντί να προσπαθούμε να παράγουμε φωνή μέσω της τροποποίησης και επεξεργασίας βασικών σημάτων με μεθόδους που προσπαθούν να εξομοιώσουν την φυσιολογία του ανθρώπινου συστήματος παραγωγής φωνής, θα ήταν πιο εύκολο να χρησιμοποιήσουμε ήδη ηχογραφημένα πραγματικά δείγματα από ήχους που θέλουμε να παράγουμε, πετυχαίνοντας έτσι παρόμοιο αποτέλεσμα. Αυτό φυσικά αποτελεί μια απλοποιημένη και αφαιρετική προσέγγιση της διαδικασίας σύνθεσης φωνής, αλλά όμως ακόμα αντιπροσωπεύει την βασική ιδέα για την πρόοδο στον τομέα της σύνθεσης φωνής τα τελευταία χρόνια.

Στις μέρες μας, χάρις την εξέλιξη της τεχνολογίας στον τομέα των επεξεργασιών, των αποθηκευτικών μονάδων και των αλγορίθμων στον τομέα της σύνθεσης φωνής, έχουμε κατορθώσει να συνθέσουμε φωνή αρκετά ικανοποιητικής ποιότητας, ώστε να μπορούμε να την χρησιμοποιήσουμε σε τηλεφωνικά κέντρα και σε πολλές άλλες εφαρμογές όπου μέχρι και σήμερα λειτουργούσαν με ανθρώπινους ομιλητές. Αυτό φυσικά δεν συνεπάγεται ότι η έρευνα και η πρόοδος τελείωσαν. Η ποιότητα της φωνής που μπορούμε να συνθέσουμε μπορεί να είναι ικανοποιητική αλλά απέχει πολύ ακόμα από τον ανθρώπινο λόγο, εφόσον καμία βάση δεδομένων δεν μπορεί να παρέχει πλήρη κάλυψη όλων των δυνατών γλωσσικών συνδυασμών.

## 1.1 Σκοπός της Εργασίας Αυτής

Η εργασία αυτή προσπαθεί να πετύχει το στόχο της βέλτιστης επιλογής κειμένων και συγκεκριμένα προτάσεων από ένα σύνολο αρχείων, οι οποίες θα χρησιμοποιηθούν για την δημιουργία της βάσης δεδομένων των τμημάτων φωνής (Speech Segment Database) για ένα σύστημα σύνθεσης φωνής. Με την υλοποίηση εξελιγμένων αλγορίθμων και τεχνικών προσπαθούμε να βελτιώσουμε την αρχιτεκτονική ενός συστήματος σύνθεσης φωνής επιτυγχάνοντας έτσι καλύτερα αποτελέσματα. Το σύστημα μας δέχεται ως είσοδο ένα κείμενο ή ένα σύνολο κειμένων, και πραγματοποιώντας κανονικοποίηση και επεξεργασία αυτού καταφέρνει να επιλέξει έναν ικανό αριθμό προτάσεων όπου από τις οποίες θα δημιουργήσουμε το σώμα λόγου (Speech Corpus). Βήμα καθοριστικό σε ένα σύστημα σύνθεσης φωνής αφού από το σώμα λόγου, έπειτα από τμηματοποίηση (segmentation) αυτού, μπορούμε να απομονώσουμε τις βασικές μονάδες για την σύνθεση φωνής, οι οποίες είναι τμήματα λέξεων, φωνήματα ή συνδυασμοί φωνημάτων και να παράγουμε έτσι την βάση δεδομένων μας (Speech Segment Database).

Παράλληλα, εφόσον έχουμε υλοποιήσει διάφορους αλγορίθμους για την επιλογή προτάσεων, παραθέτουμε τα αποτελέσματα από τη χρήση του κάθε αλγορίθμου ώστε να μπορεί ο ελάχιστος χρήστης να επιλέξει τον κατάλληλο αλγόριθμο επιλογής προτάσεων για κάθε εφαρμογή. Τέλος, παραθέτουμε στην αρχή της αναφοράς μια σύντομη άλλα περιεκτική περιγραφή της δομής και της λειτουργίας ενός συστήματος σύνθεσης φωνής, με στόχο να βοηθήσουμε τον αναγνώστη να κατανοήσει καλύτερα τα βήματα ενός τέτοιου συστήματος και να αντιληφθεί την σημασία της βέλτιστης επιλογής κειμένων.

## 1.2 Περίγραμμα

Η αναφορά αυτή είναι δομημένη ως εξής:

Στο κεφάλαιο 2, περιγράφουμε αναλυτικά τη δομή και την λειτουργία ενός ολοκληρωμένου συστήματος σύνθεσης φωνής, το οποίο δομείται από δύο κύρια υποσυστήματα: το υποσύστημα επεξεργασίας φυσικής γλώσσας (Natural Language Subsystem) και το υποσύστημα επεξεργασίας σημάτων φωνής (Speech Signal Processing Subsystem). Στη συνέχεια, αναλύουμε τα διάφορα τμήματα του κάθε υποσυστήματος εξερευνώντας τη δομή και τη λειτουργία του καθενός.

Το κεφάλαιο 3 περιγράφει την καρδιά του προβλήματος, τους κύριους στόχους και τις προτεινόμενες λύσεις. Αναφερόμαστε, δηλαδή, στις μεθόδους που υλοποιήσαμε ώστε να πετύχουμε βέλτιστη επιλογή κειμένων και συγκεκριμένα προτάσεων από ένα σύνολο αρχείων, οι οποίες θα χρησιμοποιηθούν για την δημιουργία της βάσης δεδομένων των τμημάτων φωνής (Speech Segment Database) για ένα σύστημα σύνθεσης φωνής. Εδώ περιγράφουμε αναλυτικά τα software modules που υλοποιήσαμε να για επιτευχθεί ο στόχος μας, παραθέτοντας παράλληλα, τις τυχόν δυσκολίες που συναντήσαμε και το πώς αυτές αντιμετωπίστηκαν.

Στο κεφάλαιο 4 αναλύουμε τα αποτελέσματα της κάθε μεθόδου επιλογής προτάσεων, συγκρίνοντας τον κάθε αλγόριθμο επιλογής και αξιολογώντας έτσι το όλο σύστημα μέσω της παρατήρησης και της εξέτασης διάφορων στατιστικών στοιχείων, γραφικών παραστάσεων και κατανομών.

Τέλος, στο κεφάλαιο 5 αναφέρονται τελικές παρατηρήσεις και συμπεράσματα από την όλη διαδικασία επιλογής προτάσεων και κλείνουμε την αναφορά με την εξέταση μελλοντικών δυνατοτήτων και προοπτικών στον τομέα της σύνθεσης φωνής.

---

# Συστήματα Σύνθεσης Φωνής

## 2.1 Εισαγωγή στα Συστήματα Σύνθεσης Φωνής

Η διαδικασία σύνθεσης φωνής δεν αποτελεί μια απλή διαδικασία. Αρχικά κανείς να συλλογιστεί το τι κάνει ένας άνθρωπος όταν διαβάζει δυνατά ένα γραπτό κείμενο, ώστε να γίνει σαφές το πόσο δύσκολη και σύνθετη είναι η διαδικασία της αυτοματοποιημένης σύνθεσης φωνής. Η σωστή προφορά ενός κειμένου από τον αναγνώστη είναι μόνο ένα μέρος του προβλήματος, έτσι για να ακούγεται φυσικά και για να μπορεί να κατανοήσει ο ακροατής αυτά που ακούει, ο αναγνώστης έχει ένα περίπλοκο και δύσκολο ρόλο. Δεν αρκεί μόνο να «διαβάσει» το κείμενο αλλά θα πρέπει να δώσει έμφαση σε κάποιες λέξεις, τονίζοντας τις, ενώ παράλληλα να τονίσει λιγότερο κάποιες άλλες, θα πρέπει να χωρίζει τις προτάσεις σε λογικές νοηματικές φράσεις, να ελέγχει την προσωδία της φωνής του, να προσφέρει μία λέξη με μεγαλύτερη διάρκεια, όταν αυτή εμφανίζεται σε ορισμένες θέσεις σε μία πρόταση από ότι σε κάποιες άλλες.

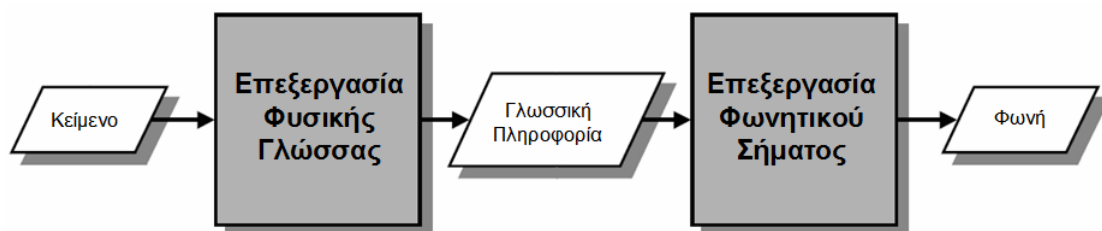
Αυτό που κάνει την σύνθεση φωνής μία επίπονη διαδικασία για μία μηχανή είναι το γεγονός ότι το γραπτό κείμενο δεν περιέχει πολλές από τις απαραίτητες πληροφορίες που είναι σημαντικές στον προφορικό λόγο. Έτσι, παρόλο που ένα γραπτό κείμενο προσδιορίζει απόλυτα τις λέξεις που το συνθέτουν, δίνει ελάχιστες πληροφορίες για την σωστή τους προφορά, με αποτέλεσμα να είναι δύσκολο στην μηχανή να αποφασίσει ποιες λέξεις θα πρέπει να τονιστούν περισσότερο από κάποιες

άλλες, πόσο άλλωστε να αποφασίσει ποια θα είναι η προσωδία της φωνής ή η διάρκεια προφοράς της εκάστοτε λέξης.

Στην συνέχεια θα περιγράψουμε αναλυτικά την δομή ενός συστήματος σύνθεσης φωνής, περιγράφοντας την λειτουργία και το ρόλο κάθε υποσυστήματος και τμήματος αυτού.

## 2.2 Δομή ενός Συστήματος Σύνθεσης Φωνής

Παρόλο που έχουν κατασκευαστεί αρκετά συστήματα σύνθεσης φωνής, όπου το καθένα προσπαθεί να προσεγγίσει το πρόβλημα τις σύνθεσης φωνής με διαφορετικό τρόπο, όλα έχουν κατά βάση μια κοινή σύνθεση και δομή. Θα μπορούσαμε να πούμε ότι το πρόβλημα της σύνθεσης φωνής διαιρείται σε δύο υπό-πρόβληματα. Το πρώτο από αυτά είναι η μετατροπή του κειμένου σε μια μορφή γλωσσικής πληροφορίας η οποία περιέχει πληροφορίες για τα φωνήματα – ήχους που θα παραχθούν, για την θέση και την διάρκεια κάθε παύσης και για την καμπύλη της θεμελιώδους συχνότητας (F0 contour) που θα χρησιμοποιήσουμε. Το δεύτερο, που ουσιαστικά αφορά το κομμάτι σύνθεσης του ήχου, προσπαθεί να δεχτεί τις πληροφορίες αυτές και να τις μετατρέψει σε ήχο.



*Εικόνα 2.1 Τα κύρια υποσυστήματα ενός συστήματος σύνθεσης φωνής.*

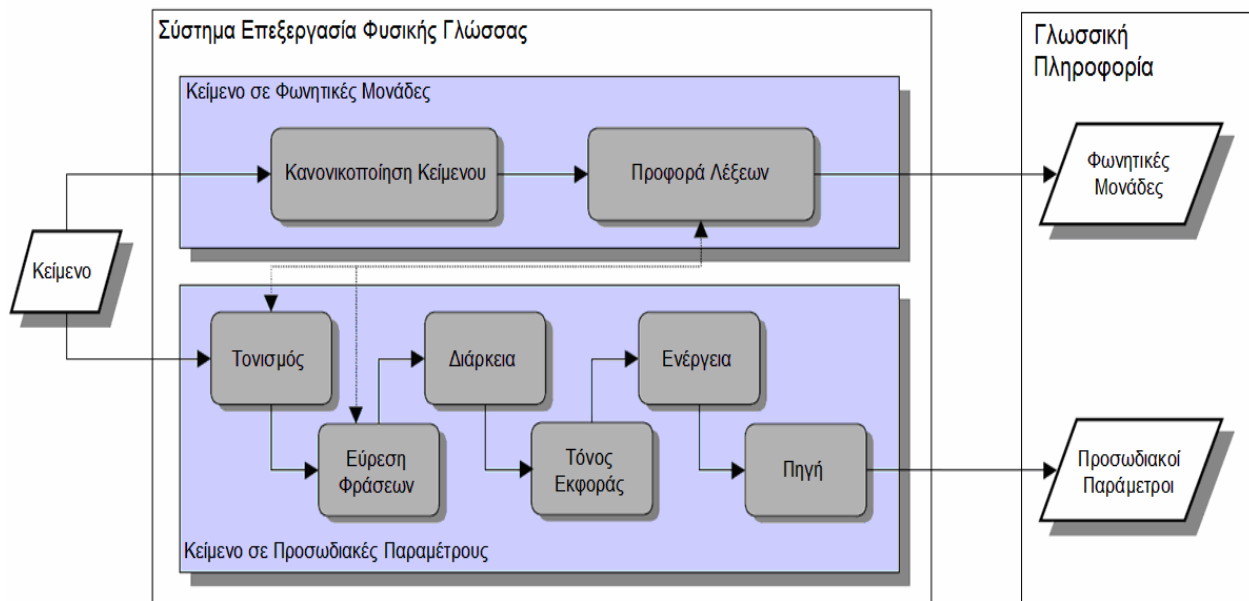
Ένα τυπικό σύστημα σύνθεσης φωνής μπορεί να διασπαστεί σε δύο υπο-συστήματα, όπως φαίνεται στην εικόνα 2.1. Το πρώτο υποσύστημα είναι μία μονάδα επεξεργασίας φυσικής γλώσσας (Natural Language Processing Module) που αναλαμβάνει να μετατρέψει το κείμενο σε μια μορφή γλωσσικής πληροφορίας η



οποία περιέχει φωνητική (phonetic) και προσωδιακή (prosodic) πληροφορία. Οι φωνητικές μονάδες (phonetic units) καθορίζουν *τι* φωνήματα – ήχοι χρειάζεται να παραχθούν, ενώ οι προσωδιακοί παράμετροι καθορίζουν το *πώς* αυτοί πρέπει να παραχθούν. Το δεύτερο υποσύστημα είναι μια μονάδα επεξεργασίας φωνητικού σήματος (Speech Signal Processing Module) που είναι υπεύθυνη για την παραγωγή των ηχητικών κυματομορφών και δέχεται ως είσοδο την γλωσσική πληροφορία από το πρώτο υποσύστημα.

Στις επόμενες δύο ενότητες θα περιγράψουμε αναλυτικά τα δύο αυτά υποσυστήματα με σκοπό να κατανοήσουμε την δομή και την λειτουργία του καθενός παρουσιάζοντας, παράλληλα, τα προβλήματα που διαχειρίζεται το κάθε ένα και τους διάφορους τρόπους αντιμετώπισης τους.

### 2.2.1 Υποσύστημα Επεξεργασίας Φυσικής Γλώσσας



*Εικόνα 2.2 Τα τμήματα του Συστήματος Επεξεργασίας Φυσικής Γλώσσας*

Όπως φαίνεται στην εικόνα 2.2, η μετατροπή του κειμένου εισόδου σε γλωσσική πληροφορία μπορεί να χωριστεί σε δύο τμήματα: την μετατροπή του κειμένου σε φωνητικές μονάδες και την μετατροπή του κειμένου σε προσωδιακές

παραμέτρους. Όπως είχαμε αναφέρει και προηγουμένως, οι φωνητικές μονάδες καθορίζουν τι φωνήματα – ήχοι πρέπει να παραχθούν ενώ, οι προσωδιακές παράμετροι καθορίζουν πώς αυτοί θα παραχθούν.

Για να είμαστε σε θέση να μπορούμε να συνθέτουμε φωνή από οποιαδήποτε κείμενο, συμπεριλαμβάνοντας νέες λέξεις και προτάσεις, οι μονάδες σύνθεσης (synthesis units) μας θα πρέπει είναι μικρότερες από λέξεις, εφόσον ο αριθμός των δυνατών λέξεων είναι πρακτικά απεριόριστος. Όπως αντιλαμβανόμαστε, οι μονάδες σύνθεσης θα πρέπει να είναι μικρές μονάδες όπως φωνήματα ή συνδυασμοί φωνημάτων ώστε να είμαστε ευέλικτοι στην σύνθεση κάθε νέας λέξης ή πρότασης. Ο ρόλος των φωνητικών μονάδων είναι πολύ σημαντικός στην διαδικασία σύνθεσης φωνής, εφόσον αυτές αποτελούν το ενδιάμεσο στάδιο ανάμεσα στις λέξεις και τελικές μονάδες σύνθεσης που θα χρησιμοποιηθούν στην δημιουργία φωνής. Η τελική μετατροπή των φωνητικών μονάδων σε μονάδες σύνθεσης γίνεται στο στάδιο της επεξεργασίας των ηχητικών σημάτων.

Με τον όρο προσωδία αναφερόμαστε στα χαρακτηριστικά της φωνής τα οποία είναι ο ρυθμός, ο τόνος και η ένταση της ή σε φυσική αντιστοιχία η διάρκεια, η ενέργεια και η θεμελιώδης συχνότητα της φωνής, παράγοντες που αποτελούν το κλειδί ώστε το αποτέλεσμα της σύνθεσης να ακούγεται φυσικό. Ο στόχος της μετατροπής του κειμένου σε προσωδιακές παραμέτρους είναι να δημιουργήσουμε μια «χρονική ροή» με τις παραμέτρους αυτούς.

Στη συνέχεια θα εξετάσουμε με περισσότερη λεπτομέρεια την μετατροπή του κειμένου σε φωνητικές μονάδες και την μετατροπή του κειμένου σε προσωδιακές παραμέτρους.

## *Μετατροπή Κειμένου σε Φωνητικές Μονάδες*

### *Κανονικοποίηση Κειμένου*

Στόχος της κανονικοποίησης του κειμένου είναι να μετασχηματιστεί το απλό κείμενο σε μια μορφή που να μπορεί να την επεξεργαστεί το υπόλοιπο σύστημα. Αυτό περιλαμβάνει τον χωρισμό του κειμένου σε προτάσεις, την εύρεση των

συντομογραφιών (όπως, σελ.), των αριθμών (όπως, ημεροπηνίες, χρονολογίες, τηλεφωνικοί αριθμοί και χρηματικά ποσά) και πιθανόν, μια μικρή γραμματική ανάλυση του κειμένου όπως, τον χαρακτηρισμό των λέξεων με διάφορα χαρακτηριστικά του λόγου, κάτι που θα διευκολύνει αργότερα στην εύρεση της προφοράς κάθε λέξης.

Πολλές προσεγγίσεις στην κανονικοποίηση του κειμένου βασίζονται σε κανόνες και σε εμπειρικές μεθόδους. Για παράδειγμα, τα σημεία στίξης, τα κεφαλαία γράμματα και το κενό μπορούν να χρησιμοποιηθούν ως σημάδια για την ανίχνευση προτάσεων και τον ορίων μιας λέξης σε αρκετές γλώσσες. Παράλληλα, μπορούν να χρησιμοποιηθούν πιο εξελιγμένοι αλγόριθμοι και εμπειρικοί κανόνες για την εύρεση των ορίων μιας λέξης σε γλώσσες όπως τα Κινέζικα, όπου οι λέξεις δεν χωρίζονται με κενό.

## Προφορά Λέξεων

Μετά την κανονικοποίηση του κειμένου, το επόμενο βήμα είναι η εύρεση και ο καθορισμός της προφοράς κάθε λέξης. Οι αποτελεσματικότεροι μέθοδοι εύρεσης των λεκτικών προφορών βασίζονται σε απλούς κανόνες. Η πιο απλή προσέγγιση είναι η αντιστοίχιση των γραμμάτων μίας λέξης σε ήχους, μετατρέποντας έτσι γραμματικές ακολουθίες σε φωνητικές ακολουθίες. Η προσέγγιση αυτή ισχύει για γλώσσες όπως τα Ελληνικά και τα Ισπανικά, όπου υπάρχει μια ένα-προς-ένα αντιστοιχία ορθογραφικής και φωνητικής γραφής, αλλά όχι σε γλώσσες όπως π.χ. τα Αγγλικά. Μια άλλη πιο σύνθετη προσέγγιση βασίζεται στην μορφολογική ανάλυση. Κάθε λέξη διαιρείται σε μορφήματα, μικρότερες γλωσσικές μονάδες, όπως συλλαβές, ρίζες, και καταλήξεις. Έπειτα, παράγουμε την προφορά κάθε μορφήματος αντιστοιχίζοντας τις μικρότερες αυτές γλωσσικές μονάδες σε ήχους, με χρήση απλών μεθόδων αντιστοίχισης και μορφικών λεξικών. Στο τέλος, ενώνουμε τα επιμέρους κομμάτια ώστε να παράγουμε την προφορά κάθε λέξης. Τέλος, αξίζει να σημειωθεί ότι υπάρχουν και προσεγγίσεις όπου η προφορά μιας λέξης προκύπτει από την ορθογραφική γραφή μέσω στατιστικών μεθόδων.

## Μετατροπή Κειμένου σε Προσωδιακές Παραμέτρους

### Τονισμός

Στον προφορικό λόγο κάποιες λέξεις τονίζονται ενώ κάποιες άλλες όχι. Το αν μια λέξη τονίζεται ή όχι καθορίζεται από την γραμματική κατηγορία της λέξης, έτσι λέξεις όπως ουσιαστικά, επίθετα και ρήματα συνήθως τονίζονται ενώ προθέσεις, βοηθητικά ρήματα δεν τονίζονται. Προβλήματα συνήθως παρουσιάζονται σε περίπλοκες φράσεις ή καταστάσεις όπου απαιτείται έμφαση ή αντίθεση.

Διάφοροι τρόποι υπάρχουν ώστε να βρεθεί ο τονισμός μιας λέξης. Οι περισσότεροι βασίζονται στην εύρεση της κατηγορίας κάθε λέξης και έπειτα χρησιμοποιώντας τη γνώση αυτή και πληροφορίες από γειτονικά συμφραζόμενα παράγεται ο τονισμός χρησιμοποιώντας ένα σύνολο κανόνων. Οι κανόνες αυτοί είναι συνήθως εμπειρικοί και κατασκευάζονται από την ανάλυση μεγάλου όγκου από δεδομένα.

### Εύρεση Φράσεων

Μια πρόταση συνήθως διαιρείται σε μικρότερες φράσεις. Η εύρεση των φράσεων αυτών είναι σημαντική στον καθορισμό των προσωδιακών παραμέτρων εφόσον στα όρια των φράσεων αυτών υπάρχουν παύσεις, γεγονός που σημαίνει ότι θα πρέπει να αλλάξει η ενέργεια και ο τόνος του ομιλητή. Το έργο της αυτοματοποιημένης εύρεσης των φράσεων αυτών βασίζεται στην ανάλυση του κειμένου.

Η πιο απλή προσέγγιση είναι να εκμεταλλευτούμε την ύπαρξη των σημείων στίξης, όπως κόμματα, άνω τελείες και παρενθέσεις ώστε να χωρίσουμε μία πρόταση σε φράσεις. Προβλήματα παρουσιάζονται σε προτάσεις αποτελούμενες από γραμματοσειρές χωρίς σημεία στίξης. Στις περιπτώσεις αυτές, χρησιμοποιούνται άλλες τεχνικές, όπως η χρησιμοποίηση ενός συνόλου από λέξεις που συνήθως βρίσκονται στα όρια των φράσεων για να χωρίσουμε επιτυχώς την πρόταση σε φράσεις.

## Διάρκεια

Εκτός από το γεγονός ότι θα πρέπει να γνωρίζουμε ποια φωνήματα θα πρέπει να παράγουμε είναι απαραίτητο να ορίζουμε τη διάρκεια προφοράς του καθενός. Υπάρχουν πολλοί παράγοντες που επηρεάζουν την διάρκεια ενός φωνήματος, όπως το ίδιο το φώνημα, τα χαρακτηριστικά των γειτονικών του φωνημάτων, η θέση του στην φράση και τα χαρακτηριστικά ομιλίας του ομιλητή, δηλαδή ο ρυθμός ομιλίας του και η διάλεκτος του.

Έτσι, γνωρίζοντας τους παραπάνω παράγοντες μπορούμε με την χρήση εμπειρικών κανόνων ανάλυσης του κειμένου να καθορίσουμε την διάρκεια προφοράς κάθε φωνήματος.

## Τόνος εκφοράς

Στόχος αυτού του τμήματος είναι να παράγει μια καμπύλη θεμελιώδους συχνότητας ( $F_0$  contour) για την πρόταση που θα συντεθεί. Ο τόνος εκφοράς (intonation) είναι ίσως ο σημαντικότερος προσωδιακός παράγοντας, εφόσον επηρεάζει σημαντικά την ποιότητα του ήχου και για τον λόγο αυτό χρησιμοποιούνται πληροφορίες από όλα τα προηγούμενα τμήματα.

Σε πολλές υλοποιήσεις του τμήματος αυτού χρησιμοποιείται μια ενδιάμεση προσωδιακή δομή (π.χ. ένα μοντέλο τρόπου εκφοράς). Το μοντέλο αυτό χρησιμοποιείται για να προβλέπει πότε η  $F_0$  ανεβαίνει ή πέφτει και σε τι επίπεδο αυτή φτάνει ανάλογα με την συντακτική δομή και τη θέση στην πρόταση. Έτσι, με τον τρόπο αυτό δημιουργούνται ζεύγη  $F_0$  - χρόνου τα οποία μετά εξομαλύνονται ώστε να πάρουμε το τελικό αποτέλεσμα. Τέλος, αξίζει να αναφέρουμε ότι στα σύγχρονα συστήματα σύνθεσης φωνής χρησιμοποιούνται πιο εξελιγμένες μέθοδοι για την πρόβλεψη της  $F_0$ , όπως νευρωνικά δίκτυα, με αποτέλεσμα να μην απαιτείται η ύπαρξη ενός ενδιάμεσου μοντέλου.

## Ενέργεια

Εκτός από τη θεμελιώδη συχνότητα, οι φωνητικές οντότητες έχουν και ένα επιπλέον χαρακτηριστικό, την ενέργεια. Έτσι, κάποια φωνήματα προφέρονται πιο έντονα από κάποια άλλα και επίσης το τέλος μιας λέξης ή φράσης είναι πιο αδύναμο από ότι η αρχή της. Παράλληλα, έχει παρατηρηθεί ότι η ενέργεια επηρεάζεται άμεσα από την  $F_0$  και πιο συγκεκριμένα, η ενέργεια μιας φωνητικής οντότητας μπορεί να υπολογισθεί αν συνδυάσουμε την κανονική ενέργεια που αντιπροσωπεύει κάθε φώνημα με την  $F_0$  του. Ο λόγος για αυτό το αποτέλεσμα είναι ότι η ένταση μίας ηχητικής μονάδος αυξάνεται όσο ανεβαίνει η  $F_0$ , με συνέπεια οι διακυμάνσεις στην ένταση να μπορούν να προβλεφθούν από τις αλλαγές της  $F_0$ .

## Πηγή Ομιλίας

Εκτός από τις συνηθισμένες προσωδιακές παραμέτρους, όπως τη θεμελιώδη συχνότητα, τη διάρκεια και την ενέργεια, υπάρχουν και επιπλέον παράγοντες που σχετίζονται με τα χαρακτηριστικά της πηγής ομιλίας (χαρακτηριστικά του ομιλητή) και παίζουν σημαντικό ρόλο στον καθορισμό της ποιότητας της φωνής, όπως λαρυγγικά χαρακτηριστικά του ομιλητή και ο θόρυβος εισπνοής – εκπνοής. Έτσι, η φιλοδοξία μας για καλύτερα αποτελέσματα στη σύνθεση φωνής, μας ωθεί στην κατασκευή πιο ρεαλιστικών μοντέλων του φωνητικού συστήματος παρατηρώντας και καταγράφοντας τα ιδιαίτερα χαρακτηριστικά του ομιλητή.

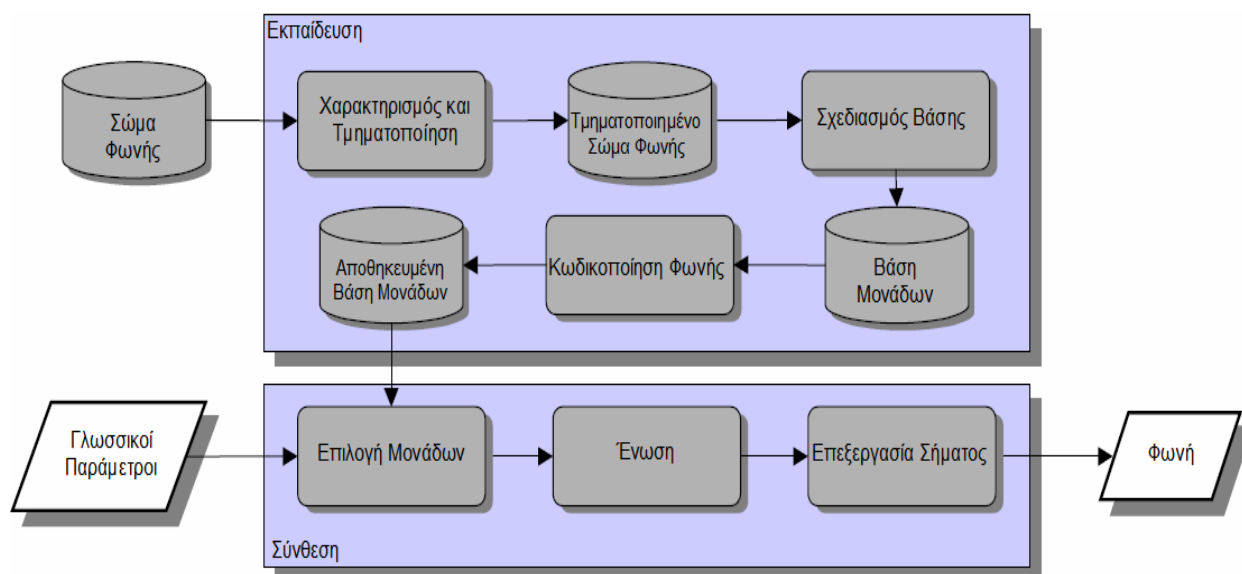
### 2.2.2 Υποσύστημα Επεξεργασίας Φωνητικού Σήματος

Το υποσύστημα επεξεργασίας φωνητικού σήματος δέχεται ως είσοδο την γλωσσική πληροφορία (φωνητική και προσωδιακή) από το υποσύστημα επεξεργασίας φυσικής γλώσσας, συνθέτει την φωνή, και βγάζει στην έξοδο τις τελικές ηχητικές κυματομορφές. Οι προσεγγίσεις της παραπάνω διαδικασίας κατηγοριοποιούνται σε δύο ομάδες: αυτές που προσπαθούν να μοντελοποιήσουν το σύστημα παραγωγής φωνής (Articulatory Synthesis) και σε αυτές που προσπαθούν να κατασκευάσουν το

σήμα φωνής (Concatenative Synthesis). Τα περισσότερα σύγχρονα συστήματα σύνθεσης φωνής ανάγονται στην δεύτερη κατηγορία και για το λόγο αυτό στην συνέχεια θα περιγράψουμε την δομή και την λειτουργία ενός concatenative συστήματος σύνθεσης φωνής.

### Concatenative Σύνθεση Φωνής

Στην concatenative σύνθεση, η φωνή παράγεται από την ανάκτηση κατάλληλων μονάδων αποθηκευμένου φυσικού λόγου, την ένωση τους και έπειτα την επεξεργασία του αποτελέσματος ώστε αυτό να εξομαλυνθεί (smoothing) και μοιάσει με τον πραγματικό λόγο.



Εικόνα 2.3 Τα τμήματα ενός Concatenative συνθέτη φωνής

Παρόλο που υπάρχουν αρκετές και διαφορετικές υλοποιήσεις concatenative συστημάτων, τα περισσότερα από αυτά έχουν μια κοινή δομή. Όπως φαίνεται και στην εικόνα 2.3, υπάρχουν δύο κύρια υποσυστήματα: ένα για την εκπαίδευση και ένα για την σύνθεση. Στην εκπαίδευση, ένα σύνολο από μονάδες σύνθεσης δημιουργείται από ένα σώμα προφορικού λόγου (Speech Corpus). Στην σύνθεση, οι φωνητικές και προσωδιακές πληροφορίες, οι οποίες αποτελούν την είσοδο του συστήματος,

καθορίζουν ποια λεκτική μονάδα θα συντεθεί. Έπειτα, οι κατάλληλες μονάδες επιλέγονται από την βάση δεδομένων μας, ενώνονται και επεξεργάζεται το τελικό αποτέλεσμα, ώστε να παραχθεί η τελική κυματομορφή. Τα υποσυστήματα εκπαίδευσης και σύνθεσης αποτελούνται από μικρότερα τμήματα όπως φαίνεται και στην εικόνα 2.3. Στην συνέχεια, θα περιγράψουμε τα τμήματα αυτά.

**Σώμα Φωνής** Το σώμα φωνής (Speech Corpus) αποτελεί την πηγή εύρεσης των

μονάδων σύνθεσης και συνήθως δημιουργείται από την ηχογράφηση κειμένου από έναν μόνο ομιλητή. Το σώμα αυτό μπορεί να είναι ένα προ-ηχογραφημένο τμήμα φωνής το οποίο δημιουργήθηκε για τις ανάγκες κάποιας άλλης εφαρμογής, όπως για ένα σύστημα αναγνώρισης φωνής ή να δημιουργήθηκε για τις ανάγκες της δικής μας εφαρμογής. Το σώμα αυτό δεν θα πρέπει να είναι τεράστιο, άλλα όσο το δυνατόν μικρότερο, ώστε να μπορεί να ηχογραφηθεί από έναν και μόνο ομιλητή χωρίς αυτός να κουραστεί, προκαλώντας έτσι αλλοίωση του αποτελέσματος. Τέλος, αξίζει να επισημάνουμε ότι για να εξασφαλίσουμε την μέγιστη ποιότητα ήχου του σώματος αυτού, η ηχογράφηση συνήθως πραγματοποιείται σε σύγχρονα δωμάτια ηχογραφήσεως με κορυφαίας ποιότητας μικρόφωνα και λαριγγογράφους.

**Τμηματοποίηση και Χαρακτηρισμός της Φωνής** Το πρώτο στάδιο στην

επεξεργασία είναι η τμηματοποίηση και ο χαρακτηρισμός (Labeling) του σώματος φωνής, έτσι ώστε να γνωρίζουμε ποιοι ήχοι υπάρχουν και πού αυτοί υπάρχουν. Παράλληλα, μπορούμε να συσχετίσουμε τα τμήματα φωνής και να τα εμπλουτίσουμε με διάφορες πληροφορίες, όπως προσωδιακά χαρακτηριστικά. Η διαδικασία της τμηματοποίησης και του χαρακτηρισμού του σώματος φωνής μπορεί να γίνει είτε χειροκίνητα είτε αυτοματοποιημένα, χρησιμοποιώντας τεχνικές αναγνώρισης φωνής.

**Σχεδιασμός της Βάσης των Μονάδων** Σκοπός της σχεδίασης της βάσης των

μονάδων είναι να συγκεντρώσουμε ένα σύνολο από τμήματα φωνής (φωνήματα, δίφωνα, συνδυασμοί φωνημάτων) που να μπορούν να συνθέσουν



κάθε είδους λέξη και πρόταση. Έτσι, επιλέγονται ένα ή και περισσότερα αντίγραφα κάθε μονάδος ώστε να πετύχουμε μέγιστη φωνητική κάλυψη. Οι μονάδες που θα επιλεγούν μπορεί να έχουν όλες το ίδιο ή διαφορετικό μέγεθος ή μία με την άλλη, ανάλογα με την αρχιτεκτονική του συστήματος. Η σχεδίαση της βάσης αυτής αποτελεί τον πιο σημαντικό παράγοντα σε ένα σύστημα σύνθεσης φωνής αφού επηρεάζει κάθε άλλο τμήμα.

### **Κωδικοποίηση φωνής και Αποθήκευση**

Στο στάδιο αυτό αποθηκεύουμε την βάση με τις μονάδες μας και παράλληλα καταχωρούμε σε ευρετήριο τις μονάδες αυτές ώστε να μπορούμε να έχουμε εύκολη και γρήγορη πρόσβαση κατά την διαδικασία σύνθεσης φωνής. Πολλά συστήματα, αναλαμβάνουν να αποθηκεύσουν τις μονάδες αυτές ως απλές κυματομορφές και άλλα πραγματοποιούν ένα είδος φωνητικής ανάλυσης (LPC analysis) στα τμήματα αυτά ενώ παράλληλα, αποθηκεύουν τα αποτελέσματα της ανάλυσης με τους παραμέτρους τους.

### **Επιλογή Μονάδων**

Κατά τη διάρκεια της σύνθεσης, θα πρέπει αυτόματα να επιλεχθούν οι κατάλληλες μονάδες από τη βάση δεδομένων μας. Τα κριτήρια επιλογής είναι φωνητικά και προσωδιακά και στόχος μας είναι να επιλεγούν οι καλύτερες μονάδες ώστε να ταιριάζουν φωνητικά και προσωδιακά με τους ήχους που επιθυμούμε να παράγουμε.

### **Ένωση Μονάδων**

Μετά την επιλογή των κατάλληλων μονάδων, το επόμενο βήμα είναι η ένωση των επιμέρους τμημάτων σε μία ενιαία ροή. Στόχος μας είναι να κάνουμε την μετάβαση από την μία μονάδα στην άλλη όσο δυνατόν πιο ομαλή, ελαχιστοποιώντας ασυνέχειες και όρια. Αυτό συνήθως επιτυγχάνεται με τεχνικές επεξεργασίας σήματος και με την κατάλληλη επιλογή μονάδων, ώστε αυτές να μπορούν να ενωθούν αρμονικά μεταξύ τους.

**Προσωδιακή Επεξεργασία Σήματος** Εκτός από την σωστή επιλογή των μονάδων,

θα πρέπει να έχουμε εξασφαλίσει και τις κατάλληλες προσωδιακές παραμέτρους, ώστε το αποτέλεσμα της σύνθεσης να ακούγεται φυσικό. Για να πετύχουν την κατάλληλη προσωδία, πολλά συστήματα εκτελούν ένα είδος επεξεργασίας σήματος, μεταβάλλοντας έτσι την διάρκεια, τη θεμελιώδη συχνότητα, την ενέργεια και κάποια κύρια χαρακτηριστικά του λόγου. Αξίζει να επισημάνουμε ότι η επεξεργασία αυτή μπορεί να μειωθεί εάν οι μονάδες που επιλέγονται, στο στάδιο της επιλογής τμημάτων, «ταιριάζουν» μεταξύ τους.

# Βέλτιστη Επιλογή Κειμένων

Τα σημερινά συστήματα σύνθεσης φωνής είναι αρκετά ικανοποιητικά στην σύνθεση λόγου αρκετά υψηλής ποιότητας, στις περισσότερες των περιπτώσεων τουλάχιστον. Ωστόσο, η ποιότητα του αποτελέσματος είναι ένα θέμα εξαρτώμενο από πολλές παραμέτρους. Όπως έχουμε προαναφέρει, καθοριστικός παράγοντας για το αποτέλεσμα της σύνθεσης φωνής είναι η σωστή σχεδίαση και η ποιότητα της βάσης δεδομένων μας. Η βάση αυτή αποτελείται από τμήματα φωνής, τα οποία προέρχονται από την τμηματοποίηση ενός σώματος φωνής (Speech Corpus). Η τμηματοποίηση του σώματος αυτού θα μας δώσει τις βασικές μονάδες φωνής, με την χρήση των οποίων θα συνθέσουμε κάθε νέα λέξη ή πρόταση. Για το λόγο αυτό, το σώμα φωνής θα πρέπει να περιέχει όσο δυνατόν μεγαλύτερη γλωσσική πληροφορία (φωνητική και προσωδική), ώστε τα τμήματα φωνής που θα πάρουμε από την τμηματοποίηση αυτού να επαρκούν και να παρέχουν πλήρη φωνητική κάλυψη για την σύνθεση κάθε νέας λέξης. Όπως αντιλαμβανόμαστε, όσο καλύτερης ποιότητας είναι τα τμήματα αυτά και όσους περισσότερους τομείς γλωσσικών τύπων καλύπτουν, τόσο καλύτερο θα είναι και το τελικό αποτέλεσμα. Στο κεφάλαιο αυτό θα περιγράψουμε αναλυτικά των τρόπο επιλογής ενός συνόλου προτάσεων, μέσω χρήσης κατάλληλων αλγορίθμων και προγραμμάτων, η ηχογράφηση των οποίων θα δημιουργήσει το σώμα φωνής που χρειαζόμαστε για ένα σύστημα σύνθεσης φωνής.

### 3.1 Προσέγγιση προβλήματος

Στόχος μας είναι να επιλεγούν όσο δυνατόν λιγότερες προτάσεις από ένα σύνολο γραπτών κειμένων που να πληρούν τα παρακάτω κριτήρια:

1. Να περιέχουν όλα τα φωνήματα της ελληνικής γλώσσας.
2. Να περιέχουν όλα τα νόμιμα δίφωνα της ελληνικής γλώσσας ( συνδυασμούς φωνημάτων - όχι κατ' ανάγκη όλους, καθώς μερικοί συνδυασμοί δεν εμφανίζονται ποτέ στη γλώσσα).
3. Να περιέχουν όσο δυνατόν περισσότερες διαφορετικές λέξεις ή ρίζες λέξεων.
4. Να περιέχουν όσο το δυνατόν ίσο - σταθμισμένο αριθμό διφώνων.
5. Να περιέχουν το πολύ επτά με δέκα λέξεις.

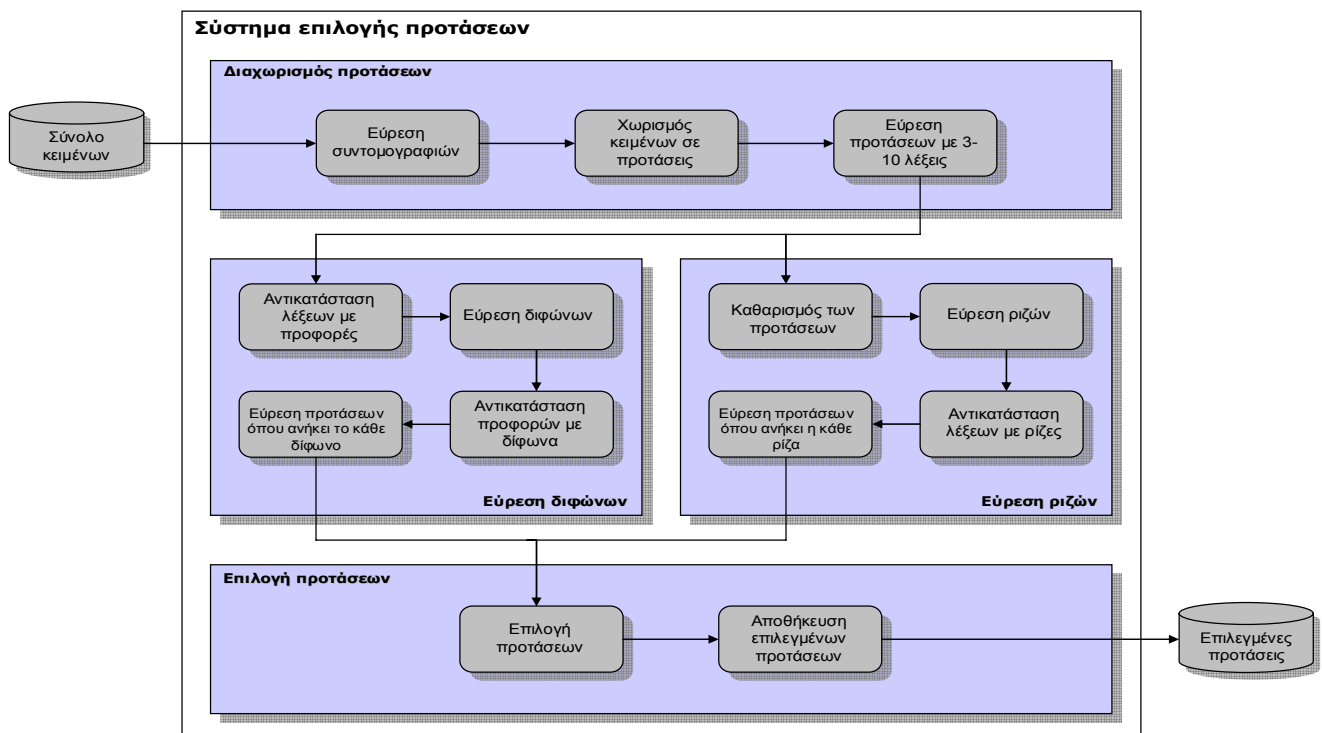
Οι προτάσεις που θα επιλέξουμε πρέπει να περιέχουν όλα τα φωνήματα της ελληνικής γλώσσας, καθώς και όλα τα νόμιμα δίφωνα, ώστε να πετύχουμε όσο το δυνατόν μεγαλύτερη λεκτική κάλυψη δημιουργώντας έτσι, μία πλήρης βάση μονάδων σύνθεσης φωνής, ικανή να συνθέσει κάθε νέα λέξη και πρόταση. Παράλληλα, επιθυμούμε οι προτάσεις αυτές να περιέχουν όσο το δυνατόν περισσότερες διαφορετικές λέξεις ή ρίζες λέξεων, ώστε να έχουμε στην διάθεση μας τον μεγαλύτερο δυνατό αριθμό φωνητικών αλληλουχιών ώστε να αντλήσουμε από αυτές περισσότερη προσωδιακή πληροφορία, καθοριστικός παράγοντας ώστε το αποτέλεσμα να ακούγεται φυσικό. Τέλος, οι προτάσεις μας θα πρέπει να είναι μικρές στο μέγεθος, αποτελούμενες το πολύ από επτά με δέκα λέξεις, ώστε να είναι εύκολη η ηχογράφηση τους χωρίς να προκληθεί κούραση στον ομιλητή, αλλοιώνοντας έτσι το τελικό αποτέλεσμα.

## 3.2 Επίλυση προβλήματος

Για την επίλυση του παραπάνω προβλήματος χρησιμοποιήθηκε μια γλώσσα υψηλού επιπέδου ιδανική για διαχείριση συμβολοσειρών, file I/O και κανονικών εκφράσεων, η Perl.

Η όλη εργασία οργανώθηκε σε software modules, όπου το καθένα εκτελεί ένα διαφορετικό μέρος της επίλυσης του προβλήματος και μπορεί να εκτελεστεί αυτόνομα από τα υπόλοιπα αρχεία. Η εργασία χωρίστηκε σε διαφορετικά τμήματα, ώστε να γίνεται πιο εύκολη η αντιμετώπιση τυχόν δυσλειτουργιών, να είναι πιο κατανοητή στον αναγνώστη η λειτουργία του καθενός μέρους, να μπορεί να αναβαθμιστεί ή να αλλαχθεί κάποιο τμήμα πιο εύκολα και τέλος, ώστε να είναι δυνατόν η συνολική εργασία να τρέξει τμηματικά, κάτι πολύ σημαντικό αν αναλογιστεί κανείς το συνολικό χρόνο που απαιτείται ώστε το παρόν πρόγραμμα να τρέξει.

Η όλη λειτουργία και η σύνδεση των επιμέρους τμημάτων του προγράμματος απεικονίζεται στην εικόνα 3.1.



Εικόνα 3.1 Διάγραμμα ροής και λειτουργικότητας επιμέρους τμημάτων του προγράμματος

Στη συνέχεια θα περιγράψουμε αναλυτικά την λειτουργικότητα του κάθε software module.

### 3.3 Σύστημα Βέλτιστης Επιλογής Προτάσεων

Ως πρώτο βήμα για την επίλυση του προβλήματος ήταν η επιλογή του συνόλου των κειμένων (text corpus) όπου από αυτό θα επιλέγονταν οι προτάσεις που θα πληρούσαν τα κριτήρια που περιγράψαμε στην παράγραφο 3.1.

Το σύνολο των κειμένων αποτελείται από το ηλεκτρονικό αρχείο της εφημερίδος «Ελευθεροτυπία» των εκδόσεων «Τεγόπουλος» από το χρονολογικό έτος 1996 έως το έτος 2001. Το ηλεκτρονικό αυτό σύνολο κειμένων βρίσκεται σε αρχεία .html, από τα οποία με κατάλληλη επεξεργασία απομακρύνουμε το markup και κρατήσαμε μόνο το κεντρικό τμήμα του αρχείου αποθηκεύοντας το σε αρχεία .raw τα οποία είναι πιο εύκολα στην επεξεργασία. Με τον τρόπο αυτό για το κάθε .html αρχείο δημιουργήθηκε ένα .raw αρχείο το οποίο ουσιαστικά περιέχει σε μορφή απλού κειμένου το κύριο τμήμα του αντίστοιχου .html αρχείου. Το λεξιλόγιο των κειμένων θεωρείται αρκετά πλούσιο, αφού δεν έχουμε κείμενα μόνο μίας θεματολογίας αλλά πολλών όπως: τέχνη, οικονομία, εσωτερική επικαιρότητα, πολιτική, αθλητισμός και εξωτερική επικαιρότητα.

Όπως προαναφέραμε, η όλη εργασία δεν υλοποιήθηκε σε ένα και μόνο αρχείο αλλά σε αρκετά software modules, το καθένα από τα οποία εκτελεί ένα διαφορετικό μέρος της επίλυσης του προβλήματος. Ακολουθεί αναλυτική περιγραφή της λειτουργικότητας, της δομής και του ρόλου του κάθε software module.

#### *Εύρεση Συντομογραφιών*

Πρωταρχικός στόχος είναι να επιλεχθεί ένα σύνολο προτάσεων από όλα τα κείμενα που να καλύπτει τα κριτήρια που περιγράφηκαν αναλυτικά στην παράγραφο 3.1. Για να επιλεγούν οι προτάσεις αυτές, θα πρέπει πρώτα να χωριστεί το σύνολο των

κειμένων σε λογικές προτάσεις (επόμενο βήμα). Οι προτάσεις αυτές δεν πρέπει να είναι απλές γραμματοσειρές οι οποίες τελειώνουν με τελεία, ερωτηματικό ή θαυμαστικό αλλά προτάσεις όπως θα τις χώριζε ένας νοήμον άνθρωπος. Για να επιτευχθεί ο στόχος αυτός πρέπει να βρεθούν εκείνες οι γραμματοσειρές που θεωρούνται ότι είναι συντομογραφίες (γραμματοσειρές που περιέχουν τελεία ή τελείες) ώστε να γίνει σωστός ο διαχωρισμός των προτάσεων. Έτσι, εάν γνωρίζουμε τις συντομογραφίες θα μπορούμε με ασφάλεια να χωρίσουμε ένα κείμενο σε λογικές προτάσεις. Για παράδειγμα, εάν στο κείμενο: «Βρισκόμαστε στην 5<sup>η</sup> σελ. της αναφοράς. Υπάρχουν αρκετές σελίδες ακόμα.» γνωρίζουμε ότι το «σελ.» είναι συντομογραφία το πρόγραμμα θα χωρίσει το κείμενο σε δύο προτάσεις και όχι σε τρεις.

Για την εύρεση των συντομογραφιών χρησιμοποιήθηκε ένα λεξικό της ελληνικής γλώσσας (αρχείο Dictionary\_GR.lex) που περιέχει τις περισσότερες ελληνικές λέξεις (δυστυχώς όχι όλες) με τις προφορές τους. Ως συντομογραφία θεωρούμε τις παρακάτω κανονικές εκφράσεις οι οποίες δεν υπάρχουν στο λεξικό ως λέξεις:

1. Γραμματοσειρές αποτελούμενες από ένα έως και πέντε γράμματα οι οποίες τελειώνουν με τελεία και η επόμενη λέξη αρχίζει με πεζό γράμμα ή αριθμό (π.χ. βλ.)
2. Γραμματοσειρές αποτελούμενες από δομές όπου η κάθε δομή περιέχει από ένα έως και τέσσερα γράμματα και τελειώνει με τελεία (π.χ. κ.τ.λ.)

Το πρόγραμμα είναι ικανό να ανιχνεύσει συντομογραφίες έστω και αν αυτές περιέχουν διάφορα σύμβολα όπως κόμμα, παρενθέσεις, εισαγωγικά κ.α.

Δυστυχώς, το πρόγραμμα ανιχνεύει και λέξεις οι οποίες δεν είναι συντομογραφίες, λόγω ατελειών και σφαλμάτων στο ίδιο το κείμενο, όπως π.χ. όταν αρχίζει μια πρόταση με μικρό γράμμα και η λέξη της προηγούμενης πρότασης αποτελείται από ένα έως και πέντε γράμματα, καθώς επίσης και λόγω της έλλειψης ύπαρξης όλων των λέξεων όπως ονόματα, μήνες, ονόματα πόλεων κ.α. από το υπάρχον λεξικό. Ενδεικτικά αναφέρω πώς λόγω των ανωτέρω προβλημάτων, στις 200 περίπου λέξεις οι οποίες επιλέχθηκαν ως συντομογραφίες οι 19 δεν ήταν. Οι περισσότερες λέξεις που εσφαλμένα επιλέγηκαν ως συντομογραφίες ήταν λέξεις που δεν περιέχονταν στο λεξικό μας, όπως ονόματα ανθρώπων, πόλεων και μηνών.

Το πρόγραμμα που αναλαμβάνει να βρει όλες τις συντομογραφίες σε ένα σύνολο κειμένων είναι το `1.Find_Syntomografies.pl`, το οποίο παίρνει ως είσοδο:

- Ένα μονοπάτι σε φάκελο (μεταβλητή `$dir_name`)
- Το λεξικό της ελληνικής γλώσσας `Dictionary_GR.lex`

και ψάχνει σε όλα τα `.raw` αρχεία που βρίσκονται στον φάκελο εισόδου ή σε υποφάκελους του και βρίσκει όλες τις συντομογραφίες, τις αποθηκεύει στο αρχείο `syntomografies.txt` με μορφή συντομογραφία ανά σειρά. Αξίζει να επισημανθεί ότι το μονοπάτι σε φάκελο (directory path) είναι στην αρχή του κώδικα σε σημείο τέτοιο ώστε να μπορεί να παραμετροποιηθεί εύκολα χωρίς να γίνουν σημαντικές αλλαγές στον κώδικα. Το μόνο που χρειάζεται να κάνει ο χρήστης εάν θελήσει να παραμετροποιήσει το directory path είναι να αλλάξει την μεταβλητή `$dir_name`.

### *Χωρισμός Κειμένων σε Προτάσεις*

Ο χωρισμός όλων των κειμένων σε λογικές προτάσεις και η αποθήκευση αυτών σε ξεχωριστό αρχείο γίνεται σε αυτό το στάδιο. Η διαδικασία αυτή θεωρείται αρκετά σύνθετη, λόγω των πολλών ατελειών και σφαλμάτων του συνόλου των κειμένων.

Το πρόγραμμα αυτό διαβάζει σειριακά όλους τους χαρακτήρες των κειμένων και όπου συναντά τελεία ή ερωτηματικό ή θαυμαστικό αλλάζει πρόταση, εκτός και αν η λέξη που τελειώνει με τελεία είναι:

- συντομογραφία (π.χ. σελ.)
- αριθμός (π.χ. 10.000)
- e-mail (π.χ. dionysis@gmail.gr)
- ηλεκτρονική διεύθυνση (π.χ. <http://www.tuc.gr>)
- ημερομηνία (π.χ. 10.1.2006)
- αριθμητική παράσταση (π.χ.  $10 + 10.25 = 20.25$ )



Παράλληλα, το πρόγραμμα είναι ικανό να ανιχνεύσει κάποιες ατέλειες ή σφάλματα του κειμένου και να μην αλλάξει πρόταση χωρίς λόγο ή να αλλάξει πρόταση στο σωστό σημείο. Για παράδειγμα, στην περίπτωση που μια λέξη τελειώνει με τελεία και η επόμενη λέξη αρχίζει με μικρό γράμμα το πρόγραμμα δεν θα αλλάξει πρόταση. Επίσης, αν μετά από μια τελεία δεν υπάρχει κενό και η λέξη δεν τελειώνει με τελεία (ώστε να θεωρείται συντομογραφία), για παράδειγμα: «...στο χειμώνα.Αυριο είναι...» το πρόγραμμα θα αλλάξει πρόταση. Ενδεικτικά αναφέρω ότι η επιτυχία σωστής αλλαγής προτάσεων είναι αρκετά υψηλή εφόσον σε ένα τυχαίο δείγμα 200 προτάσεων από κάθε θεματολογία, μόλις 5 προτάσεις αλλάχθηκαν εσφαλμένα ή δεν έγινε αλλαγή πρότασης ενώ θα έπρεπε.

Το πρόγραμμα που αναλαμβάνει να χωρίσει τα κείμενα σε λογικές προτάσεις είναι το `2.Divide_corpus_to_Sentences.pl`, το οποίο παίρνει ως είσοδο:

- Ένα μονοπάτι σε φάκελο (μεταβλητή `$dir_name`)
- Το αρχείο των συντομογραφιών, από το προηγούμενο βήμα, `syntomografies.txt`

Το πρόγραμμα ψάχνει σε όλα τα `.raw` αρχεία που βρίσκονται στον φάκελο εισόδου ή σε υποφάκελους του και χωρίζει όλα τα κείμενα σε προτάσεις τις οποίες τις αποθηκεύει στο αρχείο `sentences.txt` με μορφή πρόταση ανά σειρά.

### *Εύρεση Προτάσεων με 3-10 Λέξεις*

Στο στάδιο αυτό αναλαμβάνουμε να ξεχωρίσουμε τις προτάσεις εκείνες από το σύνολο των κειμένων που περιέχουν από τρεις μέχρι και δέκα λέξεις και να τις αποθηκεύσουμε σε ένα ξεχωριστό αρχείο.

Το πρόγραμμα που αναλαμβάνει να βρει τις προτάσεις που περιέχουν από τρεις μέχρι και δέκα λέξεις είναι το `3.Find_Sentences_3-10_Words.pl`, το οποίο παίρνει ως είσοδο το αρχείο, από το προηγούμενο βήμα, `sentences.txt` που περιέχει όλες τις λογικές προτάσεις από το σύνολο κειμένων και αποθηκεύει τις επιλεγμένες προτάσεις στο αρχείο `candidate_sentences.txt` με μορφή πρόταση ανά σειρά.

## Αντικατάσταση Λέξεων με Προφορές

Στο σημείο αυτό αναλαμβάνουμε να αντικαταστήσουμε όλες τις λέξεις με τις προφορές τους και να αποθηκεύσουμε το αποτέλεσμα σε ξεχωριστό αρχείο. Το βήμα αυτό είναι απαραίτητο, ώστε να μπορέσουμε να βρούμε όλα τα διαφορετικά νόμιμα διφωνα (συνδυασμούς φωνημάτων - όχι κατ' ανάγκη όλοι, καθώς μερικοί συνδυασμοί δεν εμφανίζονται ποτέ στη γλώσσα) του κειμένου (επόμενο βήμα). Οι προφορές των λέξεων βρίσκονται στο λεξικό που χρησιμοποιούμε, δηλαδή στο αρχείο `Dictionary_GR.lex`.

Το πρόγραμμα που αναλαμβάνει να αντικαταστήσει όλες τις λέξεις με τις προφορές τους είναι το `4.Replace_Words_with_Pronunciation.pl` το οποίο παίρνει ως είσοδο:

- Το λεξικό της ελληνικής γλώσσας `Dictionary_GR.lex`, το οποίο περιέχει τις προφορές των λέξεων
- Το αρχείο, από το προηγούμενο βήμα, `candidate_sentences.txt`

Το πρόγραμμα, διαβάζοντας σειριακά όλες τις λέξεις του, τις αντικαθιστά με την προφορά τους και αποθηκεύει το αποτέλεσμα στο αρχείο `pronunciations.txt` με μορφή πρόταση ανά σειρά. Εάν κάποια λέξη δεν έχει προφορά στο λεξικό, το πρόγραμμα αγνοεί την λέξη αυτή.

Έτσι, για παράδειγμα εάν μια πρόταση είναι η : «Το μικρό παιδί.» η προφορά της θα είναι : «t o0 m i0 . k r o1 p e0 . dh i1 #».

## Εύρεση Διφώνων

Η εύρεση όλων των διφώνων (συνδυασμούς φωνημάτων) που υπάρχουν στις επιλεγμένες προτάσεις καθώς και η αποθήκευση αυτών σε ξεχωριστό αρχείο γίνεται στο στάδιο αυτό.

Το πρόγραμμα που αναλαμβάνει να βρει όλα τα δίφωνα που υπάρχουν σε ένα σύνολο προτάσεων είναι το `5.Find_Difones.pl` το οποίο παίρνει ως είσοδο το αρχείο, από το προηγούμενο βήμα, `pronunciations.txt`, που περιέχει τις προφορές των λέξεων, και ενώνοντας τα φωνήματα ανά δυάδες δημιουργεί όλα τα δίφωνα. Τέλος αποθηκεύει όλα τα διαφορετικά δίφωνα, το ένα κάτω από το άλλο, σύμφωνα με την συχνότητα εμφάνισης του κάθε δίφωνα, με πρώτα τα πιο σπάνια δίφωνα, στο αρχείο `diphones.txt` με μορφή δίφωνο ανά σειρά.

## Αντικατάσταση Προφορών με Δίφωνα

Το επόμενο βήμα είναι να αντικατασταθούν οι προφορές με τον αύξοντα αριθμό κάθε δίφωνα και να αποθηκευτεί το αποτέλεσμα σε ξεχωριστό αρχείο.

Το πρόγραμμα που αναλαμβάνει να αντικαταστήσει όλες τις προφορές με τον αύξοντα αριθμό κάθε δίφωνα είναι το `6.Replace_Pronunciation_with_Diphones_Indexes.pl` το οποίο παίρνει ως είσοδο τα αρχεία, από προηγούμενα βήματα:

- `diphones.txt`
- `pronunciations.txt`

και αποθηκεύει το αποτέλεσμα στο αρχείο `sentences_with_diphone_indexes.txt` με μορφή πρόταση ανά σειρά, όπου η εκάστοτε πρόταση είναι ένα σύνολο από αριθμούς, χωρισμένους με κενό, όπου ο καθένας αντιπροσωπεύει ένα δίφωνο.

Έτσι για παράδειγμα, εάν μια πρόταση είναι η : «Το μικρό παιδί.» η προφορά της θα είναι : «t o0 m i0 . k r o1 p e0 . dh i1 #». Τα δίφωνα που υπάρχουν είναι τα: #

t, t o0, o0 m, m i0, i0 k, k r, r o1, o1 p, p e0, e0 dh, dh i1, i1 #. Άρα η τελική πρόταση θα έχει την μορφή: «8 5 12 58 78 45 6 1 87 23 14 89» όπου το νούμερο 8 αντιπροσωπεύει το 8<sup>ο</sup> δίφωνο του αρχείου `diphones.txt` δηλαδή το δίφωνο # t, το 5 αντιπροσωπεύει το 5<sup>ο</sup> δίφωνο του αρχείου `diphones.txt` δηλαδή το δίφωνο o0 m και ούτω το καθεξής (οι αριθμοί 8 5 12... δόθηκαν τυχαία για το συγκεκριμένο παράδειγμα).

### *Εύρεση Προτάσεων όπου Ανήκει το κάθε Δίφωνο*

Στο στάδιο αυτό βρίσκουμε σε ποιες προτάσεις ανήκει το ελάχιστοτε δίφωνο και αποθηκεύουμε το αποτέλεσμα σε ένα νέο αρχείο. Το βήμα αυτό είναι απαραίτητο για την πιο εύκολη επιλογή των τελικών προτάσεων.

Το πρόγραμμα που αναλαμβάνει να βρει σε ποιες προτάσεις υπάρχει το κάθε δίφωνο είναι το `7.Find_where_Diphones_Exist.pl` το οποίο παίρνει ως είσοδο:

- Το αρχείο με τα δίφωνα `diphones.txt`
- Το αρχείο από το προηγούμενο βήμα `sentences_with_diphone_indexes.txt`

και έχει ως έξοδο το αρχείο `diphones_in_sentences.txt`. Η μορφή του αρχείου εξόδου είναι η εξής: κάθε γραμμή του αρχείου αντιστοιχεί σε ένα δίφωνο και περιέχει τους αύξοντες αριθμούς των προτάσεων που περιέχουν το δίφωνο αυτό.

Αν, για παράδειγμα, η πρώτη γραμμή του αρχείου είναι: '53 60', σημαίνει ότι το πρώτο δίφωνο από το αρχείο `diphones.txt` βρίσκεται στις προτάσεις με αύξοντες αριθμούς 53 και 60 από το αρχείο των υποψήφιων προς επιλογή προτάσεων `candidate_sentences.txt`.

## Καθαρισμός των Προτάσεων

Στο στάδιο αυτό του προγράμματος «καθαρίζουμε» το σύνολο των προτάσεων από όλα τα πιθανά σύμβολα όπως: κόμματα, τελείες, θαυμαστικά, ερωτηματικά κ.α. Στόχος μας είναι να δημιουργήσουμε ένα σύνολο προτάσεων με καθαρές λέξεις, ώστε να βρούμε έπειτα όσο μεγαλύτερο αριθμό λεκτικών ριζών με τη χρήση άλλων προγραμμάτων, τα οποία έχουν μεγαλύτερη επιτυχία στην εύρεση λεκτικών ριζών όταν το κείμενο είναι «καθαρό».

Το πρόγραμμα που αναλαμβάνει να καθαρίσει ένα κείμενο από σύμβολα και σημεία στίξης είναι το `A.Clear_Corpus.pl`, το οποίο παίρνει ως είσοδο ένα κείμενο όπως το `candidate_sentences.txt` και αφού το «καθαρίζει» από σύμβολα και σημεία στίξης αποθηκεύει το αποτέλεσμα στο `candidate_sentences_cleared.txt` με την ίδια σειρά προτάσεων με το κείμενο εισόδου.

## Εύρεση Ριζών

Στο στάδιο αυτό αναλαμβάνουμε να βρούμε την κατάληξη και την ρίζα κάθε λέξης που βρίσκεται στο αρχείο με τις «καθαρισμένες» προτάσεις. Μας ενδιαφέρει να βρούμε τις ρίζες των λέξεων, γιατί σε επόμενα βήματα θα επιλεγούν προτάσεις με κριτήριο την ποικιλία των διαφορετικών λεκτικών ριζών που περιέχουν.

Για το σκοπό αυτό χρησιμοποιήσαμε ένα πρόγραμμα το οποίο κάνει μορφολογική ανάλυση κειμένου, το `linguistica`, το οποίο χρησιμοποιώντας στατιστικά μοντέλα είναι ικανό να διαχωρίσει από μια λέξη τη ρίζα και την κατάληξη της. Ο βαθμός επιτυχίας του προγράμματος εξαρτάται άμεσα από το μέγεθος του κειμένου που δέχεται το πρόγραμμα ως είσοδο. Έτσι, όσο μεγαλύτερο το κείμενο που δέχεται το πρόγραμμα ως είσοδο, τόσο καλύτερα αποτελέσματα έχει στην κατάτμηση των λέξεων σε ρίζες και καταλήξεις. Για τον λόγο αυτό ως είσοδο στο πρόγραμμα επιλέξαμε το αρχείο `sentences_cleared.txt`, που περιέχει όλες τις προτάσεις «καθαρισμένες» και έτσι πετύχαμε μέγιστο δυνατό και επιτυχή διαχωρισμό

καταλήξεων και ριζών. Το αποτέλεσμα του διαχωρισμού αποθηκεύτηκε στο αρχείο `linguistica.txt`, το οποίο έχει την εξής μορφή ανά σειρά:

λέξη –tab- 1 ή 2 (το 1 δηλώνει ότι για την συγκεκριμένη λέξη δεν διαχωρίστηκε η ρίζα με την κατάληξη και το 2 δηλώνει ότι διαχωρίστηκε) –tab- ρίζα λέξης (εάν το 2<sup>ο</sup> πεδίο είναι 2) –tab- κατάληξη λέξης (εάν το 2<sup>ο</sup> πεδίο είναι 2)

### Αντικατάσταση Λέξεων με τις Ρίζες

Στο στάδιο αυτό του προγράμματος γίνεται η αντικατάσταση των λέξεων από το αρχείο των προς επιλογή προτάσεων με τις ρίζες τους. Εάν μια λέξη δεν διαθέτει λεκτική ρίζα (όπως π.χ. η λέξη Ιούνιος), τότε το πρόγραμμα δεν αντικαθιστά την λέξη αυτή. Το βήμα αυτό θα μας διευκολύνει ώστε να ανιχνεύσουμε σε επόμενα στάδια σε ποιες προτάσεις βρίσκεται η κάθε λεκτική ρίζα και πόσες φορές υπάρχει αυτή στο κείμενο.

Το πρόγραμμα που αναλαμβάνει να αντικαταστήσει τις λέξεις με τις ρίζες τους είναι το `B.Replace_Words_with_Roots.pl` το οποίο παίρνει ως είσοδο:

- Το αρχείο `linguistica.txt` από το προηγούμενο βήμα, που περιέχει τις ρίζες των εκάστοτε λέξεων.
- Το αρχείο `candidate_sentences_cleared.txt` με τις «καθαρισμένες» από σύμβολα και σημεία στίξης προτάσεις

και αποθηκεύει το αποτέλεσμα στο αρχείο `candidate_sentences_root.txt` με μορφή πρόταση ανά σειρά.

## Εύρεση Προτάσεων όπου Ανήκει η κάθε Ρίζα

Στο στάδιο αυτό βρίσκουμε σε ποιες προτάσεις ανήκει η εκάστοτε ρίζα και αποθηκεύουμε το αποτέλεσμα σε ένα νέο αρχείο. Παράλληλα βρίσκουμε και πόσες φορές υπάρχει η κάθε ρίζα στο αρχείο εισόδου.

Το πρόγραμμα που αναλαμβάνει να βρει σε ποιες προτάσεις υπάρχει η κάθε ρίζα και πόσες φορές μέσα στο αρχείο εισόδου, είναι το `C.Find_where_Roots_Exist.pl`, το οποίο παίρνει ως είσοδο το αρχείο από το προηγούμενο βήμα `candidate_sentences_root.txt`, με τις υποψήφιες προς επιλογή προτάσεις με τις ρίζες των λέξεων και αποθηκεύει το αποτέλεσμα στο αρχείο `roots_in_sentences.txt` με μορφή ανά σειρά:

ρίζα -tab- αύξων αριθμός των προτάσεων που υπάρχει η ρίζα χωρισμένες με κόμμα -tab- τον συνολικό αριθμό που υπάρχει η ρίζα μέσα στο αρχείο εισόδου.

## Επιλογή Προτάσεων

Στο στάδιο αυτό της επεξεργασίας γίνεται η επιλογή ενός ελάχιστου ή ενός επιθυμητού (προαιρετικό) αριθμού προτάσεων από το αρχείο των υποψηφίων προς επιλογή προτάσεων που να πληρούν τα παρακάτω κριτήρια:

1. Να περιέχουν όλα τα φωνήματα της ελληνικής γλώσσας.
2. Να περιέχουν όλα τα υπαρκτά δίφωνα της ελληνικής γλώσσας που βρίσκονται στο σύνολο των κειμένων.
3. Να περιέχουν όσο δυνατόν περισσότερες διαφορετικές λεκτικές ρίζες.
4. Να περιέχουν όσο το δυνατόν ίσο - σταθμισμένο αριθμό διφώνων.
5. Να περιέχουν κάποιες συγκεκριμένες λέξεις (προαιρετικό).

Για την επίτευξη των παραπάνω στόχων δημιουργήθηκαν τρία προγράμματα τα οποία υλοποιούν τρεις διαφορετικούς αλγορίθμους για την επιλογή των τελικών προτάσεων.

Τα προγράμματα αυτά είναι:

- 8.Select\_Sentences.pl
- 8.Select\_Sentences\_V2++.pl
- 8.Select\_Sentences\_V2-.pl

Ακολουθεί αναλυτική περιγραφή του τρόπου επιλογής των τελικών προτάσεων από το εκάστοτε πρόγραμμα.

Στο **πρώτο** πρόγραμμα, 8.Select\_Sentences.pl, πρωταρχικός στόχος είναι η επιλογή ενός ελάχιστου αριθμού προτάσεων που να περιέχει όλα τα υπαρκτά δίφωνα της ελληνικής γλώσσας που βρίσκονται στο σύνολο των κειμένων. Για να επιτευχθεί ο στόχος αυτός, το πρόγραμμα παίρνει σειριακά όλα τα δίφωνα που βρέθηκαν στο σύνολο των κειμένων και για κάθε δίφωνο επιλέγει τη πρόταση με τη μεγαλύτερη βαθμολογία που να το περιέχει και την αποθηκεύει στο αρχείο των τελικών προτάσεων.

Το κάθε πρόγραμμα βαθμολογεί τις προτάσεις με διαφορετικό τρόπο. Το 8.Select\_Sentences.pl βρίσκει πόσα διαφορετικά δίφωνα περιέχει κάθε πρόταση και αναθέτει σε αυτή βαθμολογία ίση με τον αριθμό των διαφορετικών διφώνων που αυτή περιέχει. Όταν μια πρόταση επιλεγεί, το πρόγραμμα βρίσκει τα διαφορετικά δίφωνα που συγκροτούν την πρόταση αυτή και μειώνει ένα βαθμό από τη βαθμολογία κάθε πρότασης που περιέχει κάποιο από τα δίφωνα που συγκροτούν την επιλεγμένη πρόταση. Έτσι, στο τέλος του προγράμματος, όλες οι προτάσεις έχουν βαθμολογία ίση με το μηδέν, απόλυτα λογικό και αναμενόμενο εφόσον επιλέχθηκαν όλα τα δίφωνα από τουλάχιστον μία φορά το καθένα.

Το 8.Select\_Sentences.pl παίρνει ως είσοδο:

- Το αρχείο diphones\_in\_sentences.txt, που δείχνει σε ποιες προτάσεις ανήκει το εκάστοτε δίφωνο.
- Το αρχείο sentences\_with\_diphone\_indexes.txt, που δείχνει ποια δίφωνα περιέχει η εκάστοτε πρόταση.

και έχει ως έξοδο το αρχείο selected\_sentences\_with\_indexes.txt που περιέχει τις επιλεγμένες προτάσεις με μορφή: αύξων αριθμός πρότασης (από το αρχείο των υποψήφιων προς επιλογή προτάσεων candidate\_sentences.txt) ανά γραμμή.



Στο **δεύτερο** πρόγραμμα 8.Select\_Sentences\_V2++.pl πάλι πρωταρχικός στόχος είναι η επιλογή ενός ελάχιστου αριθμού προτάσεων που να περιέχουν όλα τα υπαρκτά δίφωνα της ελληνικής γλώσσας που βρίσκονται στο σύνολο των κειμένων, όπως άλλωστε και στο πρώτο πρόγραμμα. Το μόνο που διαφέρει είναι ο τρόπος βαθμολόγησης κάθε πρότασης.

Εδώ, στην αρχή του προγράμματος υπολογίζεται ο αριθμός των συνολικών διφώνων που περιέχονται στο αρχείο με τις υποψήφιες προς επιλογή προτάσεις και επίσης τον αριθμό ύπαρξης κάθε διφώνου στο αρχείο αυτό. Η βαθμολογία κάθε πρότασης γίνεται ως εξής:

1. Το πρόγραμμα βρίσκει πόσα διαφορετικά δίφωνα περιέχει κάθε πρόταση.
2. Για κάθε δίφωνο προσθέτει στη βαθμολογία της ελάχιστοτε πρότασης το αποτέλεσμα της διαίρεσης του αριθμού των συνολικών διφώνων που περιέχονται στο αρχείο με τις υποψήφιες για επιλογή προτάσεις προς τον αριθμό συνολικών εμφανίσεων του διφώνου αυτού.
3. Όταν μια πρόταση επιλεγεί, το πρόγραμμα βρίσκει τα διαφορετικά δίφωνα που συγκροτούν την πρόταση αυτή και μειώνει τόσους βαθμούς από τη βαθμολογία κάθε πρότασης που περιέχει κάποιο από τα δίφωνα που συγκροτούν την επιλεγμένη πρόταση όσο είναι το αποτέλεσμα της διαίρεσης του αριθμού των συνολικών διφώνων που περιέχονται στο αρχείο με τις υποψήφιες για επιλογή προτάσεις προς τον αριθμό συνολικών εμφανίσεων του διφώνου αυτού.

Έτσι για παράδειγμα, εάν το αρχείο με τις υποψήφιες προς επιλογή προτάσεις περιέχει 1000 δίφωνα και το δίφωνο «α ε» υπάρχει στο αρχείο 10 φορές, εάν επιλεγεί μια πρόταση που περιέχει το δίφωνο αυτό τότε θα αφαιρεθούν από κάθε πρόταση που έχει το δίφωνο «α ε» 1000/10 βαθμοί. Με τον τρόπο αυτό, στο τέλος του προγράμματος όλες οι προτάσεις έχουν βαθμολογία ίση με το μηδέν, κάτι απόλυτα λογικό εφόσον επιλέχθηκαν όλα τα δίφωνα από τουλάχιστον μία φορά το καθένα.

Παράλληλα, το πρόγραμμα δίνει την δυνατότητα να καθορίσει ο χρήστης και τον επιθυμητό αριθμό επιλεγμένων προτάσεων, καθορίζοντας την μεταβλητή \$desirable\_num\_of\_senetences. Όταν η τιμή της μεταβλητής αυτής είναι μηδέν ή μικρότερη από τον αριθμό των ελάχιστων δυνατών επιλεγμένων προτάσεων, τότε το πρόγραμμα αγνοεί την ύπαρξή της. Οι επιπλέον προτάσεις που θα επιλεγούν με την

μέθοδο αυτή έχουν μπροστά από τον αύξων αριθμό της εκάστοτε πρότασης στο αρχείο εξόδου το σύμβολο «-».

Τέλος το πρόγραμμα δίνει την δυνατότητα στο χρήστη να μπορεί να επιλέξει ένα σύνολο προτάσεων που να περιέχουν οπωσδήποτε κάποιες επιθυμητές λέξεις ή ρίζες λέξεων. Οι προτάσεις αυτές θα προστεθούν στις υπόλοιπες επιλεγμένες προτάσεις. Οι λέξεις ή οι ρίζες αυτές καθορίζονται από το χρήστη και πρέπει να βρίσκονται στο αρχείο `desirable_words.txt` με μορφή λέξη - ρίζα ανά σειρά. Οι επιπλέον προτάσεις που θα επιλεγούν με την μέθοδο αυτή έχουν μπροστά από τον αύξων αριθμό της εκάστοτε πρότασης στο αρχείο εξόδου το σύμβολο «+».

Το `8.Select_Sentences_V2++.pl` παίρνει ως είσοδο:

- Το αρχείο `diphones_in_sentences.txt`, που δείχνει σε ποιες προτάσεις ανήκει το εκάστοτε δίφωνο.
- Το αρχείο `sentences_with_diphone_indexes.txt`, που δείχνει ποια δίφωνα περιέχει η εκάστοτε πρόταση.
- Το αρχείο `candidate_sentences.txt` με τις προς επιλογή προτάσεις.
- Το αρχείο `desirable_words.txt` με τις επιθυμητές λέξεις (προαιρετικό).

και έχει ως έξοδο το αρχείο `selected_sentences_with_indexes.txt` που περιέχει τις επιλεγμένες προτάσεις με μορφή: αύξων αριθμός πρότασης (από το αρχείο των υποψήφιων προς επιλογή προτάσεων `candidate_sentences.txt`) ανά γραμμή.

Στο **τρίτο** πρόγραμμα `8.Select_Sentences_V2-.pl` και πάλι ο πρωταρχικός στόχος είναι η επιλογή ενός ελάχιστου αριθμού προτάσεων που να περιέχει όλα τα υπαρκτά δίφωνα της ελληνικής γλώσσας που βρίσκονται στο σύνολο των κειμένων, όπως άλλωστε και προηγούμενα πρόγραμμα, με τη διαφορά ότι εδώ έχουμε διαφορετικό τρόπο βαθμολόγησης κάθε πρότασης.

Το παρόν πρόγραμμα βαθμολογεί κάθε πρόταση με δύο διαφορετικά κριτήρια:

1. Το πρώτο είναι η ποικιλότητα και η διαφορετικότητα των διφώνων σε κάθε πρόταση. Η βαθμολόγηση κάθε πρότασης με αυτό το κριτήριο είναι ίδια με αυτή του προγράμματος `8.Select_Sentences_V2++.pl`.
2. Το δεύτερο είναι η ποικιλότητα και η διαφορετικότητα των λεκτικών ριζών που περιέχει κάθε πρόταση. Από τα προηγούμενα βήματα γνωρίζουμε πόσες φορές υπάρχει η κάθε λεκτική ρίζα στο σύνολο των προτάσεων, καθώς επίσης και σε ποιες προτάσεις βρίσκεται η κάθε ρίζα. Η βαθμολογία της εκάστοτε πρότασης με αυτό το κριτήριο γίνεται ως εξής: το πρόγραμμα βρίσκει πόσες διαφορετικές ρίζες έχει η εκάστοτε πρόταση και για κάθε ρίζα προσθέτει στην βαθμολογία της πρότασης τόσους βαθμούς όσες και οι φορές που υπάρχει η τρέχουσα ρίζα στο σύνολο των προτάσεων. Όταν μια πρόταση επιλεγεί, το πρόγραμμα βρίσκει τις διαφορετικές ρίζες που συγκροτούν την πρόταση αυτή και μειώνει τόσους βαθμούς από τη βαθμολογία κάθε πρότασης που περιέχει κάποια από τις ρίζες που συγκροτούν την επιλεγμένη πρόταση όσες και οι φορές που υπάρχει η κάθε λεκτική ρίζα στο σύνολο των κειμένων.

Η τελική βαθμολογία της κάθε πρότασης προκύπτει από το άθροισμα των δύο επιμέρους βαθμολογιών (μία από το κάθε κριτήριο). Επειδή όμως παρατηρήθηκε ότι η βαθμολογία της κάθε πρότασης με κριτήριο τη διαφορετικότητα των διφώνων είναι πολύ μεγαλύτερη από τη βαθμολογία με κριτήριο τη ποικιλότητα και τη διαφορετικότητα των λεκτικών ριζών, η τελική βαθμολογία προκύπτει από το κανονικοποιημένο άθροισμα των δύο βαθμολογιών. Πιο συγκεκριμένα, το πρόγραμμα υπολογίζει τους μέσους όρους βαθμολόγησης των προτάσεων από το κάθε κριτήριο. Έπειτα διαιρεί τον μέσο όρο της βαθμολόγησης των προτάσεων από το κριτήριο της διαφορετικότητας των διφώνων με το μέσο όρο της βαθμολόγησης των προτάσεων από το κριτήριο της διαφορετικότητας των ριζών και με το αποτέλεσμα πολλαπλασιάζει την βαθμολογία κάθε πρότασης από το κριτήριο της διαφορετικότητας των ριζών. Η τελική βαθμολογία κάθε πρότασης προκύπτει από το άθροισμα της βαθμολογίας της από το κριτήριο της διαφορετικότητας των διφώνων

με την κανονικοποιημένη βαθμολογία από κριτήριο της διαφορετικότητας των λεκτικών ριζών (το γινόμενο που περιγράψαμε πιο πάνω). Για παράδειγμα, εάν μια πρόταση έχει βαθμολογία 23 από το κριτήριο της διαφορετικότητας των διφώνων, βαθμολογία 3 από κριτήριο της διαφορετικότητας των λεκτικών ριζών και ο μέσος όρος των βαθμολογιών των προτάσεων με κριτήριο τη διαφορετικότητα των διφώνων είναι 16 ενώ ο μέσος όρος των βαθμολογιών των προτάσεων με κριτήριο τη διαφορετικότητα των ριζών είναι 2, τότε η τελική βαθμολογία της πρότασης θα είναι:  $23 + (16/2)*3$ .

Κανονικά, τα δύο παραπάνω κριτήρια βαθμολόγησης θα έπρεπε να κανονικοποιηθούν με βάση τις διακυμάνσεις και όχι τις μέσες τιμές τις κάθε ποσότητας. Για το λόγο αυτό, ο χρήστης έχει την δυνατότητα να αυξομειώσει κάθε ένα κριτήριο βαθμολόγησης ξεχωριστά αλλάζοντας τις μεταβλητές που βρίσκονται στη αρχή του προγράμματος: `$diphones_weight` (βαρύτητα βαθμολόγησης του 1<sup>ου</sup> κριτηρίου) , `$roots_weight` (βαρύτητα βαθμολόγησης του 2<sup>ου</sup> κριτηρίου) αλλάζοντας έτσι τον τρόπο βαθμολόγησης κάθε πρότασης. Έτσι με τον τρόπο αυτό, μπορεί ο χρήστης να επιλέξει προτάσεις με περισσότερα διαφορετικά δίφωνα ή περισσότερες διαφορετικές λεκτικές ρίζες. Αξίζει να σημειωθεί ότι, εξ αρχής τα δύο παραπάνω κριτήρια βαθμολογούν την κάθε πρόταση με βαρύτητα το κάθε ένα 50% και 50% αντίστοιχα.

Παράλληλα, το πρόγραμμα δίνει την δυνατότητα στο χρήστη να επιλέξει ένα σύνολο προτάσεων που να περιέχουν οπωσδήποτε κάποιες επιθυμητές λέξεις ή ρίζες λέξεων ή και να καθορίσει τον επιθυμητό αριθμό επιλεγμένων προτάσεων, όπως ακριβώς και στο πρόγραμμα `8.Select_Sentences_V2++.pl`.

Το `8.Select_Sentences_V2-.pl` παίρνει ως είσοδο:

- Το αρχείο `diphones_in_sentences.txt`, που δείχνει σε ποιες προτάσεις ανήκει το εκάστοτε δίφωνο.
- Το αρχείο `sentences_with_diphone_indexes.txt`, που δείχνει ποια δίφωνα περιέχει η εκάστοτε πρόταση.
- Το αρχείο `candidate_sentences.txt` με τις προς επιλογή προτάσεις.
- Το αρχείο `desirable_words.txt` με τις επιθυμητές προς επιλογή λέξεις (προαιρετικό).

- Το αρχείο `candidate_sentences_root.txt` με τις προς επιλογή προτάσεις με μορφή ριζών.
- Το αρχείο `roots_in_sentences.txt` με πληροφορίες σχετικά με τις λεκτικές ρίζες.

και έχει ως έξοδο το αρχείο `selected_sentences_with_indexes.txt` που περιέχει τις επιλεγμένες προτάσεις με μορφή: αύξων αριθμός πρότασης (από το αρχείο των υποψήφιων προς επιλογή προτάσεων `candidate_sentences.txt`) ανά γραμμή.

### *Αποθήκευση Επιλεγμένων Προτάσεων*

Στο τελικό αυτό στάδιο γίνεται η αποθήκευση των επιλεγμένων προτάσεων, από το προηγούμενο βήμα, με τις προφορές σε ένα ξεχωριστό αρχείο. Το βήμα αυτό απαιτείται εφόσον οι επιλεγμένες προτάσεις από το προηγούμενο βήμα είναι με μορφή: αύξων αριθμός για κάθε πρόταση.

Το πρόγραμμα που αναλαμβάνει να κάνει την παραπάνω διαδικασία είναι το `9.Find_Sentences_from_Indexes.pl`, το οποίο παίρνει ως είσοδο τα αρχεία από προηγούμενα βήματα:

- `pronunciations.txt`.
- `selected_sentences_with_indexes.txt`.

και διαβάζοντας σειριακά τους αύξοντες αριθμούς των επιλεγμένων προτάσεων από το αρχείο εισόδου `selected_sentences_with_indexes.txt` βρίσκει ποια πρόταση αντιστοιχεί στον ελάχιστο αύξοντα αριθμό και την αποθηκεύει με την προφορά της στο αρχείο `selected_sentences.txt` με μορφή: προφορά πρότασης ανά γραμμή.

Τέλος, για λόγους λειτουργικότητας και ευκολίας, δημιουργήθηκε ένα πρόγραμμα, το `make_file_program.pl` το οποίο τρέχει όλα τα παραπάνω προγράμματα με τη σειρά και παράγει το επιθυμητό αποτέλεσμα ώστε να μην είναι αναγκασμένος ο χρήστης να τρέξει ένα προς ένα όλα τα παραπάνω προγράμματα.

---

# Αποτελέσματα

Όπως περιγράψαμε στο προηγούμενο κεφάλαιο, καταφέραμε με τη χρήση τεσσάρων διαφορετικών αλγορίθμων να επιλέξουμε έναν ελάχιστο αριθμό προτάσεων, από τις οποίες θα ηχογραφήσουμε το σώμα φωνής. Εφόσον δεν ηχογραφήθηκε ακόμα το σώμα φωνής, ώστε να μπορούμε να δούμε τα οφέλη αυτής της βέλτιστης επιλογής κειμένων σε ένα σύστημα σύνθεσης φωνής, προσπαθήσαμε μέσω των αποτελεσμάτων των αλγορίθμων αυτών να αξιολογήσουμε το σύστημα μας. Έτσι, στο κεφάλαιο αυτό θα αναλύσουμε τα αποτελέσματα του κάθε αλγορίθμου, συγκρίνοντας και αξιολογώντας τις κατανομές διφώνων και λεκτικών ριζών από την κάθε μέθοδο επιλογής προτάσεων. Παράλληλα, με τον υπολογισμό της εντροπίας και της σχετικής εντροπίας των κατανομών αυτών, θα μπορέσουμε να συγκρίνουμε κατά πόσο «μοιάζει» το αποτέλεσμα κάθε μεθόδου επιλογής προτάσεων με το σύνολο κειμένων που δέχεται το πρόγραμμα μας ως είσοδο, στον τομέα της ποικιλότητας και τις διαφορετικότητας λεκτικών ριζών και διφώνων. Τέλος, επειδή κάθε αλγόριθμος επιλέγει έναν διαφορετικό αριθμό προτάσεων θα συγκρίνουμε και την απόδοση κάθε αλγορίθμου, εξαναγκάζοντας κάθε μέθοδο επιλογής να επιλέξει ένα συγκεκριμένο αριθμό προτάσεων, ώστε να μπορέσουμε να συγκρίνουμε την περιεκτικότητα των επιλεγμένων προτάσεων σε λεκτικές ρίζες και σε δίφωνα όταν έχουμε τον ίδιο αριθμό τελικών προτάσεων.

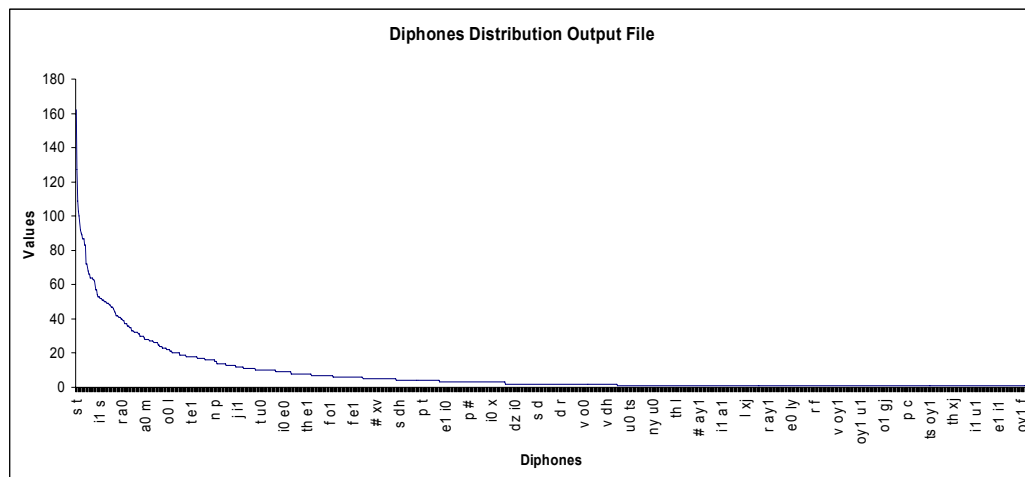
## 4.1 Αποτελεσμάτα κάθε Αλγορίθμου Επιλογής Προτάσεων

### Αποτελέσματα πρώτου αλγόριθμου

Ο πρώτος αλγόριθμος επιλογής προτάσεων επιλέγει τις τελικές προτάσεις από ένα σύνολο κειμένων με μοναδικό κριτήριο επιλογής την διαφορετικότητα των διφώνων. Έτσι, επιλέγει έναν ελάχιστο αριθμό προτάσεων που να περιέχουν όλα τα υπαρκτά δίφωνα της ελληνικής γλώσσας που βρίσκονται στο σύνολο των κειμένων και έναν αριθμό διαφορετικών λεκτικών ριζών.

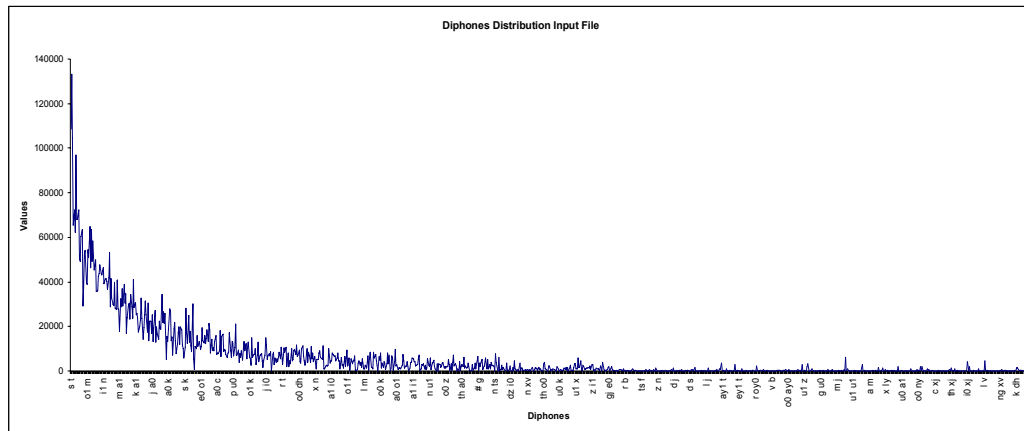
Ο αλγόριθμος αυτός επιλέγει τελικά 260 προτάσεις από ένα σύνολο 1.165.868 προτάσεων οι οποίες περιέχουν έστω και μία φορά όλα τα 1.114 δίφωνα που υπήρχαν στο σύνολο των αρχείων, οι προτάσεις αυτές περιέχουν επίσης 1.221 διαφορετικές λεκτικές ρίζες.

Η ακόλουθη κατανομή (εικόνα 4.1), μας παρουσιάζει τα δίφωνα που βρίσκονται στο αρχείο με τις επιλεγμένες προτάσεις (αρχείο εξόδου) σε συνάρτηση με τις πόσες φορές υπάρχει το κάθε δίφωνο στο αρχείο αυτό.



Εικόνα 4.1 Κατανομή διφώνων στο αρχείο εξόδου

Ακολουθεί η αντίστοιχη κατανομή για το αρχείο εισόδου, εικόνα 4.2 (σώμα κειμένων).



*Εικόνα 4.2 Κατανομή διφώνων στο αρχείο εισόδου*

Παράλληλα, υπολογίσαμε την μέση ποσότητα πληροφορίας ανά δίφωνο, δηλαδή την εντροπία κάθε κατανομής, ώστε να μπορούμε να συγκρίνουμε τις δύο αυτές κατανομές (αρχείο εξόδου – αρχείο εισόδου), με την χρήση του μαθηματικού τύπου:

$$H(x) = -\sum_{n=1}^N p(x_n) \log_2 p(x_n) \quad (4.1)$$

Τα αποτελέσματα παραθέτονται στον πίνακα 4.1:

Εντροπία - Αρχείο Εισόδου	Εντροπία - Αρχείο Εξόδου
8.337 bits/diphone	8.733 bits/diphone

*Πίνακας 4.1 Μέση πληροφορία ανά δίφωνο*

Τέλος, με την χρήση του παρακάτω μαθηματικού τύπου υπολογίσαμε την μέση «απόσταση» ανάμεσα στις δύο κατανομές (αρχείο εξόδου – αρχείο εισόδου), υπολογίζοντας την σχετική εντροπία τους:

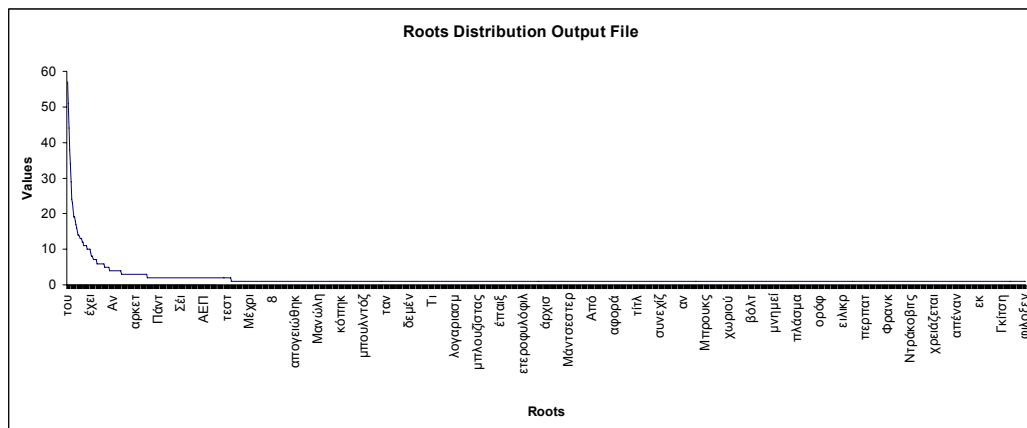
$$D(p \parallel q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)} \quad (4.2)$$



Όπου  $p$  είναι η κατανομή διφώνων του αρχείου εξόδου και  $q$  είναι η κατανομή διφώνων του αρχείου εισόδου. Η  $D(p \parallel q)$  τελικά είναι 0.347 bits.

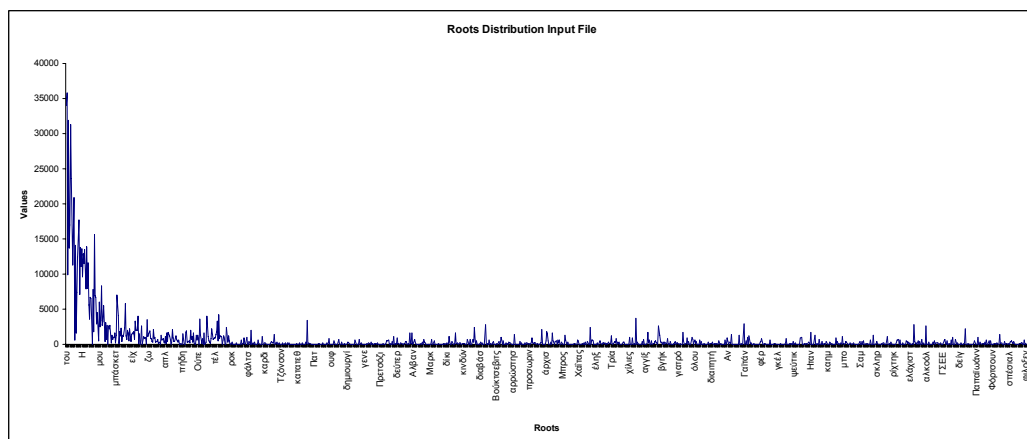
Στην συνέχεια παραθέτουμε τα αντίστοιχα αποτελέσματα για τις κατανομές των λεκτικών ριζών.

Η ακόλουθη κατανομή, εικόνα 4.3, μας παρουσιάζει τις λεκτικές ρίζες που βρίσκονται στο αρχείο με τις επιλεγμένες προτάσεις (αρχείο εξόδου) σε συνάρτηση με τις πόσες φορές υπάρχει η κάθε ρίζα στο αρχείο αυτό.



**Εικόνα 4.3** Κατανομή λεκτικών ριζών στο αρχείο εξόδου

Ακολουθεί η αντίστοιχη κατανομή για το αρχείο εισόδου, εικόνα 4.4 (σώμα κειμένων).



**Εικόνα 4.4** Κατανομή λεκτικών ριζών στο αρχείο εισόδου

Η μέση πληροφορία ανά λεκτική ρίζα παρατίθεται στον πίνακα 4.2, και υπολογίστηκε με την χρήση του μαθηματικού τύπου 4.1:

Εντροπία - Αρχείο Εισόδου	Εντροπία - Αρχείο Εξόδου
11.282 bits/root	9.373 bits/root

*Πίνακας 4.2 Μέση πληροφορία ανά λεκτική ρίζα*

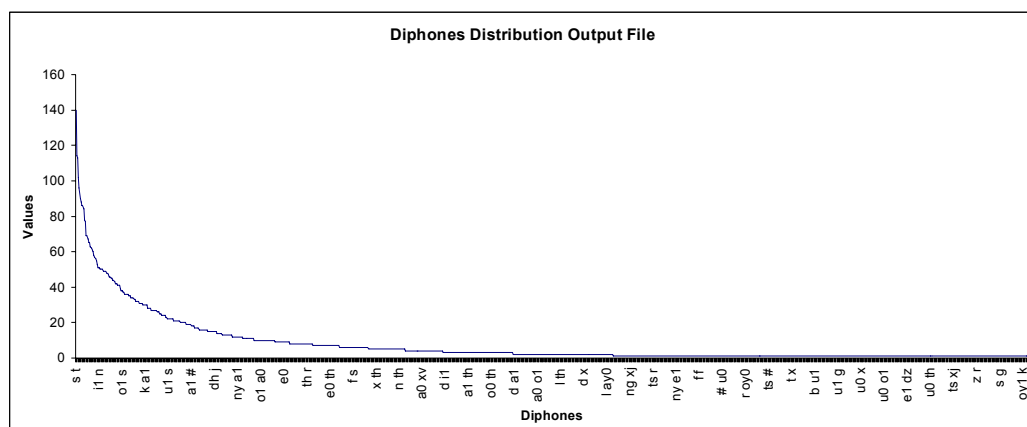
Τέλος, η σχετική εντροπία των δύο κατανομών είναι  $D(p \parallel q) = 2.644$  bits, όπου  $p$  είναι η κατανομή λεκτικών ριζών του αρχείου εξόδου και  $q$  είναι η κατανομή λεκτικών ριζών του αρχείου εισόδου και υπολογίστηκε από τον μαθηματικό τύπο 4.2.

### Αποτελέσματα δευτέρου αλγόριθμου

Ο δεύτερος αλγόριθμος επιλογής, όπως έχουμε προαναφέρει, επιλέγει έναν αριθμό προτάσεων με μοναδικό κριτήριο την διαφορετικότητα των διφώνων αλλά, με έναν τρόπο πιο περίπλοκο και πιο σύνθετο από τον πρώτο αλγόριθμο.

Ο αλγόριθμος αυτός επιλέγει τελικά 250 προτάσεις από ένα σύνολο 1.165.868 προτάσεων οι οποίες περιέχουν έστω και μία φορά όλα τα 1.114 δίφωνα που υπήρχαν στο σύνολο των αρχείων και 1.219 διαφορετικές λεκτικές ρίζες.

Η ακόλουθη κατανομή, εικόνα 4.5, μας παρουσιάζει τα δίφωνα που βρίσκονται στο αρχείο με τις επιλεγμένες προτάσεις (αρχείο εξόδου) σε συνάρτηση με τις πόσες φορές υπάρχει το κάθε δίφωνο στο αρχείο αυτό.

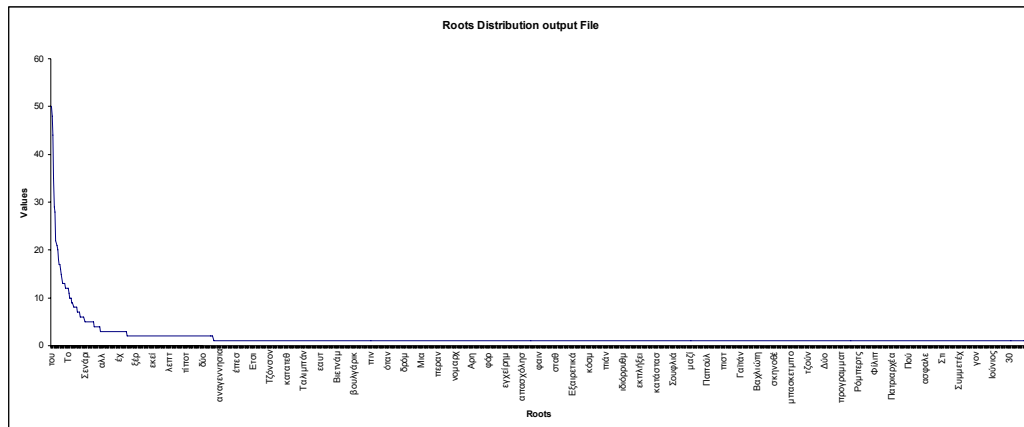


*Εικόνα 4.5 Κατανομή διφώνων στο αρχείο εξόδου*

Η μέση πληροφορία ανά διφώνιο είναι  $H(x) = 8.764$  bits/difone, ενώ η σχετική εντροπία για τις δύο κατανομές (εισόδου και εξόδου) είναι  $D(p \parallel q) = 0.360$  bits, όπου  $p$  είναι η κατανομή διφώνων του αρχείου εξόδου και  $q$  είναι η κατανομή διφώνων του αρχείου εισόδου.

Στην συνέχεια παραθέτουμε τα αντίστοιχα αποτελέσματα για τις κατανομές των λεκτικών ριζών.

Η ακόλουθη κατανομή, εικόνα 4.6, μας παρουσιάζει τις λεκτικές ρίζες που βρίσκονται στο αρχείο με τις επιλεγμένες προτάσεις (αρχείο εξόδου) σε συνάρτηση με τις πόσες φορές υπάρχει η κάθε ρίζα στο αρχείο αυτό.



**Εικόνα 4.6** Κατανομή λεκτικών ριζών στο αρχείο εξόδου

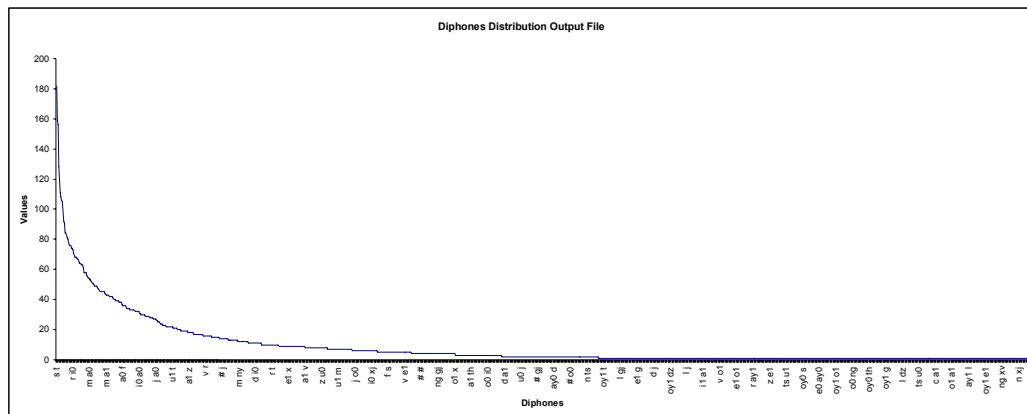
Η μέση πληροφορία ανά λεκτική ρίζα είναι  $H(x) = 9.419$  bits/root, ενώ η σχετική εντροπία για τις δύο κατανομές (εισόδου και εξόδου) είναι  $D(p \parallel q) = 2.808$  bits, όπου  $p$  είναι η κατανομή ριζών του αρχείου εξόδου και  $q$  είναι η κατανομή ριζών του αρχείου εισόδου.

### Αποτελέσματα τρίτου αλγόριθμου

Ο τρίτος αλγόριθμος επιλέγει έναν αριθμό προτάσεων με δύο κριτήρια, την διαφορετικότητα των διφώνων και την διαφορετικότητα των λεκτικών ριζών, δίνοντας βαρύτητα 50% και 50% αντίστοιχα σε κάθε κριτήριο.

Ο αλγόριθμος αυτός επιλέγει τελικά 277 προτάσεις από ένα σύνολο 1.165.868 προτάσεων οι οποίες περιέχουν έστω και μία φορά όλα τα 1.114 δίφωνα που υπήρχαν στο σύνολο των αρχείων και 1.456 διαφορετικές λεκτικές ρίζες.

Η ακόλουθη κατανομή, εικόνα 4.7, μας παρουσιάζει τα δίφωνα που βρίσκονται στο αρχείο με τις επιλεγμένες προτάσεις (αρχείο εξόδου) σε συνάρτηση με πόσες φορές υπάρχει το κάθε δίφωνο στο αρχείο αυτό.

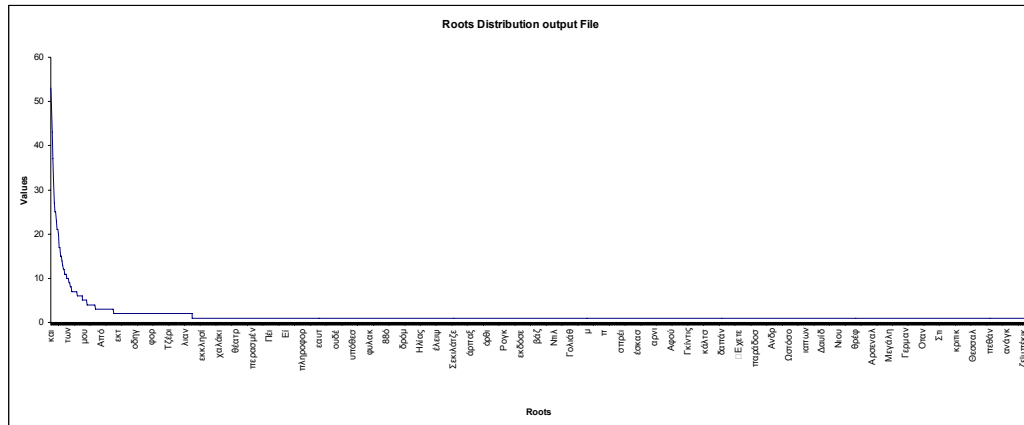


**Εικόνα 4.7** Κατανομή διφώνων στο αρχείο εξόδου

Η μέση πληροφορία ανά δίφωνο είναι  $H(x) = 8.681$  bits/difone, ενώ η σχετική εντροπία για τις δύο κατανομές (εισόδου και εξόδου) είναι  $D(p \parallel q) = 0.289$  bits, όπου  $p$  είναι η κατανομή διφώνων του αρχείου εξόδου και  $q$  είναι η κατανομή διφώνων του αρχείου εισόδου.

Στην συνέχεια παραθέτουμε τα αντίστοιχα αποτελέσματα για τις κατανομές των λεκτικών ριζών.

Η ακόλουθη κατανομή, εικόνα 4.8, μας παρουσιάζει τις λεκτικές ρίζες που βρίσκονται στο αρχείο με τις επιλεγμένες προτάσεις (αρχείο εξόδου) σε συνάρτηση με τις πόσες φορές υπάρχει η κάθε ρίζα στο αρχείο αυτό.



Εικόνα 4.8 Κατανομή λεκτικών ριζών στο αρχείο εξόδου

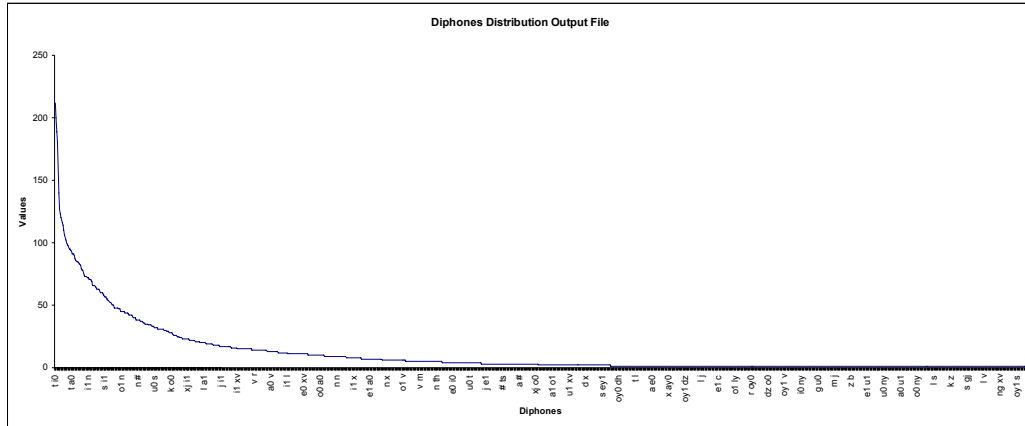
Η μέση πληροφορία ανά λεκτική ρίζα είναι  $H(x) = 9.695$  bits/root, ενώ η σχετική εντροπία για τις δύο κατανομές (εισόδου και εξόδου) είναι  $D(p \parallel q) = 2.460$  bits, όπου  $p$  είναι η κατανομή ριζών του αρχείου εξόδου και  $q$  είναι η κατανομή ριζών του αρχείου εισόδου.

### Αποτελέσματα τετάρτου αλγόριθμου

Ο τέταρτος αλγόριθμος επιλογής προτάσεων είναι ουσιαστικά μια παραλλαγή του τρίτου αλγορίθμου. Πιο συγκεκριμένα, όπως είχαμε προαναφέρει, ο τρίτος αλγόριθμος επιλέγει προτάσεις με κριτήριο την διαφορετικότητα των ριζών και τον διφώνων δίνοντας βαρύτητα 50% και 50% αντίστοιχα για κάθε κριτήριο. Ο τέταρτος αλγόριθμος, επιλέγει προτάσεις με μοναδικό κριτήριο την διαφορετικότητα και την ποικιλία σε λεκτικές ρίζες, δηλαδή βαρύτητα 100% στις λεκτικές ρίζες, και σταματάει μόλις επιλεγούν τόσες προτάσεις ώστε να περιέχονται σε αυτές όλα τα διαφορετικά δίφωνα της γλώσσας. Με λίγα λόγια, διατηρεί το πρωταρχικό μας στόχο, την επιλογή ενός συνόλου προτάσεων που να περιέχουν όλα τα διαφορετικά δίφωνα της ελληνικής γλώσσας, ενώ παράλληλα προσπαθεί να πετύχει το στόχο αυτό με το να επιλέξει προτάσεις που να περιέχουν όσο το δυνατόν περισσότερες διαφορετικές ρίζες λέξεων.

Ο αλγόριθμος αυτός επιλέγει τελικά 338 προτάσεις από ένα σύνολο 1.165.868 προτάσεων οι οποίες περιέχουν έστω και μία φορά όλα τα 1.114 δίφωνα που υπήρχαν στο σύνολο των αρχείων και 1.742 διαφορετικές λεκτικές ρίζες.

Η ακόλουθη κατανομή, εικόνα 4.9, μας παρουσιάζει τα δίφωνα που βρίσκονται στο αρχείο με τις επιλεγμένες προτάσεις (αρχείο εξόδου) σε συνάρτηση με τις πόσες φορές υπάρχει το κάθε δίφωνο στο αρχείο αυτό.

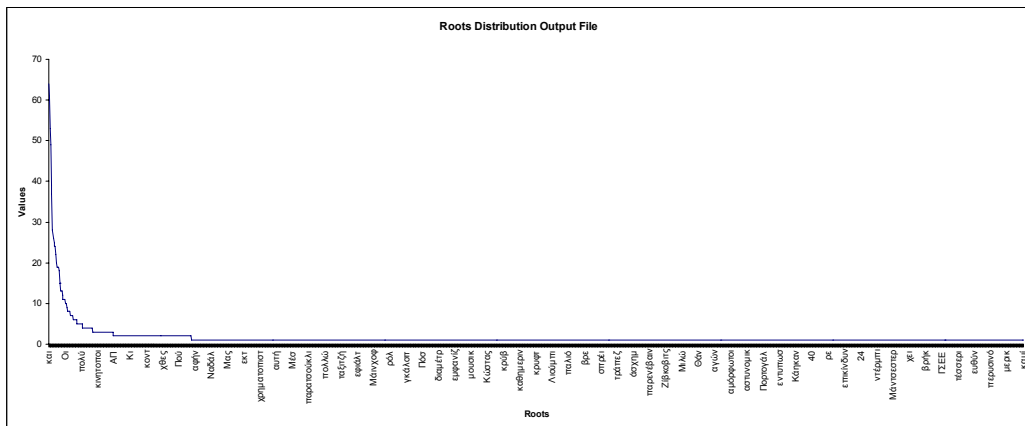


**Εικόνα 4.9** Κατανομή διφώνων στο αρχείο εξόδου

Η μέση πληροφορία ανά δίφωνο είναι  $H(x) = 8.623$  bits/difone, ενώ η σχετική εντροπία για τις δύο κατανομές (εισόδου και εξόδου) είναι  $D(p \parallel q) = 0.205$  bits, όπου  $p$  είναι η κατανομή διφώνων του αρχείου εξόδου και  $q$  είναι η κατανομή διφώνων του αρχείου εισόδου.

Στην συνέχεια παραθέτουμε τα αντίστοιχα αποτελέσματα για τις κατανομές των λεκτικών ριζών.

Η ακόλουθη κατανομή, εικόνα 4.10, μας παρουσιάζει τις λεκτικές ρίζες που βρίσκονται στο αρχείο με τις επιλεγμένες προτάσεις (αρχείο εξόδου) σε συνάρτηση με πόσες φορές υπάρχει η κάθε ρίζα στο αρχείο αυτό.



**Εικόνα 4.10** Κατανομή λεκτικών ριζών στο αρχείο εξόδου

Η μέση πληροφορία ανά λεκτική ρίζα είναι  $H(x) = 9.868$  bits/root, ενώ η σχετική εντροπία για τις δύο κατανομές (εισόδου και εξόδου) είναι  $D(p \parallel q) = 1.973$  bits, όπου  $p$  είναι η κατανομή ριζών του αρχείου εξόδου και  $q$  είναι η κατανομή ριζών του αρχείου εισόδου.

## 4.2 Σύγκριση των Αλγορίθμων Επιλογής Προτάσεων

Παρόλο, που παραθέσαμε αναλυτικά τα αποτελέσματα κάθε αλγορίθμου επιλογής προτάσεων, θα ήταν σκόπιμο να συγκρίνουμε τα αποτελέσματα αυτά ώστε να κατανοήσει καλύτερα ο αναγνώστης την λειτουργία των μεθόδων επιλογής προτάσεων. Στην συνέχεια, με τη χρήση πινάκων και γραφημάτων, θα συγκεντρώσουμε όλα τα παραπάνω αποτελέσματα ώστε να γίνει ευκολότερα κατανοητή η λειτουργία κάθε αλγορίθμου.

Πρωταρχικός στόχος κάθε αλγόριθμου είναι να επιλέξει έναν αριθμό προτάσεων που να περιέχει όλα τα δυνατά φωνήματα και δίφωνα της ελληνικής γλώσσας. Ο στόχος αυτός επιτυγχάνεται από όλους τους αλγορίθμους, όπως φαίνεται και στον πίνακα 4.3, και έτσι, το αρχείο εξόδου περιέχει τον ίδιο ακριβός αριθμό διαφορετικών διφώνων με το αρχείο εισόδου ανεξάρτητα από την μέθοδο επιλογής που χρησιμοποιήσαμε.

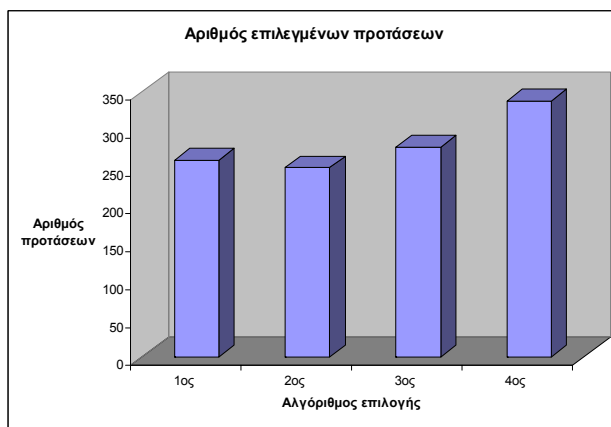
Αλγόριθμος επιλογής	Αριθμός διαφορετικών διφώνων – Αρχείο εξόδου	Αριθμός διαφορετικών διφώνων – Αρχείο εισόδου
1 <sup>ος</sup>	1114	1114
2 <sup>ος</sup>	1114	
3 <sup>ος</sup>	1114	
4 <sup>ος</sup>	1114	

**Πίνακας 4.3** Αριθμός διαφορετικών διφώνων σε κάθε αρχείο και μέθοδο επιλογής

Για να πετύχει το στόχο του ο κάθε αλγόριθμος επιλέγει έναν αριθμό τελικών προτάσεων. Ανάλογα με το ποια είναι τα κριτήρια επιλογής κάθε αλγορίθμου διαφέρει και ο αριθμός των επιλεγμένων προτάσεων. Έτσι, παρατηρούμε ότι για τους δύο πρώτους αλγορίθμους που το μοναδικό κριτήριο επιλογής προτάσεων είναι η διαφορετικότητα των διφώνων ο αριθμός των επιλεγμένων προτάσεων είναι μικρότερος από τον αριθμό των προτάσεων που επιλέγουν οι άλλοι δύο αλγόριθμοι που έχουν δύο κριτήρια επιλογής προτάσεων (δίφωνα – λεκτικές ρίζες). Όπως αυτό φαίνεται και στον πίνακα 4.4, ο 1<sup>ος</sup> αλγόριθμος επιλέγει 260 προτάσεις, ενώ ο 2<sup>ος</sup> που χρησιμοποιεί πιο εξελιγμένο τρόπο επιλογής προτάσεων επιλέγει 250 πετυχαίνοντας το ίδιο αποτέλεσμα. Ο 3<sup>ος</sup> αλγόριθμος που έχει ένα επιπλέον κριτήριο επιλογής προτάσεων, την διαφορετικότητα των λεκτικών ριζών, επιλέγει 277 προτάσεις ώστε να πετύχει το στόχο αυτό, ενώ τέλος, ο 4<sup>ος</sup> αλγόριθμος ο οποίος επιλέγει προτάσεις με μοναδικό κριτήριο την διαφορετικότητα των λεκτικών ριζών επιλέγει αρκετές παραπάνω προτάσεις σε σχέση με τους προηγούμενους αλγορίθμους και συγκεκριμένα επιλέγει 338 προτάσεις.

Αλγόριθμος επιλογής	Αριθμός επιλεγμένων προτάσεων – Αρχείο εξόδου	Αριθμός διαφορετικών προτάσεων – Αρχείο εισόδου
1 <sup>ος</sup>	260	1165868
2 <sup>ος</sup>	250	
3 <sup>ος</sup>	277	
4 <sup>ος</sup>	338	

**Πίνακας 4.4** Αριθμός επιλεγμένων προτάσεων με κάθε αλγόριθμο



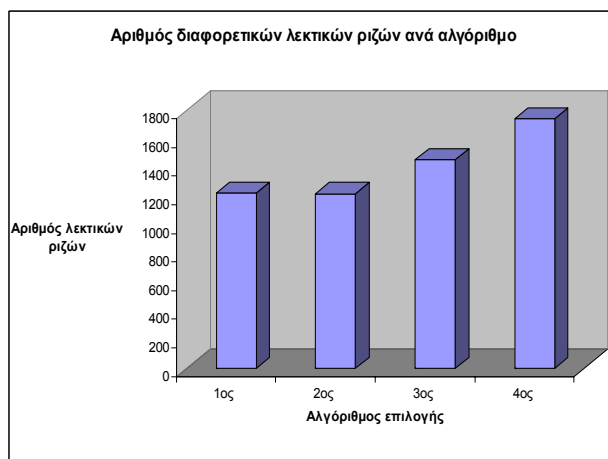
**Εικόνα 4.11** Αριθμός επιλεγμένων προτάσεων με κάθε αλγόριθμο



Όπως έχουμε προαναφέρει, οι δύο πρώτοι αλγόριθμοι επιλέγουν προτάσεις με μοναδικό κριτήριο επιλογής την διαφορετικότητα των διφώνων, ενώ ο 3<sup>ος</sup> και ο 4<sup>ος</sup> αλγόριθμος επιλέγουν προτάσεις με κριτήρια την διαφορετικότητα των διφώνων και την διαφορετικότητα των λεκτικών ριζών. Όπως είναι αναμενόμενο, οι προτάσεις που επιλέγηκαν από τους δύο πρώτους αλγορίθμους θα πρέπει να περιέχουν περίπου ίδιο αριθμό διαφορετικών ριζών στο σύνολο τους, ενώ οι προτάσεις που επιλέγηκαν από τους αλγορίθμους 3 και 4 θα πρέπει να περιέχουν συνολικά αρκετά περισσότερες λεκτικές ρίζες. Παράλληλα, ο 4<sup>ος</sup> αλγόριθμος επιλογής λόγω του ότι δίνει μέγιστη βαρύτητα στην επιλογή διαφορετικών ριζών, θα πρέπει οι προτάσεις που επιλέγει να περιέχουν συνολικά περισσότερες διαφορετικές ρίζες από κάθε άλλο αλγόριθμο. Τα στοιχεία αυτά επιβεβαιώνονται απόλυτα από τον πίνακα 4.5 και από το γράφημα 4.12.

Αλγόριθμος επιλογής	Αριθμός διαφορετικών ριζών – Αρχείο εξόδου	Αριθμός διαφορετικών ριζών – Αρχείο εισόδου
1 <sup>ος</sup>	1221	88590
2 <sup>ος</sup>	1219	
3 <sup>ος</sup>	1456	
4 <sup>ος</sup>	1742	

**Πίνακας 4.5** Αριθμός διαφορετικών λεκτικών ριζών με κάθε αλγόριθμο



**Εικόνα 4.12** Αριθμός διαφορετικών λεκτικών ριζών με κάθε αλγόριθμο

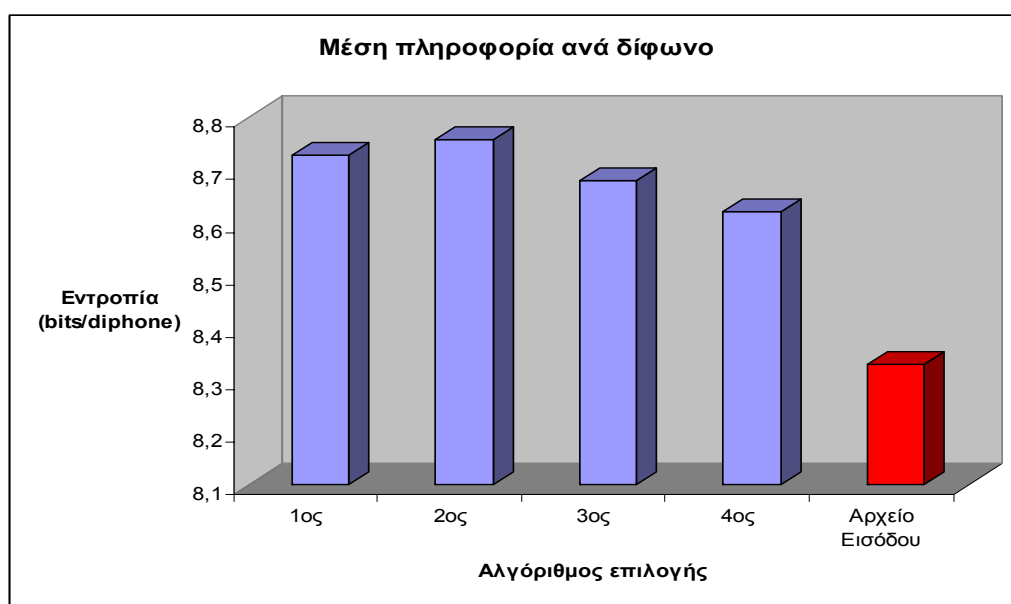
Για να μπορέσουμε να συγκρίνουμε με μεγαλύτερη ακρίβεια το αποτέλεσμα κάθε μεθόδου επιλογής προτάσεων και να βρούμε την μετρική απόσταση ανάμεσα στο αρχείο εισόδου και στο αρχείο με τις επιλεγμένες προτάσεις για κάθε μέθοδο επιλογής, υπολογίσαμε την εντροπία και την σχετική εντροπία ανάμεσα στις κατανομές των διφώνων και των λεκτικών ριζών για κάθε μέθοδο επιλογής.

Οι παρακάτω πίνακες και οι γραφικές παραστάσεις θα μας επιτρέψουν να δούμε τα στοιχεία αυτά συγκεντρωμένα ώστε να συγκρίνουμε καλύτερα τις διάφορες μεθόδους επιλογής προτάσεων.

### Μέση Πληροφορία ανά Δίφωνο

Αλγόριθμος επιλογής	Εντροπία διφώνων – Αρχείο εξόδου	Εντροπία διφώνων – Αρχείο εισόδου
1 <sup>ος</sup>	8,73	8,33
2 <sup>ος</sup>	8,76	
3 <sup>ος</sup>	8,68	
4 <sup>ος</sup>	8,62	

*Πίνακας 4.6 Μέση πληροφορία ανά δίφωνο για κάθε αλγόριθμο*



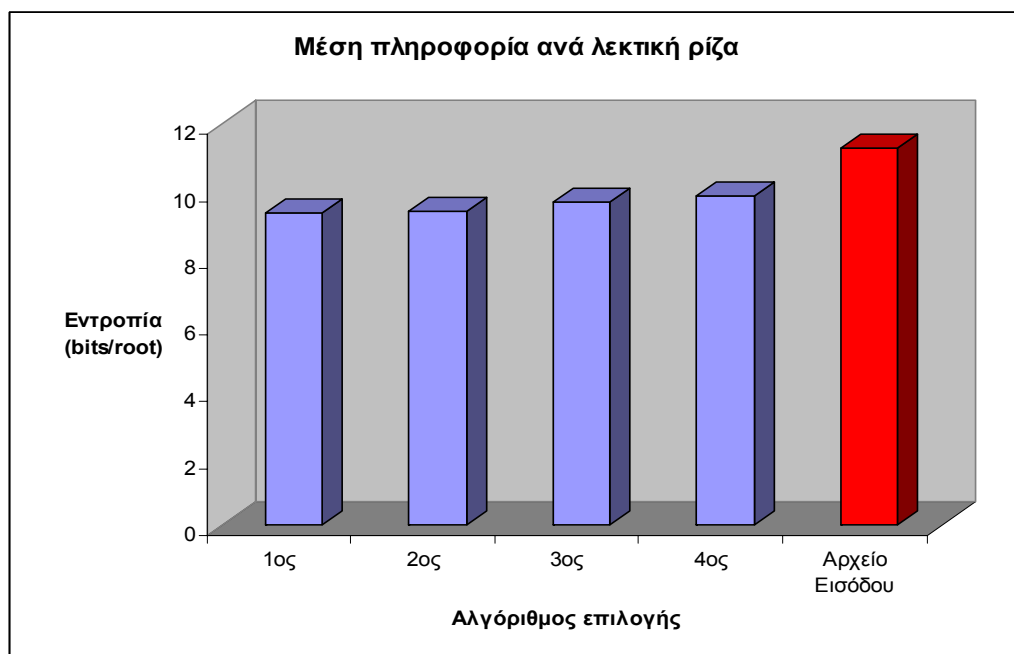
*Εικόνα 4.13 Μέση πληροφορία ανά δίφωνο για κάθε αλγόριθμο*

Όπως παρατηρούμε από τον πίνακα 4.6 και την εικόνα 4.13, η μέση πληροφορία ανά δίφωνο είναι μικρότερη στο αρχείο με τις επιλεγμένες προτάσεις που προκύπτει από την 4<sup>η</sup> μέθοδο επιλογής και άρα πιο κοντά στην μέση πληροφορία ανά δίφωνο για το αρχείο εισόδου.

### Μέση Πληροφορία ανά Λεκτική Ρίζα

Αλγόριθμος επιλογής	Εντροπία λεκτικών ριζών – Αρχείο εξόδου	Εντροπία λεκτικών ριζών – Αρχείο εισόδου
1 <sup>ος</sup>	9,37	11,28
2 <sup>ος</sup>	9,41	
3 <sup>ος</sup>	9,69	
4 <sup>ος</sup>	9,86	

**Πίνακας 4.7** Μέση πληροφορία ανά λεκτική ρίζα για κάθε αλγόριθμο



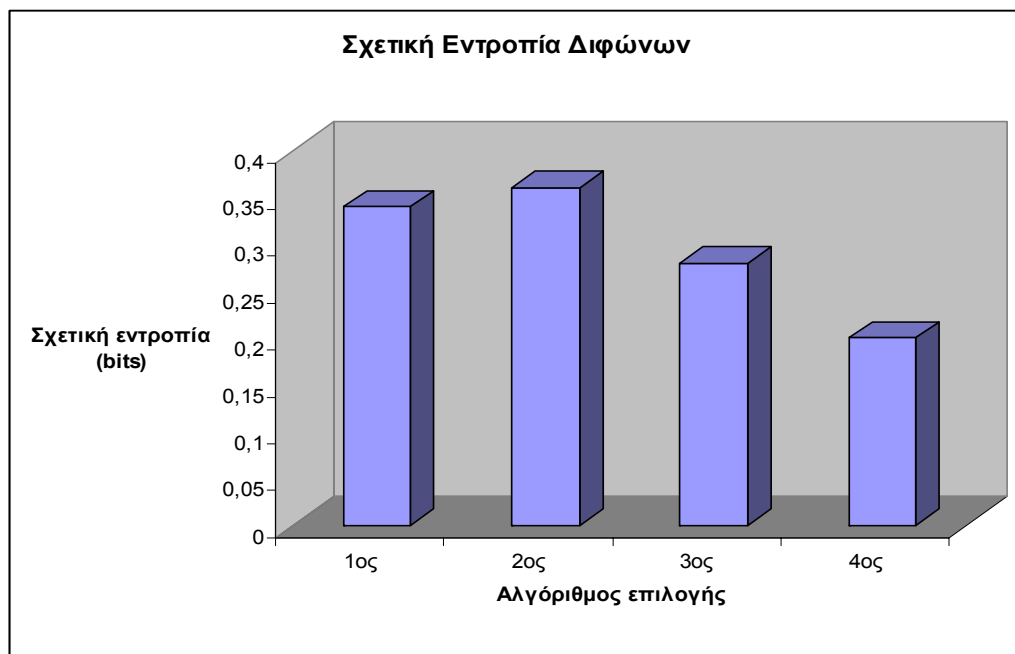
**Εικόνα 4.14** Μέση πληροφορία ανά λεκτική ρίζα για κάθε αλγόριθμο

Όπως φαίνεται και στον πίνακα 4.7 και στην εικόνα 4.14, η μέση πληροφορία ανά λεκτική ρίζα είναι μεγαλύτερη στο αρχείο με τις επιλεγμένες προτάσεις που προκύπτει από την 4<sup>η</sup> μέθοδο επιλογής και άρα πιο κοντά στην μέση πληροφορία ανά λεκτική ρίζα για το αρχείο εισόδου.

### Σχετική Εντροπία Ανάμεσα στις Κατανομές των Διφώνων

Αλγόριθμος επιλογής	Σχετική Εντροπία Διφώνων (bits)
1 <sup>ος</sup>	0,34
2 <sup>ος</sup>	0,36
3 <sup>ος</sup>	0,28
4 <sup>ος</sup>	0,20

**Πίνακας 4.8** Σχετική εντροπία για τις κατανομές διφώνων για κάθε αλγόριθμο



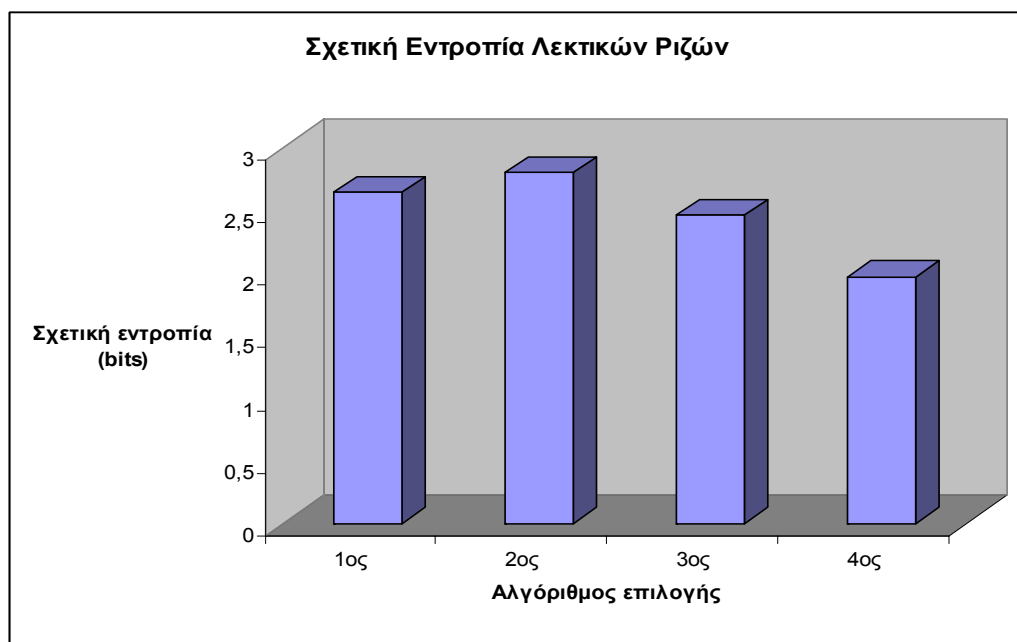
**Εικόνα 4.15** Σχετική εντροπία για τις κατανομές διφώνων για κάθε αλγόριθμο

Όπως παρατηρούμε και από την σχετική εντροπία  $D(p \parallel q)$ , όπου  $p$  είναι η κατανομή διφώνων του αρχείου εξόδου και  $q$  είναι η κατανομή διφώνων του αρχείου εισόδου, για κάθε μέθοδο επιλογής προτάσεων, η 4<sup>η</sup> μέθοδος επιλέγει προτάσεις που η κατανομή των διφώνων τους είναι πιο κοντά στην κατανομή των διφώνων στο αρχείο εισόδου.

### Σχετική Εντροπία Ανάμεσα στις Κατανομές των Λεκτικών Ριζών

Αλγόριθμος επιλογής	Σχετική Εντροπία Λεκτικών Ριζών (bits)
1 <sup>ος</sup>	2,64
2 <sup>ος</sup>	2,80
3 <sup>ος</sup>	2,46
4 <sup>ος</sup>	1,97

**Πίνακας 4.9** Σχετική εντροπία για τις κατανομές λεκτικών ριζών για κάθε αλγόριθμο



**Εικόνα 4.16** Σχετική εντροπία για τις κατανομές λεκτικών ριζών για κάθε αλγόριθμο

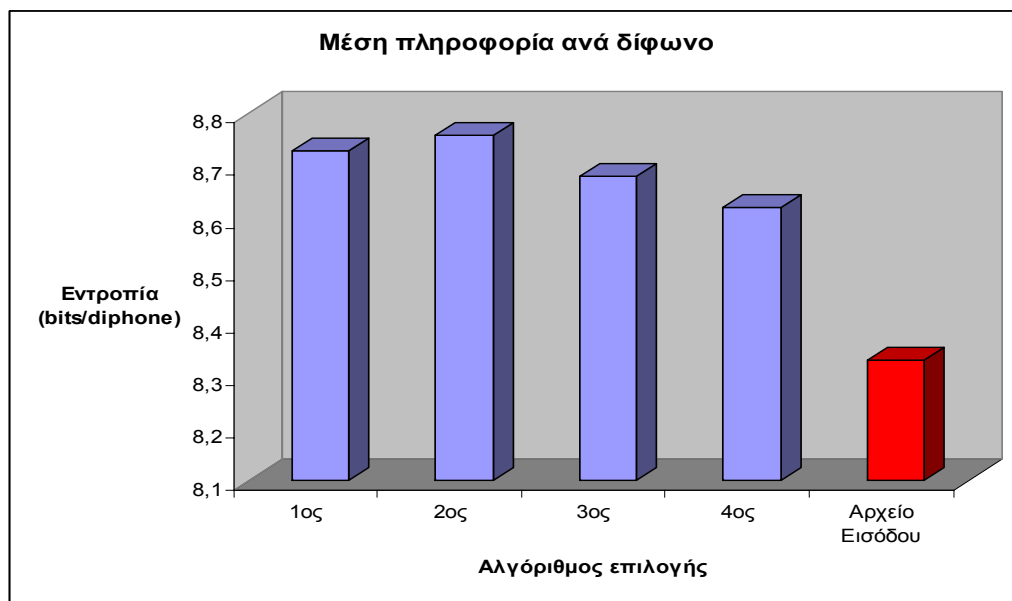
Όπως παρατηρούμε και από την σχετική εντροπία  $D(p \parallel q)$ , όπου  $p$  είναι η κατανομή λεκτικών ριζών του αρχείου εξόδου και  $q$  είναι η κατανομή λεκτικών ριζών του αρχείου εισόδου, για κάθε μέθοδο επιλογής προτάσεων, η 4<sup>η</sup> μέθοδος επιλέγει προτάσεις που η κατανομή των λεκτικών ριζών τους είναι πιο κοντά στην κατανομή των λεκτικών ριζών στο αρχείο εισόδου.

Από τα παραπάνω στατιστικά στοιχεία και κυρίως από τις εντροπίες και σχετικές εντροπίες παρατηρούμε ότι η 4<sup>η</sup> μέθοδος επιλογής προτάσεων καταφέρνει να επιλέξει ένα σύνολο προτάσεων που οι κατανομές διφώνων του και λεκτικών ριζών του «μοιάζουν» με τις κατανομές διφώνων και λεκτικών ριζών του αρχείου εισόδου περισσότερο από κάθε άλλη μέθοδο επιλογής προτάσεων. Αρά, μπορούμε να πούμε με βεβαιότητα ότι η 4<sup>η</sup> μέθοδος επιλέγει ένα σύνολο προτάσεων που περιέχουν περισσότερη λεκτική πληροφορία από το αποτέλεσμα κάθε άλλης μεθόδου. Το στόχο αυτό τον επιτυγχάνει η μέθοδος αυτή επιλέγοντας περισσότερες προτάσεις από κάθε άλλη μέθοδο. Για να μπορέσουμε να πειστούμε για την αξιοπιστία της μεθόδου αυτής, επαναλάβουμε τα παραπάνω πειράματα με τον υπολογισμό των εντροπίων και των σχετικών εντροπίων αναγκάζοντας όλες τις μεθόδους να επιλέξουν έναν ελάχιστο αριθμό προτάσεων και συγκεκριμένα το μικρότερο δυνατό, δηλαδή 250 προτάσεις. Στην συνέχεια θα παραθέσουμε αναλυτικά τα στοιχεία αυτά σε μορφή πινάκων και διαγραμμάτων ενώ τέλος θα τα αναλύσουμε.

### Μέση Πληροφορία ανά Δίφωνο

Αλγόριθμος επιλογής	Εντροπία διφώνων – Αρχείο εξόδου	Εντροπία διφώνων – Αρχείο εισόδου
1 <sup>ος</sup>	8,72	8,33
2 <sup>ος</sup>	8,76	
3 <sup>ος</sup>	8,68	
4 <sup>ος</sup>	8,62	

**Πίνακας 4.10** Μέση πληροφορία ανά δίφωνο για κάθε αλγόριθμο



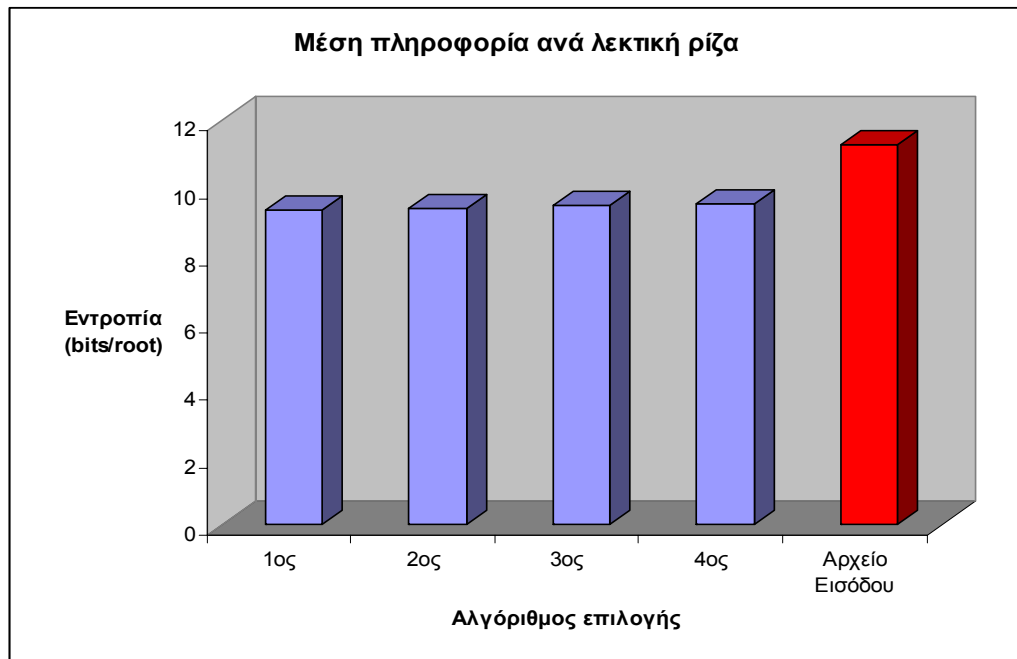
*Εικόνα 4.17 Μέση πληροφορία ανά δίφωνο για κάθε αλγόριθμο*

Όπως παρατηρούμε από τον πίνακα 4.10 και την εικόνα 4.17, η μέση πληροφορία ανά δίφωνο είναι ακόμα μικρότερη στο αρχείο με τις επιλεγμένες προτάσεις που προκύπτει από την 4<sup>η</sup> μέθοδο επιλογής και άρα πιο κοντά στην μέση πληροφορία ανά δίφωνο για το αρχείο εισόδου. Σημειώνουμε ότι δεν παρατηρήσαμε αισθητή διαφορά στα αποτελέσματα παρόλο που περιορίσαμε τον αριθμό των τελικών προτάσεων.

### Μέση Πληροφορία ανά Λεκτική Ρίζα

Αλγόριθμος επιλογής	Εντροπία λεκτικών ριζών – Αρχείο εξόδου	Εντροπία λεκτικών ριζών – Αρχείο εισόδου
1 <sup>ος</sup>	9,34	11,28
2 <sup>ος</sup>	9,41	
3 <sup>ος</sup>	9,51	
4 <sup>ος</sup>	9,53	

*Πίνακας 4.11 Μέση πληροφορία ανά λεκτική ρίζα για κάθε αλγόριθμο*



**Εικόνα 4.18** Μέση πληροφορία ανά λεκτική ρίζα για κάθε αλγόριθμο

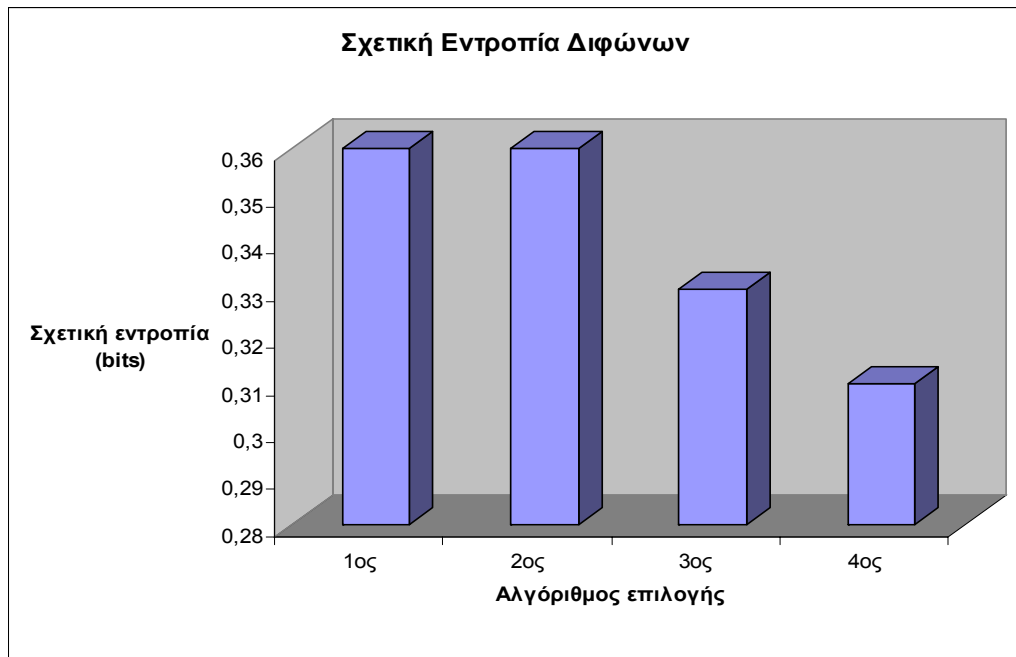
Όπως φαίνεται και στον πίνακα 4.11 και στην εικόνα 4.18, η μέση πληροφορία ανά λεκτική ρίζα είναι οριακά μεγαλύτερη στο αρχείο με τις επιλεγμένες προτάσεις που προκύπτει από την 4<sup>η</sup> μέθοδο επιλογής και άρα πιο κοντά στην μέση πληροφορία ανά λεκτική ρίζα για το αρχείο εισόδου. Στην περίπτωση αυτή παρατηρήθηκε αλλαγή των αποτελεσμάτων λόγω των λιγότερων επιλεγμένων προτάσεων αλλά ακόμα και τώρα ο 4<sup>ος</sup> αλγόριθμος επιλογής είχε την «κοντινότερη» εντροπία με το αρχείο εισόδου σε σχέση με τους άλλους τρεις αλγορίθμους.

### Σχετική Εντροπία Ανάμεσα στις Κατανομές των Διφώνων

Αλγόριθμος επιλογής	Σχετική Εντροπία Διφώνων (bits)
1 <sup>ος</sup>	0,36
2 <sup>ος</sup>	0,36
3 <sup>ος</sup>	0,33
4 <sup>ος</sup>	0,31

**Πίνακας 4.12** Σχετική εντροπία για τις κατανομές διφώνων για κάθε αλγόριθμο





**Εικόνα 4.19** Σχετική εντροπία για τις κατανομές διφώνων για κάθε αλγόριθμο

Όπως παρατηρούμε και από την σχετική εντροπία  $D(p \parallel q)$ , όπου  $p$  είναι η κατανομή διφώνων του αρχείου εξόδου και  $q$  είναι η κατανομή διφώνων του αρχείου εισόδου, για κάθε μέθοδο επιλογής προτάσεων, η 4<sup>η</sup> μέθοδος επιλέγει προτάσεις που η κατανομή των διφώνων τους είναι πιο κοντά στην κατανομή των διφώνων στο αρχείο εισόδου. Όπως συμπεραίνουμε και σε αυτήν την περίπτωση, παρόλο τον περιορισμό των προτάσεων και ενώ αυξήθηκε η μετρική απόσταση ανάμεσα στις κατανομές, η 4<sup>η</sup> μέθοδος επιλογής υπερσχύει έναντι των άλλων μεθόδων.

### Σχετική Εντροπία Ανάμεσα στις Κατανομές των Λεκτικών Ριζών

Αλγόριθμος επιλογής	Σχετική Εντροπία Λεκτικών Ριζών (bits)
1 <sup>ος</sup>	2,68
2 <sup>ος</sup>	2,80
3 <sup>ος</sup>	2,66
4 <sup>ος</sup>	2,42

**Πίνακας 4.13** Σχετική εντροπία για τις κατανομές λεκτικών ριζών για κάθε αλγόριθμο



*Εικόνα 4.20 Σχετική εντροπία για τις κατανομές λεκτικών ριζών για κάθε αλγόριθμο*

Όπως παρατηρούμε και από την σχετική εντροπία  $D(p \parallel q)$ , όπου  $p$  είναι η κατανομή λεκτικών ριζών του αρχείου εξόδου και  $q$  είναι η κατανομή λεκτικών ριζών του αρχείου εισόδου, για κάθε μέθοδο επιλογής προτάσεων, η 4<sup>η</sup> μέθοδος καταφέρνει να επιλέξει προτάσεις που η κατανομή των λεκτικών ριζών τους είναι πιο κοντά στην κατανομή των λεκτικών ριζών στο αρχείο εισόδου παρόλο τον περιορισμό στις προτάσεις που μπορεί να επιλέξει.

Έτσι, παρατηρώντας τα παραπάνω αποτελέσματα σχετικά με τις εντροπίες και τις σχετικές εντροπίες για τις διάφορες μεθόδους επιλογής προτάσεων, με περιορισμό στην επιλογή αυτών, παρατηρούμε ότι παρόλο που η μετρική απόσταση ανάμεσα στις κατανομές των αρχείων εξόδου και τις κατανομές του αρχείου εισόδου αυξήθηκε, η 4<sup>η</sup> μέθοδος επιλογής προτάσεων εξακολουθεί να υπερισχύει έναντι των τριών άλλων στο θέμα τις επιλογής προτάσεων που να περιέχουν λεκτική πληροφορία πολύ κοντά στην λεκτική πληροφορία του αρχείου εξόδου. Το συμπέρασμα αυτό πιστοποιεί ότι η μέθοδος αυτή υπερισχύει των άλλων μεθόδων ανεξαρτήτως του αριθμού των προτάσεων που επιλέγονται. Με άλλα λόγια, οι προτάσεις που επιλέγονται από τον 4<sup>ο</sup> αλγόριθμο είναι πλουσιότερες σε λεκτική πληροφορία από τις προτάσεις που επιλέγονται με κάθε άλλη μέθοδο.

---

# Τελικά συμπεράσματα

## 5.1 Επίλογος

Το σύστημα το οποίο περιγράψαμε στα προηγούμενα κεφάλαια δεν μπορεί να χαρακτηριστεί σε καμία περίπτωση τέλειο. Παρά τις ατέλειες και τα μειονεκτήματα του καταφέρνει να επιλέξει επιτυχώς ένα σύνολο προτάσεων, από τις οποίες θα μπορούσαμε να ηχογραφήσουμε στο μέλλον, το σώμα φωνής και έπειτα με την επεξεργασία και την τμηματοποίηση αυτού να συνθέσουμε κάθε νέα λέξη και πρόταση. Για την επίτευξη του στόχου αυτού, υλοποιήσαμε τέσσερις διαφορετικούς αλγορίθμους επιλογής προτάσεων και συγκρίναμε τα αποτελέσματα κάθε αλγορίθμου παραθέτοντας τα πλεονεκτήματα και τα μειονεκτήματα του κάθε αλγορίθμου. Από την σύγκριση αυτή, καταλήξαμε ότι ο τέταρτος αλγόριθμος επιλογής προτάσεων επιλέγει προτάσεις με περισσότερη λεκτική πληροφορία από κάθε άλλο αλγόριθμο.

Όπως είχε πει και ο Bill Toomey (χρυσός Ολυμπιονίκης και έξι φορές παγκόσμιος πρωταθλητής στο δέκαθλο), «Είναι πάντα δυνατό να βελτιώνεσαι». Στην περίπτωση μας ο όρος «βελτίωση» έχει πολλά πρόσωπα, όπως: η εύρεση ενός πλουσιότερου σε λεκτικές πληροφορίες σώματος κειμένου, ώστε να μπορούμε να αντλήσουμε από αυτό περισσότερη γλωσσική πληροφορία, η εύρεση ενός λεξικού της ελληνικής γλώσσας που να περιέχει περισσότερες ελληνικές λέξεις, καθώς επίσης και ονόματα ανθρώπων, πόλεων, κρατών και ημερομηνιών ώστε να μπορούμε να πετύχουμε ακριβέστερη ανίχνευση συντομογραφιών με αποτέλεσμα καλύτερο διαχωρισμό προτάσεων και τέλος στην ανάπτυξη καλύτερων αλγορίθμων στον τομέα

της επιλογής προτάσεων, ώστε να επιλέξουμε προτάσεις με την μέγιστη δυνατή λεκτική πληροφορία.

Τέλος, ως μελλοντική έρευνα, θα μπορούσαμε να υλοποιήσουμε ένα σύστημα το οποίο θα δεχόταν ως είσοδο τις προτάσεις που επιλέξαμε και θα συνέθετε φωνή, γεγονός που θα μας επέτρεπε να κατανοήσουμε και να δούμε τα πραγματικά αποτελέσματα και πλεονεκτήματα της βέλτιστης επιλογής προτάσεων για ένα σύστημα σύνθεσης φωνής.

---

## Βιβλιογραφία

- [1] A. W. Black, *Perfect Synthesis for All the People All of the Time*, IEEE TTS Workshop 2002, 2002.
- [2] J. Allen, S. Hunnicutt, and D. Klatt, *The MITalk System*, Cambridge University Press, Cambridge, 1987.
- [3] A. W. Black and P. Taylor, *CHATR: A generic speech synthesis system*, COLING94 (Kyoto, Japan), vol. 2, 1994, pp. 983 – 986.
- [4] R. Sproat, *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.
- [5] K. Ng, *Survey of Data-Driven Approaches to Speech Synthesis*, Area survey, Spoken Language Systems Group, Massachusetts Institute of Technology, Cambridge, MA, 1998.
- [6] N. Campbell and A. W. Black, *Progress in Speech Synthesis*, ch. Prosody and the Selection of Source Units for Concatenative Speech Synthesis, Springer Verlag, New York, 1995.
- [7] K. Lenzo and A. Black, *Customized Synthesis: Blending and Tiering*, in AVIOS2002, San Jose, CA, 2002.

- [8] A. J. Hunt and A. W. Black, *Unit Selection in a Concatenative Speech Synthesis System Using a Large Speech Database*, ICASSP '96 (Atlanta, GA), vol. 1, May 1996, pp. 373–376.
- [9] A. Black, P. Taylor, and R. Caley, *The Festival speech synthesis system*, Technical report hcrc/tr-83, Human Computer Communication Research Centre, Edinburgh University, 1997.
- [10] A. Black, *Chatr version 0.8: A generic Speech Synthesizer*, System documentation, ATR Interpreting Telecommunications Laboratories, Kyoto, Japan, March 1996.
- [11] T. Christiansen, N. Torkington, O'Reilly - *Perl Cookbook 2nd Edition*, August 2003.
- [12] <http://humanities.uchicago.edu/faculty/goldsmith/Linguistica2000/>.
- [13] M. Beutnagel, A. Conkie, J. Schroeter, Y. Stylianou, and A. Syrdal, *The AT&T Next-Gen TTS system*, Joint Meeting of ASA, EAA, and DAGA (Berlin, Germany), March 1999.
- [14] A. Black and K. Lenzo, *Limited domain Synthesis*, in ICSLP2000, Beijing, China, 2000.
- [15] C. Vosnidis, *Robust Speech Synthesis*, Technical University of Crete, Chania, June 2004.
- [16] T. Dutoit, *An introduction to Text-to-Speech Synthesis*, Kluwer Academic Publishers, 1997.
- [17] R. Spoak and J. Olive, *An Approach to Text-to-Speech Synthesis in Speech Coding and Synthesis*, (W. Kleijn and K. Paliwal, eds), pp. 611-633, Amsterdam, Holland: Elsevier Science, 1995.