ΕΘΝΙΚΟ &ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ
ΗΛΕΚΤΡΟΝΙΚΟΥ ΑΥΤΟΜΑΤΙΣΜΟΥ

Νικολάου Λύρα (Α.Μ. 97514)

Το Λεξιλογίο στα Συστήματα Αναγνώρισης Ομιλίας

Εργασία στο μάθημα: Επικοινωνία με Ομιλία
Διδάσκων: Γεώργιος Κουρουπέτρογλου

Αθήνα 1999

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας εργασίας είναι η ανάλυση και η παρουσίαση θεμάτων που σχετίζονται με το λεξιλόγιο το οποίο χρησιμοποιείται στα συστήματα αναγνώρισης ομιλίας. Αρχικά περιγράφονται οι επικρατέστερες τεχνολογικές μέθοδοι που χρησιμοποιούνται για την αναγνώριση λέξεων τόσο σε μεγάλα όσο και σε μικρά λεξιλόγια. Αυτές είναι, τα ίχνη, τα κρυφά μοντέλα Markov (HMM), οι ακολουθίες φωνημάτων και οι ακολουθίες υπολέξεων. Η αναφορά στις μεθόδους αυτές περιλαμβάνει ανάλυση του τρόπου λειτουργίας κάθε μεθόδου, επισήμανση των πλεονεκτημάτων και των αδυναμιών που εμφανίζουν καθώς και μία σύντομη ιστορική αναδρομή. Στην συνέχεια εξετάζονται τρόποι και μέθοδοι επέκτασης του λεξιλογίου από τους κατασκευαστές και προβάλλεται το θέμα της αποτελεσματικής αναζήτησης σε συστήματα μεγάλου λεξιλογίου με την παρουσίαση σχετικών αλγόριθμων αναζήτησης, όπως ο αλγόριθμος αναζήτησης δέσμης, ο αλγόριθμος αναζήτησης στοίβας και ο αλγόριθμος γρήγορου ταιριάσματος. Ακολουθεί η εστίαση σε άλλες πλευρές της έννοιας της λέξης, που είναι ο προσδιοριστής λέξης, η μετάφραση και η έννοια της σημασίας σε επίπεδο λέξης (λεξιλογική σημασιολογία). Ακόμη εισάγεται η έννοια του ολικού λεξιλογίου. Ακολούθως παρουσιάζονται μέθοδοι σχεδίασης λεξιλογίου. Αυτές είναι: η υλοποίηση λεξικών από τους κατασκευαστές, η δημιουργία εφαρμογών σχεδιασμού λεξιλογίου σε επίπεδο λέξης, η δημιουργία εφαρμογών σχεδιασμού λεξιλογίου σε επίπεδο υπολέξης, η δημιουργία λεξιλογίου από τελικούς χρήστες, η αυτόματη εξαγωγή λεξιλογίου από το λεξιλόγιο. Στην συνέχεια γίνεται λόγος για ειδικά θέματα λεξιλογίου όπως το ενεργό λεξιλόγιο, οι συγχεόμενες λέξεις, το αλφάβητο, και οι αριθμοί. Η ολοκλήρωση των θεμάτων που ασχολούνται με το λεξιλόγιο γίνεται με μία αποτίμηση των απαιτήσεων σε λεξιλόγιο σε μία εφαρμογή. Ακολουθεί παρουσίαση θεμάτων που σχετίζονται με το λεξιλόγιο όπως είναι η έννοια της γραμματικής, η έννοια της περιπλοκής της γλώσσας, η ανάλυση δύο μεθόδων συντακτικής ανάλυσης της γλώσσας (Top-down και Bottom-up ανάλυση) και τέλος παρουσιάζονται και άλλα μοντέλα γραμματικής που έχουν υιοθετηθεί στα συστήματα αναγνώρισης ομιλίας. Η εργασία καταλήγει με μία αναφορά στα συστήματα αναγνώρισης ομιλίας, σε εκείνα που διαδραμάτισαν κάποιο σημαντικό ρόλο στο παρελθόν αλλά και σε αυτά που πλαισιώνουν το παρόν.

ΕΙΣΑΓΩΓΗ

Σε γενικές γραμμές τα σύγχρονα συστήματα αναγνώρισης ομιλίας ανήκουν σε μία από τις τρεις γενικές κατηγορίες:

1. Αυτά με μικρό λεξιλόγιο (10 έως 100 λέξεις).
2. Αυτά στα οποία οι λέξεις ομιλούνται απομονωμένα η μία από την άλλη (το λεξιλόγιο εκτείνεται μέχρι και τις 10000 λέξεις).
3. Αυτά που μπορούν να δεχθούν συνεχή ομιλία (το λεξιλόγιο εκτείνεται από 1000 έως 5000 λέξεις)

Ένας από τους διαρκείς στόχους της τεχνολογίας αναγνώρισης ομιλίας είναι να δωθεί η δυνατότητα στους χρήστες αυτών των συστημάτων να εκφράσουν αυτό που θέλουν χρησιμοποιώντας οποιεσδήποτε λέξεις χρειάζονται για να υλοποιήσουν τη διαδικασία αυτή. Το γεγονός αυτό πολύ συχνά μεταφράζεται στην ανάγκη ύπαρξης ενός πολύ μεγάλου και εκτενούς λεξιλογίου. Στην πράξη το λεξιλόγιο που απαιτείται κατά τις διάφορες εφαρμογές ποικίλει από μερικές λέξεις έως και πολλές εκατοντάδες λέξεων. Πολλές από τις εφαρμογές απαιτούν πενήντα ή και λιγότερες λέξεις, όμως οι κατασκευαστές αντιλαμβάνονται ότι με το να δημιουργήσουν συστήματα ικανά για επεξεργασία μεγάλων σε έκταση λεξιλογίων θα επεκτείνουν αυτόματα το πεδίο και τους τύπους των εφαρμογών, που σχετίζονται με τον τομέα της αναγνώρισης ομιλίας.

Αν και τα φωνήματα είναι τα βασικά τμήματα ήχου μίας γλώσσας τα περισσότερα εμπορικά συστήματα αναγνώρισης ομιλίας χρησιμοποιούν τις λέξεις, παρά τα φωνήματα, σαν την βασική μονάδα αναγνώρισης ομιλίας. Το τμήμα το οποίο εστιάζει στην τεχνολογία περιγράφει τις επικρατέστερες τεχνολογικές μεθόδους που χρησιμοποιούνται για την αναγνώριση λέξεων τόσο σε μεγάλα όσο και σε μικρά λεξιλόγια:

- Ίχνη
- Κρυφά Μοντέλα Markov (HMM)
- Ακολουθίες Φωνημάτων
- Ακολουθίες Υπολέξεων

Κάθε μέθοδος περιγράφεται από την άποψη του πως ορίζει την έννοια της ***λέξης*** και στο πως αντιμετωπίζει την μεταβλητότητα στον τρόπο με τον οποίο η λέξη ομιλείται. Στόχος, στην συνέχεια είναι, η επέκταση στην εξέταση τεχνολογικών θεμάτων που εμφανίζονται στον σχεδιασμό μεγάλων λεξιλογίων.

Στο τμήμα που εστιάζει στις εφαρμογές η έννοια της λέξης επανεξετάζεται από την πλευρά της εφαρμογής, έχοντας σαν πηγές σύγκρισης είτε πρότυπα λεξικά είτε δείγματα από την

καθημερινή χρήση των λέξεων. Ακόμη περιγράφονται επαναληπτικές μέθοδοι που χρησιμοποιούνται στις εφαρμογές ανάπτυξης λεξιλογίου και ακολουθεί η εξέταση του προβλήματος χειρισμού ακουστικά ομοίων λέξεων (λέγονται και ***συγχεόμενες*** λέξεις). Ακολουθεί μία επισκόπηση των απαιτήσεων σε λεξιλόγιο στις διάφορες εφαρμογές . Αφού ολοκληρωθεί η παρουσίαση και η ανάλυση των θεμάτων που αφορούν το λεξιλόγιο γίνεται αναφορά στην γραμματική της γλώσσας στις μεθόδους ανάλυσης που χρησιμοποιούνται καθώς και σε συστήματα γραμματικής που περιλαμβάνονται στα συστήματα αναγνώρισης ομιλίας. Η παρούσα εργασία καταλήγει με μία επισκόπηση των σημαντικότερων συστημάτων αναγνώρισης ομιλίας που έχουν παρουσιαστεί έως σήμερα.

ΤΙ ΕΙΝΑΙ ΜΙΑ ΛΕΞΗ;

Το να οριστεί τι είναι λέξη μπορεί να θεωρηθεί ως μη αναγκαίο. Για ένα ανθρώπινο ον μία ομιλούμενη λέξη είναι μία ακολουθία από ήχους συνδεδεμένους κατά τέτοιο τρόπο ώστε να αποδίδουν κάποια έννοια. Σε ένα μεγάλο βαθμό, μία συγκεκριμένη λέξη ή φράση είναι δυνατόν να προσδιοριστεί διότι αφενός μεν είναι μία έννοια ήδη γνωστή, αφετέρου γιατί η ακολουθία των ήχων από τους οποίους συντίθεται εκφράζεται σε ένα πλαίσιο από συμφραζόμενες λέξεις ή προτάσεις το οποίο έχει έννοια. Η εκ των προτέρων γνώση της λέξης ή της φράσης και η υποβοήθηση από τα συμφραζόμενα καθιστούν ικανούς τους ανθρωπούς στο να ερμηνευούν την ερώτηση “Jeet?” ως “Did you eat?”, όταν αυτή λέγεται το μεσημέρι ή στις 6.00 μ.μ.

Όταν η λέξη ή φράση είναι ασυνήθιστη ή όταν δεν μπορούμε να χρησιμοποιήσουμε τα συμφραζόμενα για να προσδιορίσουμε τι πιθανώς έχει λεχθεί, τότε η μόνη πηγή δεδομένων είναι μία ακολουθία από ήχους. Αυτή η ακολουθία ήχων είναι γεμάτη παραμόρφωση που οφείλεται στην συνάρθρωση και σε άλλους παράγοντες, κάνοντας έτσι δύσκολο τον προσδιορισμό των φωνημάτων και των λέξεων που δρούν σαν ετικέτες. Αυτές είναι οι συνθήκες κάτω από τις οποίες γεννιούνται παρανοήσεις όπως για παράδειγμα στην περίπτωση που κάποιος ακούσει “pullet surprise” αντί για “Pulitzer Prize”.

Τα συστήματα αναγνώρισης ομιλίας λειτουργούν όπως τον άνθρωπο που άκουσε “pullet surprise”. Η αρχική τους πηγή δεδομένων είναι μία προεπεξεργασμένη ακολουθία από ακουστικές παραμέτρους οι οποίες θα πρέπει να μεταφραστούν σε μία λέξη ή σε μία σειρά από λέξεις. Για να επιτελεστεί αυτή η διεργασία κατά την εκτέλεση μίας εφαρμογής τα συστήματα αναγνώρισης ομιλίας έχουν αποθηκευμένα μοντέλα λέξεων (μερικές φορές καλούνται ***μοντέλα αναφοράς***). Κατά την διαδικασία της αναγνώρισης συγκρίνεται η ακολουθία των ακουστικών παραμέτρων που εκφράζεται από τον χρήστη με αυτή των αποθηκευμένων μοντέλων.

Οι κατασκευαστές διαφωνούν στον τρόπο με τον οποίο τα μοντέλα αναφοράς αναπαρίστανται. Οι μέθοδοι αναπαράστασης που χρησιμοποιούνται συνήθως είναι:

- Ίχνη
- Κρυφά Μοντέλα Markov (HMM)
- Ακολουθίες Φωνημάτων
- Ακολουθίες Υπολέξεων

Η επιλογή που ο κατασκευαστής κάνει δηλώνει το πως η έννοια "λέξη" ορίζεται και την μεθοδολογία αναγνώρισης που θα εφαρμοστεί.

Πολλές από τις παραπάνω μεθόδους αναπαράστασης αποτελούν εργαλεία που έχουν ήδη χρησιμοποιηθεί κατά την διαδικασία αναγνώρισης ομιλίας. Τα κοινά χαρακτηριστικά ανάμεσα στην εστίαση πάνω στην αναπαράσταση και στην ενασχόληση με το λεξιλόγιο απορρέουν από το γεγονός ότι η αναγνώριση γενικά εκτελείται σε επίπεδο λέξης.

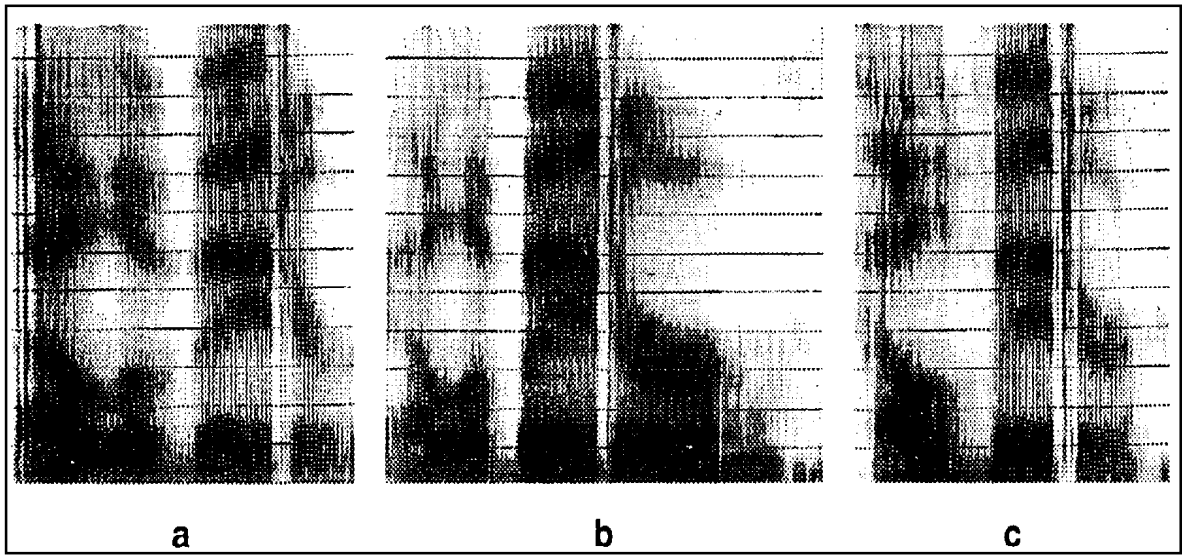
Μεταβλητότητα

Όλες οι μέθοδοι για την αναπαράσταση του λεξιλογίου θα πρέπει να αντμετωπίζουν το θέμα της μεταβλητότητας σχετικά με τον τρόπο με τον οποίο οι λέξεις ομιλούνται. Η μεταβλητότητα προκύπτει από μία σειρά από πηγές. Μία από αυτές, η συνάρθρωση αναφέρεται στην επίδραση των περιβαλλόντων ήχων. Η μεταβλητότητα είναι επίσης το αποτέλεσμα διαφορών στην ομιλία μεταξύ των ομιλητών. Κάθε ομιλητής έχει μία μοναδική διαμόρφωση της φωνητικής του οδού. Οι ομιλητές διαφέρουν ως προς το μέγεθος και το σχήμα του στόματος, της γλώσσας, του λαιμού, και των δοντιών. Οι ομιλητές διαφέρουν επίσης στον τρόπο με τον οποίο προφέρουν τα φωνήματα και τις λέξεις. Μερικά πρότυπα προφοράς μεγάλου πληθυσμού ομιλητών μπορούν να ομαδοποιηθούν με βάση τις τοπικές, ταξικές, ηλικιακές διαφορές, ή της διαφοράς στο φύλο. Άλλα πρότυπα σχετίζονται με την ιδιοσυγκρασία. Μερικοί άνθρωποι, για παράδειγμα, προφέρουν το φώνημα *λ* με το πίσω μέρος της γλώσσας τους, ενώ άλλοι με την άκρη της γλώσσας τους. Μερικοί ομιλητές μιλούν πιο αργά από άλλους ή με ένρρινο τρόπο.

Η μεταβλητότητα υπάρχει όχι μόνο ανάμεσα στους ομιλητές, αλλά και στην ομιλία ενός ομιλητή. Οι λέξεις ομιλούνται με διαφορετική ταχύτητα και ακουστότητα. Αντανακλούν τα διαφορετικά επίπεδα των συναισθημάτων και της κόπωσης. Όπως φαίνεται από το σχήμα 1, η έννοια της μεταβλητότητας είναι ένα θεμελιώδες χαρακτηριστικό της ανθρώπινης ομιλίας, γεγονός που ενισχύεται από το ότι ένα άτομο σπάνια προφέρει την ίδια λέξη με ακριβώς τον ίδιο τρόπο δύο φορές.

Η συνάρθρωση, οι διαφορές στην ομιλία μεταξύ των ομιλητών, και οι ανακολουθίες στην ομιλία του ίδιου ομιλητή, αλληλεπιδρούν με τέτοιο τρόπο ώστε να παράγουν ένα σύνθετο πρότυπο μεταβλητότητας το οποίο θα πρέπει να αντιμετωπιστεί. Η ικανότητα ενός συστήματος αναγνώρισης ομιλίας να λειτουργεί σωστά σε συνθήκες όπου εμφανίζεται μεταβλητότητα στην

ομιλία, είναι μία από τις προϋποθέσεις της ***ευρωστείας***. Η άλλη κύρια προϋπόθεση της ευρωστείας είναι η ακριβής αναγνώριση κάτω από μεταβλητές συνθήκες θορύβου.



Σχήμα 1. Φασματογράφημμα από τις τρεις εκφράσεις της λέξεις "elevator" από τον ίδιο ομιλητή.

ΤΟ ΗΧΗΤΙΚΟ ΦΑΣΜΑΤΟΓΡΑΦΗΜΜΑ

Πριν από την χρήση του φασματογραφήμματος δεν υπήρχε τρόπος αναπαράστασης των συχνοτήτων που ήταν ενσωματωμένες στην ομιλία. Τα διαγράμματα του παλμογράφου όπου αποτυπώνεται η ηχητική πίεση των κυματομορφών ομιλίας καλύπτουν συνηθισμένες μορφές και παρουσιάζουν εκφράσεις της ίδιας λέξεις που έχει λεχθεί από τον ίδιο ομιλητή να εμφανίζεται εντελώς διαφορετική. Αντίθετα, το φασματογράφημμα παράγει μία λεπτομερή εκτύπωση των φωνοσυντονισμικών μορφών της ομιλίας, η οποία πραγματοποιεί αλλαγές στα πρότυπα αυτά άμεσα εμφανείς στο ανθρώπινο μάτι.

Το ηχητικό φασματογράφημμα ανακαλύφθηκε κατά τη διάρκεια του δευτέρου παγκοσμίου πολέμου από τον Ralph Potter και τους συνεργάτες του στα Bell Laboratories. Αρχικά το ηχητικό φασματογράφημμα χρησιμοποιήθηκε σαν εργαλείο από γιατρούς με σκοπό την καλυτέρευση της επικοινωνίας σε ανθρώπους με σοβαρά ακουστικά προβλήματα. Χρησιμοποιήθηκε εκτενώς από γλωσσολόγους, θεραπευτές, και ερευνητές επεξεργασίας ομιλίας για την κατανόηση και την περιγραφή προτύπων ομιλίας.

Η πρώτη περιγραφή του ηχητικού φασματογραφήμματος παρουσιάστηκε στην δημοσίευση 2654 του Science τον Νοέμβριο του 1945. Ο Potter και δύο συνεργάτες του δημοσίευσαν αργότερα το Visible Speech (1947), ένα βιβλίο που περιγράφει το φασματογράφημμα και την χρήση του.

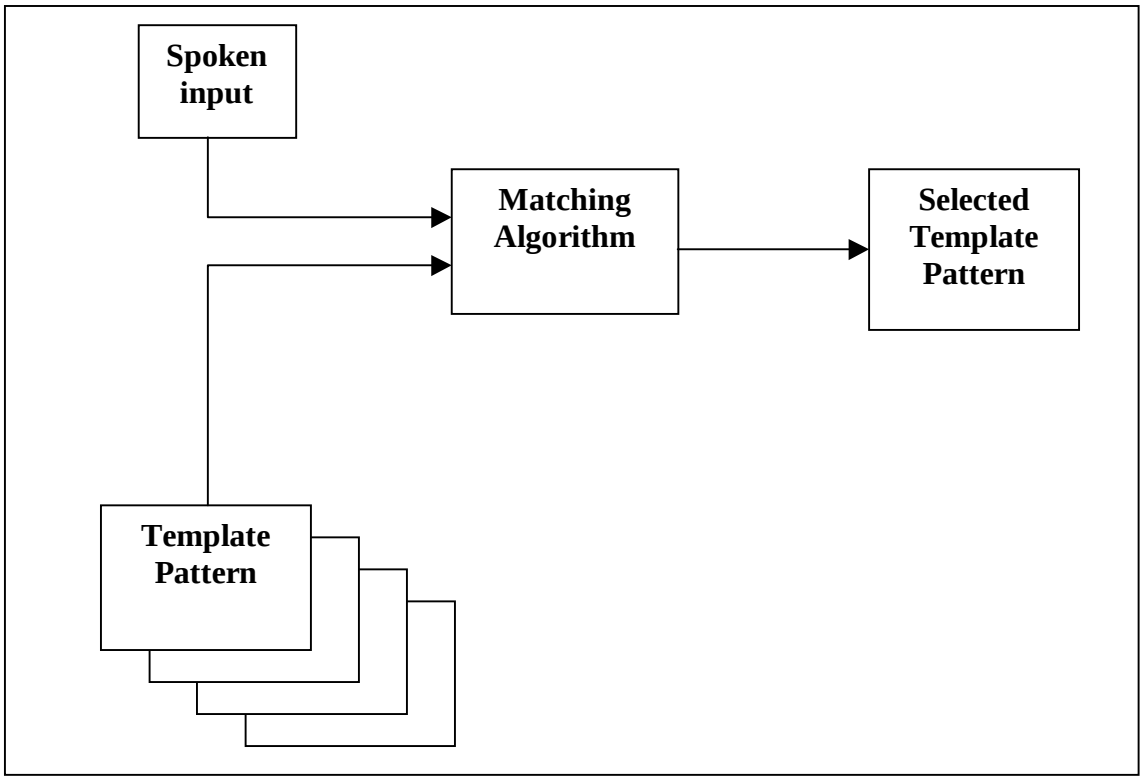
Ίχνη

Το ταίριασμα των ιχνών είναι μία μορφή αναγνώρισης προτύπων. Η διαδικασία ταιριάσματος των ιχνών εξισώνει μία λέξη με τα ίχνη αναπαράστασης της. Αναπαριστά τα δεδομένα ομιλίας σαν σύνολα παραμετρικών διανυσμάτων τα οποία καλούνται ***ίχνη***. Με άλλα λόγια ένα ίχνος είναι μία ακολουθία από ανύσματα. Κάθε άνυσμα περιέχει ένα σύνολο από τιμές για τις παραμέτρους που χρησιμοποιούνται από το σύστημα για την αναπαράσταση της ομιλίας. Η αναπαράσταση είναι απλή, άμεση και εύκολη στο να παραχθεί. Δεν υπάρχει εσωτερική δομή πέρα από την ακολουθία των ανυσμάτων που παράγονται με την διαδικασία της κωδικοποίησης. Δεν γίνεται προσπάθεια ώστε να αναλυθούν γλωσσικές ή φωνητικές σχέσεις οι οποίες μπορεί να υπάρχουν εντός της λέξης.

Η διαδικασία ταιριάσματος των ιχνών πραγματοποιείται ως εξής**:** σε μία εφαρμογή κάθε λέξη ή φράση αποθηκεύεται σαν ένα ξεχωριστό ίχνος. Η ομιλία των τελικών χρηστών στην είσοδο είναι οργανωμένη σε ίχνη πριν την εκτέλεση της διαδικασίας της αναγνώρισης. Στη συνέχεια η είσοδος συγκρίνεται με τα αποθηκευμένα ίχνη, και όπως φαίνεται από το σχήμα 2, τα αποθηκευμένα ίχνη που ταιριάζουν περισσότερο με τα εισερχόμενα πρότυπα ομιλίας, αναγνωρίζονται σαν την λέξη ή φράση εισόδου. Το ίχνος που επιλέγεται ονομάζεται ***καλύτερο ταίριασμα*** για την συγκεκριμένη είσοδο. Η διαδικασία του ταιριάσματος των ιχνών απαιτεί την τμήμα προς τμήμα σύγκριση των φασματικών προτύπων και παράγει μία ολική εκτίμηση της ομοιότητας για κάθε ίχνος (συχνά καλείται ***μετρική απόσταση***). Η σύγκριση δεν αναμένεται να παράγει πανομοιότυπο ταίριασμα. Μεμονωμένες εκφράσεις της ίδιας λέξης, εκφωνημένες ακόμη και από το ίδιο άτομο, συχνά διαφέρουν ως προς το χρονικό μήκος. Η διακύμανση αυτή μπορεί να οφείλεται σε μία σειρά από παράγοντες, όπως την διαφορά στον ρυθμό με τον οποίο ένα άτομο μιλάει, την έμφαση που δίνεται στην ομιλία και την συναισθηματική κατάσταση του ατόμου. Όποιο και να είναι το αίτιο θα πρέπει να βρεθεί τρόπος έτσι ώστε να ελαχιστοποιηθούν οι χρονικές διαφορές μεταξύ των προτύπων, με αποτέλεσμα τόσο οι αργές όσο και οι γρήγορες εκφράσεις της ίδιας λέξης να μην προσδιορίζονται σαν ξεχωριστές λέξεις. Η διαδικασία ελαχιστοποίησης των διαφορών του χρονικού μήκους των λέξεων καλείται ***χρονική ευθυγράμμιση.*** Η προσέγγιση που χρησιμοποιείται πιο συχνά για την εκτέλεση της χρονικής ευθυγράμμισης στην διαδικασία ταιριάσματος των ιχνών είναι μία τεχνική ταιριάσματος προτύπων που καλείται ***δυναμική στρεύλωση χρόνου (DTW)*** **.** Η μέθοδος αυτή συνιστά την βέλτιστη χρονική ευθυγράμμιση ενός συνόλου από διανύσματα (ίχνη) με ένα άλλο.

Τα περισσότερα συστήματα ταιριάσματος των ιχνών έχουν ένα προκαθορισμένο κατώφλι αποδεκτικότητας. Αυτή η λειτουργία αποτρέπει τον θόρυβο και τις λέξεις που δεν ανήκουν στο λεξιλόγιο της εφαρμογής να αναγνωριστούν λαθεμένα σαν αποδεκτά δεδομένα ομιλίας στην είσοδο. Αν κανένα από τα ταιριάσματα ίχνους δεν υπερβεί το κατώφλι αποδεκτικότητας, τότε δεν

καταγράφεται αναγνώριση. Ο τρόπος με τον οποίο τα γεγονότα μη αναγνώρισης αντιμετωπίζονται διαφέρει από σύστημα σε σύστημα. Για παράδειγμα πόλλα συστήματα ζητούν από τον χρήστη να επαναλάβει την λέξη ή την φράση.



Σχήμα 2. Αναγνώριση ομιλίας με την χρήση της μεθόδου ταιριάσματος των ιχνών.

Στα αρχικά συστήματα ιχνών κάθε ίχνος αναπαριστούσε ένα παράδειγμα (λεγόταν ***εκφώνηση***) λέξης παραγόμενης από τον ομιλητή. Δέκα εκφωνήσεις αποθηκεύονταν σαν δέκα διαφορετικά ίχνη και καθένα από αυτά τα ίχνη ήταν "παγωμένο" και αμετάβλητο. Το γεγονός αυτό είχε σαν αποτέλεσμα η μεταβλητότητα στην ομιλία μεταξύ των ομιλητών και η μεταβλητότητα στην ομιλία του ίδιου ομιλητή να περιοριστεί μέσω της δημιουργίας των συνόλων των ιχνών για κάθε λέξη του λεξιλογίου της εφαρμογής.

Αυτή είναι η απλούστερη προσέγγιση των συναρτήσεων μεταβλητότητας, αποδεκτή για συστήματα μικρού λεξιλογίου που περιέχουν ισχυρά διαφοροποιημένες λέξεις. Τα στοιχεία που απαιτούνται για αποθήκευση και αναζήτηση δια μέσω των ιχνών αυξάνουν γραμμικά με το μέγεθος του λεξιλογίου. Χρησιμοποιώντας αυτή την προσέγγιση, το ταίριασμα των ιχνών για λεξιλόγια μεγαλύτερα απο εκατό λέξεις καθίσταται μη αποδεκτό σε ότι αφορά την δαπάνη χρόνου και την τάση εμφάνισης λάθους.

Πιο πρόσφατες εφαρμογές ιχνών χρησιμοποιούν εύρωστα ίχνη. Τα εύρωστα ίχνη δημιουργούνται από περισσότερες από μία εκφωνήσεις μίας λέξης χρησιμοποιώντας μαθηματικούς μέσους όρους και στατιστικές τεχνικές συσταδοποίησης. Αυτές οι διαδικασίες

επικεντρώνονται στο να κανονικοποιήσουν την συχνότητα και τα χρονικά πρότυπα των εκφωνήσεων μίας λέξης. Επειδή συμπεριλαμβάνουν δεδομένα από περισσότερες εκφωνήσεις κατορθώνουν καλύτερα να χειριστούν την μεταβλητότητα στην ομιλία μεταξύ των ομιλητών. Η χρήση των εύρωστων ιχνών μειώνει επίσης την γρήγορη ανάπτυξη απαιτήσεων αποθήκευσης οι οποίες συνήθως σχετίζονται με τα συστήματα ταιριάσματος των ιχνών.

Η χρήση της μεθόδου ταιριάσματος των ιχνών είναι σήμερα περιορισμένη, αλλά βρίσκει ακόμη εφαρμογή σε μερικά εμπορικά συστήματα. Είναι πολύ αποτελεσματική σε εφαρμογές μικρού λεξιλογίου, αλλά δεν είναι αρκετά ακριβής για να κάνει τους λεπτούς διαχωρισμούς που απαιτούνται στις εφαρμογές αναγνώρισης που χρησιμοποιούν μεγάλο λεξιλόγιο. Εφαρμογές που απαιτούν λεξιλόγιο μεσαίου μεγέθους, δηλαδή αυτό κυμαίνεται από χίλιες έως δέκα χιλιάδες λέξεις, μπορούν να χρησιμοποιούν την μέθοδο ταιριάσματος των ιχνών, αν ο αριθμός των επιλογών λεξιλογίου σε κάθε σημείο της εφαρμογής κρατείται στο ελάχιστο επίπεδο. Ένα άλλο πρόβλημα είναι η αύξηση των απαιτήσεων αποθήκευσης με ένα γραμμικό τρόπο μαζί με την αύξηση του μεγέθους του λεξιλογίου. Το γεγονός αυτό παράγει μία ανάλογη αύξηση της υπολογιστικής πολυπλοκότητας.

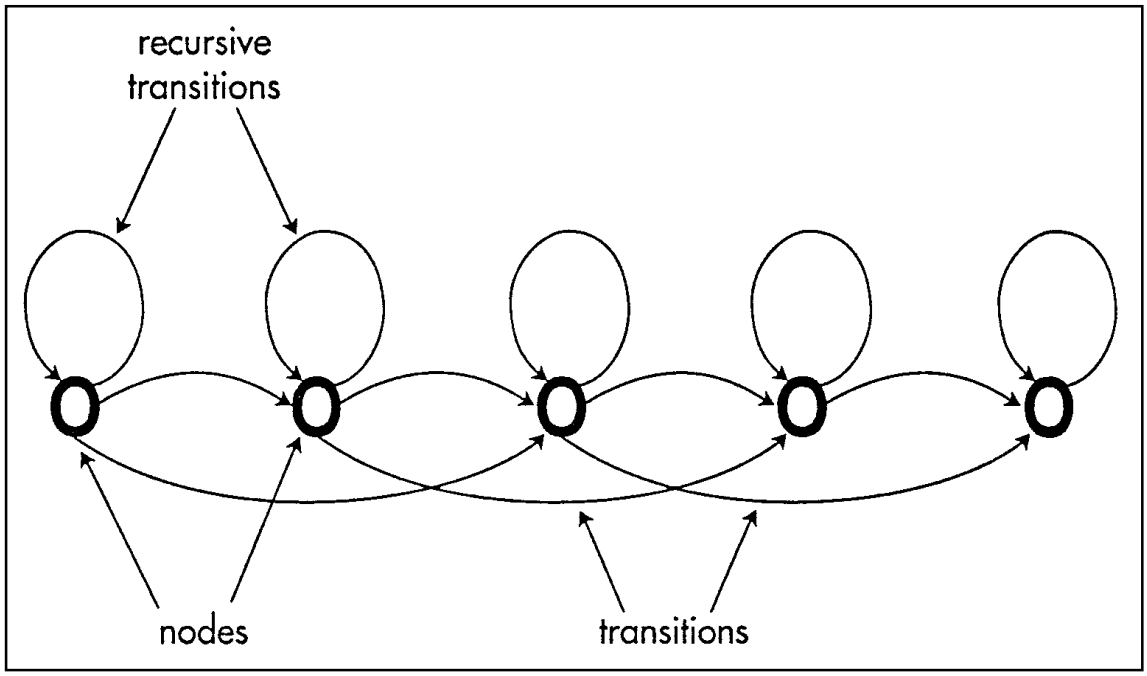
ΤΑΙΡΙΑΣΜΑ ΤΩΝ ΙΧΝΩΝ

Το ταίριασμα των ιχνών ήταν η επικρατέστερη μεθοδολογία αναγνώρισης κατά την δεκαετία του 1950 και 1960. Η αδυναμία της στο να παράγει ακριβή αναγνώριση, σε συνδιασμό με το αυξανόμενο ενδιαφέρον για τις ακουστοφωνητικές τεχνικές συντέλεσαν στο να χάσει την αίγλη της. Κατά την δεκαετία του 1970 επιχειρήθηκε μία ανανέωση της μεθόδου λόγω της απογοήτευσης που ακολούθησε από την εμφάνιση των ακουστοφωνητικών εφαρμογών. Εν τω μεταξύ είχε βελτιωθεί με την πρόσθεση τεχνικών όπως αυτή της ***δυναμικής στρέβλωσης του χρόνου (DTW)***. Οι νέοι αλγόριθμοι ταιριάσματος των ιχνών ήταν γρήγοροι, εύρωστοι, και αρκετά ακριβείς για τις απαιτήσεις των εμπορικών συστημάτων στις αρχές της δεκαετίας του 1980. Η επικράτηση τους συνεχίστηκε μέχρι το τέλος της δεκαετίας του 1980 όπου και αντικαταστάθηκαν από στοχαστικές μεθόδους επεξεργασίας. Αν και η μέθοδος ταιριάσματος των ιχνών, σαν βασική μέθοδος αναγνώρισης, είναι σε πτώση, χρησιμοποιήθηκε σε ***εφαρμογές στόχευσης λέξεων***. Ακόμη αποτελεί τη βασική τεχνολογική μέθοδο που εφαρμόζεται για την εξακρίβωση της ταυτότητας ενός ομιλητή.

Οι Leedham (1992), Moore (1984), και οι Rabiner & Jyang (1993) παρέχουν άριστη τεχνική ανάλυση της μεθόδου ταιριάσματος των ιχνών.

Κρυφά Μοντέλα Markov (HMM)

Ο όρος ***στοχαστικός*** αναφέρεται στην διαδικασία υλοποίησης μίας ακολουθίας από ***μη ντιτερμινιστικές*** επιλογές ανάμεσα από ένα σύνολο από εναλλακτικές επιλογές. Είναι μη ντιτερμινιστικές γιατί οι επιλογές κατά την διαδιακασία της αναγνώρισης κατευθύνονται από τα χαρακτηριστικά της εισόδου και δεν προσδιορίζονται εκ των προτέρων. Η χρήση των στοχαστικών μοντέλων και της διαδικασίας της στοχαστικής επεξεργασίας έχει διεισδύσει σημαντικά στην διαδικασία αναγνώρισης ομιλίας. Όπως η διαδικασία ταιριάσματος των ιχνών, η στοχαστική επεξεργασία απαιτεί την δημιουργία και την αποθήκευση μοντέλων καθενός από τα προς αναγνώριση στοιχεία. Στο σημείο αυτό οι δύο προσεγγίσεις αποκλίνουν. Η στοχαστική επεξεργασία δεν περιλαμβάνει άμεσο ταίριασμα ανάμεσα στα αποθηκευμένα μοντέλα και στα δεδομένα εισόδου. Αντίθετα, βασίζεται σε σύνθετη στατιστική και πιθανοκρατική ανάλυση η οποία γίνεται καλύτερα κατανοητή εξετάζοντας την δικτυακή δομή στην οποία αυτές οι στατιστικές μορφές αποθηκεύονται: Τα Κρυφά Μοντέλα Markov (HMM).



Σχήμα 3. Τυπική δομή ενός Κρυφού Μοντέλου Markov.

Ένα HMM, όπως αυτό του σχήματος 3 αποτελείται από μία ακολουθία από καταστάσεις συνδεδεμένες με μεταβάσεις. Οι καταστάσεις αναπαριστούν τις εναλλακτικές καταστάσεις της στοχαστικής διαδικασίας και οι μεταβάσεις περιέχουν πιθανοκρατικά και άλλου είδους δεδομένα τα οποία χρησιμοποιούνται για να καθορίσουν ποια κατάσταση θα πρέπει να επιλεγεί στην συνέχεια. Οι καταστάσεις του HMM στο σχήμα 3 αναπαρίστανται με κύκλους και οι μεταβάσεις με βέλη. Έτσι για παράδειγμα, οι δυνατές μεταβάσεις από την πρώτη κατάσταση του HMM μπορούν να είναι, στην πρώτη κατάσταση (καλείται ***αναδρομική μετάβαση***), στην επόμενη

κατάσταση, ή στην τρίτη κατάσταση του HMM. Ας υποθέσουμε ότι το HMM του σχήματος 3 είναι ένα μοντέλο αποθήκευσης της λέξης "five", θα καλείται ***μοντέλο αναφοράς*** για τη λέξη "πέντε" και θα περιέχει στατιστικά στοιχεία για όλα τα εκφωνημένα δείγματα της λέξης τα οποία χρησιμοποιούνται για να δημιουργήσουν το μοντέλο αναφοράς. Οι εκφωνήσεις είναι δυνατό να παρέχονται από μεμονωμένους χρήστες, ομάδες ομιλητών ή από προϋπάρχουσα βάση δεδομένων (καλείται ***σώμα κειμένων***) ψηφιοποιημένης ομιλίας. Κάθε κατάσταση ενός HMM περιέχει στατιστικά στοιχεία για κάθε τμήμα της λέξης. Τα στατιστικά στοιχεία περιγράφουν τις τιμές και την διακύμανση των παραμέτρων που περιέχονται στα δείγματα της λέξης. Ένα σύστημα αναγνώρισης μπορεί να έχει πολυάριθμα HMM, όπως αυτό του σχήματος 3, ή μπορεί να τα ενοποιήσει σε ένα δίκτυο καταστάσεων και μεταβάσεων.

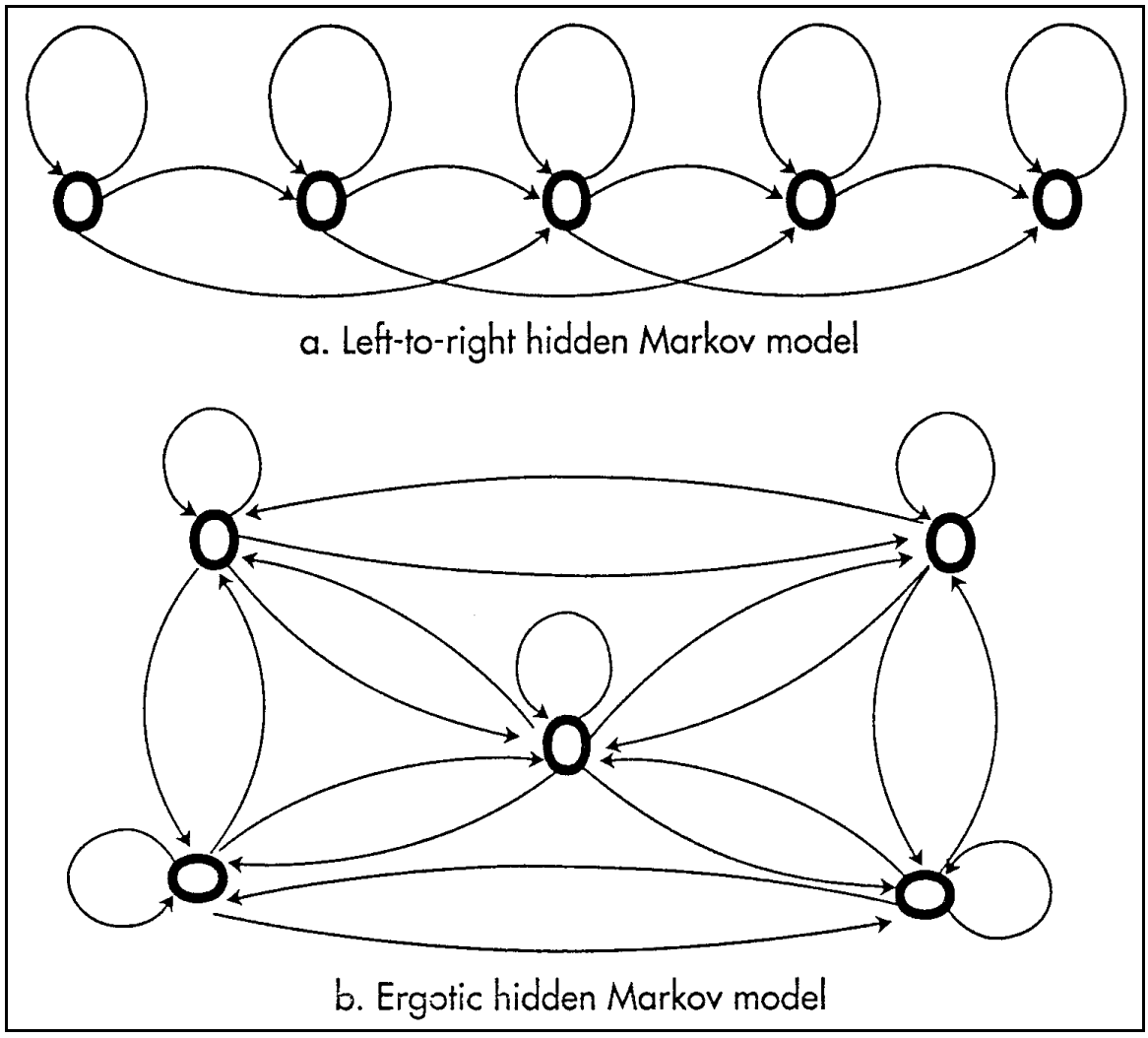
Το σύστημα αναγνώρισης προβαίνει στην σύγκριση των δεδομένων στην είσοδο με τα αποθηκευμένα μοντέλα του συστήματος. Εάν ο χρήστης επρόκειτο να προφέρει "fayv" τότε το σύστημα πιθανόν να επέλεγε το HMM που απεικονίζεται στο σχήμα 3 σαν ένα από τα αποθηκευμένα μοντέλα για να πραγματοποιήσει την σύγκριση με τα δεδομένα στην είσοδο από τον χρήστη. Εάν ο χρήστης παρατείνει την εκφώνηση του f κατά το ξεκίνημα της προφοράς της λέξης εισόδου, είναι πιθανόν όταν η διάταξη αναγνώρισης συγκρίνει την είσοδο με το HMM του σχήματος 3 να υπάρξει τουλάχιστον μία αναδρομική μετάβαση στην πρώτη κατάσταση του HMM.

Οι συγκρίσεις αυτές παράγουν μία τιμή πιθανότητας δείχνοντας έτσι την πιθανοφάνεια ένα συγκεκριμένο αποθηκευμένο μοντέλο αναφοράς να αποτελεί το καλύτερο ταίριασμα για τα δεδομένα εισόδου. Αυτή η προσέγγιση καλείται ***αλγόριθμος μέγιστης πιθανοφάνειας των Baun-Welch.*** Μία άλλη συνήθης μέθοδος που χρησιμοποιείται στην στοχαστική αναγνώριση είναι ο ***αλγόριθμος Viterbi.*** Ο ***αλγόριθμος Viterbi*** ψάχνει μέσα από το δίκτυο των κόμβων για μία ακολουθία από καταστάσεις του HMM οι οποίες αντιστοιχούν περισσότερο στην είσοδο. Αυτό ονομάζεται το ***καλύτερο μονοπάτι.***

Η απόδοση ενός HMM εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων που χρησιμοποιούνται για την δημιουργία των εσωτερικών πιθανοτήτων. Οι πιθανότητες αυτές ορίζουν την φύση και την έκταση της διακύμανσης η οποία σχετίζεται με μία λέξη ή με κάποια άλλη είσοδο. Κατευθύνουν την επιλογή του καλύτερου μονοπατιού ή του καλύτερου ταιριάσματος. Αν αυτές οι πιθανότητες προέρχονται από ανεπαρκή δεδομένα ή από εκφωνήσεις που δεν αντικατοπτρίζουν την ομιλία του αναμενόμενου πληθυσμού χρηστών, τότε η αναγνώριση με την χρήση του HMM θα είναι μη αποτελεσματική.

Τα HMM είναι γρήγορα, αποτελεσματικά, ακριβή, και ευέλικτα. Ένα κριτήριο για τα HMM είναι το ότι οι καταστάσεις τους είναι λειτουργικά ανεξάρτητες.

Το μειωνέκτημα των HMM είναι η μη ακριβή υπόθεση του Markov, συγκεκριμένα ότι οι ακουστικές πραγματοποιήσεις και η διάρκεια μίας κατάστασης εξαρτάται μόνο από την τρέχουσα κατάσταση και είναι υπό συνθήκη ανεξάρτητη από το παρελθόν … τα HMM δεν παρέχουν σε μεγάλο βαθμό επίγνωση της διαδικασίας αναγνώρισης. Σαν αποτέλεσμα, είναι δύσκολο να αναλυθούν τα λάθη ενός συστήματος HMM σε μία προσπάθεια να βελτιωθεί η απόδοση του (Alex Waibel, Carnegie Mellon University &Kai-Fu Lee, Carnegie Mellon University [υπό τις παρούσες συνθήκες Apple Computer] Ερμηνεία της Αναγνώρισης Ομιλίας, 1990, σελ.263).



Σχήμα 4. Δύο τυπικές αρχιτεκτονικές Κρυφών Μοντέλων Markov.

Τα HMM μπορούν να δομηθούν με ποικίλους τρόπους. Ο πιο συνηθισμένη HMM αρχιτεκτονική που συναντάται στα συστήματα αναγνώρισης ομιλίας είναι το HMM με διεύθυνση από αριστερά προς τα δεξία, όπως φαίνεται στο σχήμα 4.α. Το εργοδικό HMM που φαίνεται στο σχήμα 4.β, συχνά χρησιμοποιείται για την αναπαράσταση δικτύων από φωνήματα ή υπολεξέων. Τα εργοδικά HMM είναι χωρίς διεύθυνση και συνδέουν κάθε κατάσταση με κάθε άλλη κατάσταση.

Στην ορολογία της γλωσσικής θεωρίας ένα HMM μπορεί να θεωρηθεί σαν ένα ***Αυτόματο Πεπερασμένων Καταστάσεων*** (FSA). Ο όρος FSA για ένα HMM χρησιμοποιείται για την μοντελοποίηση γλωσσικών πληροφοριών, ενώ ο όρος HMM χρησιμοποιείται στα διάφορα μοντέλα σε ακουστικό επίπεδο. Τα δύο μοντέλα είναι ισοδύναμα αν και χρησιμοποιούνται για να μοντελοποιήσουν διαφορετικά φαινόμενα και γενικά δεν υπάρχει σταθερή χρήση των συγκεκριμένων όρων στην βιβλιογραφία αναγνώρισης ομιλίας.

Οι Moore (1984), Paul (1990), και Rabiner & Juang (1986) παρέχουν λεπτομηρείς τεχνικές περιγραφές των Κρυφών μοντέλων Markov. Η δημοσίευση του Jame Baker (1975) παρουσιάζει την στοχαστική επεξεργασία στην διαδικασία αναγνώρισης ομιλίας.

ΚΡΥΦΑ ΜΟΝΤΕΛΑ MARKOV

Το 1913 ο A. A Markov περιέγραψε ένα μοντέλο δικτύου ικανό να παράγει ακολουθίες Ρωσσικών γραμμάτων ή να προβλέπει ακολουθίες γραμμάτων χρησιμοποιώντας ακολουθίες πιθανοτήτων οι οποίες εκθέτονται σε ρωσσικά κείμενα. Χρησιμοποιήθηκε σαν βάση σε υπολογιστικά μοντέλα, σε μία ευρεία περιοχή πεδίων, συμπεριλαμβανομένων και μοντέλων της ανθρώπινης γλώσσας.

Στην δεκαετία του 1960 και στις αρχές της δεκαετίας του 1970 τα μοντέλα Markov εφαρμόστηκαν σε πολυεπίπεδα, σε ιεραρχικές δομές από τον Baum και άλλους ερευνητές. Εφόσον οι πιθανοκρατικοί υπολογισμοί των υποκείμενων επιπέδων δεν θεωρούνται σαν μέρη των ακολουθιών υψηλών επιπέδων, τα μοντέλα αυτά ονομάστηκαν Κρυφά μοντέλα Markov (HMM).

Οι ερευνητές άρχισαν να διερευνούν την χρήση των HMM στην αναγνώριση ομιλίας στις αρχές της δεκαετίας του 1970. Ένας από τους πρώτους που υιοθέτησε τα HMM ήταν ο James Baker του πανεπιστημίου Carnegie Mellon (CMU), ο οποίος τα χρησιμοποιήσε για να αναπτύξει το σύστημα DRAGON του πανεπιστημίου CMU για το πρόγραμμα ARPA SUR. Ένας άλλος υποστηρικτής των HMM ήταν ο Frederick Jelinek του οποίου η ερευνητική ομάδα στην IBM συνέβαλλε στην ανάπτυξη της τεχνολογίας των HMM. Το 1982, ο James και η Janet Baker ίδρυσαν τα συστήματα Dragon και αργότερα κατασκεύασαν το σύστημα DragonScribe, ένα από τα πρώτα εμπορικά προϊόντα που χρησιμοποίησαν την τεχνολογία HMM.

Η τεχνολογία HMM δεν απέκτησε ευρεία αποδοχή σε εμπορικά συστήματα μέχρι τα τέλη της δεκαετίας 1980, αλλά από τις αρχές της δεκαετίας του 1990 τα HMM έγιναν η κύρια προσέγγιση στην αναγνώριση ομιλίας.

Για πρόσθετες πληροφορίες συνίσταται στον αναγνώστη να ανατρέξει στις δημοσιεύσεις των James Baker (1975α και 1975β) και Baum (1972).

Μοντέλα Φωνημάτων

Σύμφωνα με την ακουστοφωνητική προσέγγιση κατά την διαδικασία της αναγνώρισης οι λέξεις θεωρούνται σαν ακολουθίες από φωνήματα. Η γοητεία των φωνημάτων είναι πολύπλευρη:

- Οι φωνητικές μονάδες έχουν το πλεονέκτημα ότι προσφέρουν μία οικονομική αναπαράσταση διότι οι περισσότερες γλώσσες έχουν μόνο πενήντα ή και λιγότερα φωνήματα.
- Η αύξηση του λεξιλογίου δεν παράγει την αντίστοιχη γραμμική αύξηση στις απαιτήσεις αποθήκευσης καθώς και στις υπολογιστικές απαιτήσεις που σχετίζονται με άλλες μεθοδολογίες.
- Το γεγονός ότι τα φωνήματα είναι τα βασικά στοιχεία της ανθρώπινης γλώσσας υποδηλώνει ότι μπορεί να γίνει ευκολότερη η μεταφορά εφαρμογών από μία γλώσσα σε μία άλλη και έτσι θα μπορούν να χρησιμοποιηθούν για αυτόματη αύξηση του λεξιλογίου.

Υπάρχουν πάρα πολλοί τρόποι για την δημιουργία μοντέλων φωνημάτων. Η παραδοσιακή προσέγγιση εξαρτάται από γλωσσικούς κανόνες σχεδιασμένους από τον ίδιο τον κατασκευαστή και εκφράζουν τις σκέψεις του για τις εντυπωσιακές ιδιότητες των φωνημάτων. Η ανάγκη για τη δημιούργια μεγάλων, επεκτάσιμων, γενικής χρήσης λεξιλογίων μετακίνησε τους κατασκευαστές προς τις υπολογιστικές προσεγγίσεις, όπως τα νευρωνικά δίκτυα και την στατιστική ανάλυση. Οι μέθοδοι που υλοποιούνται με το χέρι και οι υπολογιστικές μέθοδοι μπορούν να συνδιαστούν και να παρέχουν λεπτομερείς αναπαραστάσεις ικανές για αυτοματοποιημένη αύξηση ή αναπαραγωγή. Για παράδειγμα, το σύστημα MIT VOYAGER συνδίασε εργαλεία κατασκευασμένα με το χέρι και υπολογιστικά εργαλεία για να δημιουργήσει φωνητικά ***μοντέλα προέλευσης*** για τα Αγγλικά και τα Ιαπωνικά. Τα στατιστικά παραγόμενα φωνήματα καλούνται μονάδες που μοιάζουν με φωνήματα (ή ***PLU*** ή ***ανεξάρτητα από το περιεχόμενο***) και αναπαρίστανται συνήθως χρησιμοποιώντας εργοδικά HMM, όπου κάθε PLU μπορεί να αναπαρασταθεί σαν μία HMM κατάσταση.

Υπάρχουν πολλά προβλήματα με την ακουστοφωνητική προσέγγιση. Ένα από αύτα είναι ότι η εντατική με το χέρι υλοποιήση, γιατί για την κατάρτηση του συστήματος φωνητικής ταξινόμησης απαιτείται η δημιουργία ενός μεγάλου αριθμού από φωνητικά δεδομένα στα οποία η τοποθέτηση των ετικετών υλοποιείται με το χέρι. Η επιμονή στην κατασκευή αυτού του τμήματος με το χέρι είναι συνέπεια της εξάρτησης από την γλωσσική ανάλυση και την έλλειψη κοινά αποδεκτών κριτηρίων για την ακουστοφωνητική ανάλυση. Οι λεπτομέρειες της αναπαράστασης είναι σημαντικές γιατί χρησιμοποιούνται από τους κανόνες του συστήματος.

Ωστόσο οι ερευνητές διαφοροποιούνται ως προς τα ακουστοφωνητικά χαρακτηριστικά που θα πρέπει να αναπαρασταθούν.

Ένα άλλο πρόβλημα που αντιμετωπίζουν τα ακουστοφωνητικά συστήματα είναι η αναπαράσταση της διακύμανσης. Σε θεωρητικές διατυπώσεις τα φωνήματα αναπαρίστανται σαν σταθερές οντότητες, αλλά κατά την ομιλία παρουσιάζουν σε μεγάλο βαθμό μεταβλητότητα. Μία μέθοδος για την αναπαράσταση της μεταβλητότητας είναι η χρήση του ***φθόγγου***. Ένας φθόγγος περιέχει την ακουστική πληροφορία για μία μοναδική έκφραση του φωνήματος και αποτελεί άριστο τρόπο αναπαράστασης της τυχαίας μεταβλητότητας. Όμως η έννοια του φθόγγου δεν μπορεί να καλύψει τις περιπτώσεις εκείνες όπου εμφανίζονται αναμενόμενες μορφές μεταβλητότητας παραγόμενες από διαλέκτους καθώς και από το φωνητικό περιεχόμενο (συνέπειες της συνάρθρωσης).

Η πιο συνηθισμένη προσέγγιση στον χειρισμό της αναμενόμενης μεταβλητότητας, όπως οι συνέπειες της συνάρθωσης, είναι ο σχεδιασμός ενός συνόλου από κανόνες που συνοδεύουν ανεξάρτητα από το περιεχόμενο φωνήματα χρησιμοποιούμενα στην ακουστοφωνητική αναπαράσταση. Οι κανόνες ενός ακουστοφωνητικού συστήματος θα πρέπει να καλύπτουν τις σύνθετες μορφές της συνάρθρωσης και τις άλλες μορφές της μεταβλητότητας και να τις εκφράζουν από την πλευρά των ακουστοφωνητικών χαρακτηριστικών που χρησιμοποιούνται στην αναπαράσταση. Αν το ακουστοφωνητικό σύστημα είναι κατασκευασμένο από μία συγκεκριμένη γλωσσική θεωρία, τότε η θεωρία αυτή θα πρέπει να είναι πολύ καλά ορισμένη. Αν δεν υπάρχει κάποια γλωσσική θεωρία που να υπόκειται στην αναπαράσταση οι κατασκευαστές θα πρέπει να εκτελέσουν εκτενή με το χέρι ανάλυση ή να σχεδιάσουν εργαλεία για να προσδιορίσουν τις μορφές της διακύμανσης, τέτοια που να μπορούν να μετατραπούν σε κανόνες ή χαρακτηριστικά της αναπαράστασης.

Οι Cole (1986), Mercier, και άλλοι, (1989), και Zue (1985) παρέχουν καλές τεχνικές περιγραφές για τα μοντέλα φωνημάτων. Οι Ljolje & Levinson (1991) περιγράφουν ένα παράδειγμα στο οποίο χρησιμοποιείται ένα εργοδικό HMM για την αναπαράσταση των φωνημάτων. Ο Oshika, και άλλοι, (1975) παρέχουν ένα παράδειγμα γλωσσικής ανάλυσης για την αναγνώριση ομιλίας.

Υπολέξεις

> Η μοντελοποίηση των υπολέξεων αποτέλεσε ένα ολοένα αυξανόμενης σημασίας θέμα γιατί καθώς η χωρητικότητα των συστημάτων αναγνώρισης αυξάνει, γίνεται δύσκολη αν όχι αδύνατη η εκμάθηση των μοντέλων ολόκληρου του κόσμου (Hsiao-Wuen & Kai-Fu Lee, Carnegie Mellon University [υπό τις παρούσες συνθήκες Apple Computer], "Μοντελοποίηση Ομιλίας πάνω σε Ανεξάρτητο Λεξιλόγιο," 1990, p.725).

Το πρόβλημα αυτό αναφέρεται σε συστήματα που ορίζουν λέξεις από την πλευρά των ακουστοφωνητικών στοιχείων τους (φωνήματα) ή άλλες μονάδες υπολέξεων. Η χρήση των φωνημάτων είναι ελκυστική, αλλά οι προσεγγίσεις οι βασιζόμενες σε φωνήματα παρέχουν μικρή βοήθεια στον χειρισμό των επιδράσεων της συναρθρώσης. Ένας αριθμός από εναλλακτικές μονάδες υπολέξεων προτάθηκε, οι οποίες αντλούν τις πληροφορίες τους από ομιλούμενα δεδομένα παρά από θεωρητικές δημιουργίες. Περιέχουν φθόγγους, PLU, συλλαβές, ημισυλλαβές (το μισό της συλλαβής) και δίφθογγους. Η πιο επιτυχημένη μονάδα υπολέξης είναι ο ***τρίφθογγος*** (επίσης καλείται ***ευαίσθητο από το περιεχόμενο PLU*** (CS-PLU) και ***φώνημα στο περιεχόμενο*** (PIC)) γιατί είναι πολύ εύρωστη και γιατί είναι εφικτή η κατασκευή της μέσα στα όρια της υπάρχουσας τεχνολογίας των υπολογιστών.

ΦΩΝΗΜΑΤΑ ΣΤΟ ΠΕΡΙΕΧΟΜΕΝΟ

Στις αρχές της δεκαετίας του 1980 οι ερευνητές άρχισαν να διατυπώνουν εναλλακτικά μοντέλα ως προς το ανεξάρτητο από το περιεχόμενο μοντέλο φωνήματος. Ένας από τους πρώτους που κατασκεύασε ένα μοντέλο φωνήματος στο περιεχόμενο ήταν ο Boil Beranek και ο Newman (BBN). Το 1985, οι BBN πρόσφεραν μία λεπτομερή πρόταση για την χρήση των φωνημάτων στο περιεχόμενο και την ολοκλήρωσαν με το BYBLOS, ένα σύστημα αναγνώρισης μεγάλου λεξιλογίου από τους BBN. Αυτός ο διαφορετικός τύπος φωνήματος ονομάστηκε τρίφθογγος και εκφράζει την πρόσθεση τόσο δεξιών όσο και αριστερών περιεχομένων.

Διαφορετικοί τύποι τριφθόγγων κατασκευάστηκαν από την IBM, την Dragon Systems, το πανεπιστήμιο CMU και άλλους οργανισμούς που εργάζονται πάνω σε συστήματα μεγάλου λεξιλογίου. Το τρίφθογγο μοντέλο είναι μία προκαθορισμένη προσέγγιση στα συστήματα αναγνώρισης ομιλίας μεγάλου λεξιλογίου. Η εκτεταμένη χρήση του καθιστά ικανούς και τους κατασκευαστές συστημάτων μικρού λεξιλογίου να εισέλθουν στην αγορά των συστημάτων μεγάλου λεξιλογίου.

Για την παροχή περισσότερων πληροφοριών πάνω στο θέμα αυτό συνίσταται η δημοσιεύσεις του Schwartz, και άλλων, (1985), καθώς επίσης και το βιβλίο του Kai-Fu Lee (1989) για την περιγραφή του συστήματος SPHINX από το πανεπιστήμιο CMU.

Δημιουργία Τριφθόγγων και Υπολέξεων

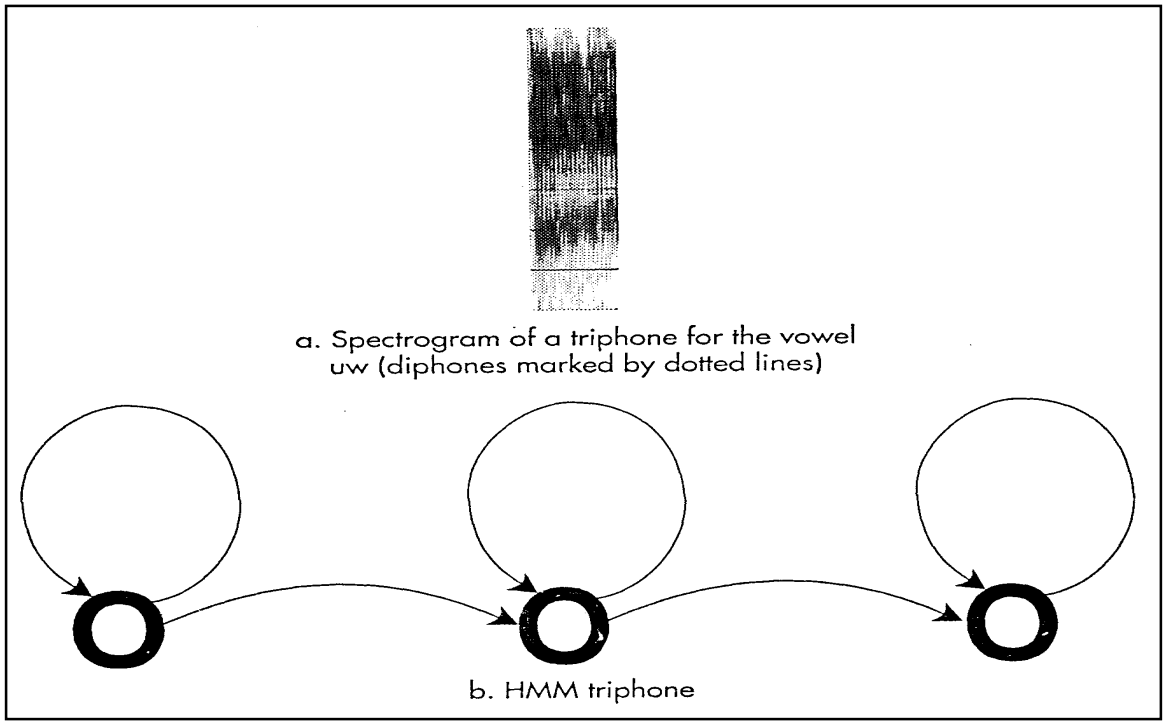
Ένας τρίφθογγος δεν είναι ένα φώνημα, παρά το γεγονός ότι μερικοί κατασκευαστές αναφέρουν τα PCI ή τους τριφθόγγους σαν φωνήματα. Ένας τρίφθογγος αποτελείται από ένα φώνημα ή ένα PLU και περιστοιχίζεται και από τις δύο πλευρές από πληροφορίες που προκύπτουν από τα συμφραζόμενα. Ένας τρίφθογγος για το φώνημα uw απεικονίζεται στο σχήμα 5.α. Όπως τα PLU, τα τρίφθογγα μοντέλα είναι στατιστικές κατασκευές που βασίζονται

στην ανάλυση μεγάλου αριθμού δειγμάτων. Γενικά, αναπαρίστανται σαν HMM, όπως αυτό που αποικονίζεται στο σχήμα 5.β. Συνήθως περιέχουν τρεις καταστάσεις αναπαριστώντας:

- Την μετάβαση από το προηγούμενο φώνημα
- Το φώνημα
- Την μετάβαση από το ένα φώνημα στο άλλο φώνημα

Η στατιστική πληροφορία που αναπαρίσταται σε ένα τρίφθογγο HMM περιέχει μεταβλητότητα προερχόμενη από την συνάρθρωση. Αντίθετα με τα ακουστοφωνητικά μοντέλα η ανάλυση της επίδρασης της συνάρθρωσης στην ομιλία εισόδου μπορεί να γίνει χρησιμοποιώντας αποκλειστικά ένα τρίφθογγο μοντέλο παρά συμπεριλαμβάνοντας μεγάλο αριθμό από κανόνες συνάρθρωσης.

Την ίδια στιγμή η ενσωμάτωση της συνάρθρωσης σε ένα τρίφθογγο, συνεπάγεται γενικά την δημιουργία περισσότερων από ένος τρίφθογγων μοντέλων για ένα μόνο φώνημα. Το φώνημα t, για παράδειγμα μπορεί να έχει ένα τρίφθογγο μοντέλο για φωνητικά περιεχόμενα τα οποία απαιτούν στρογγιλεμένο στόμα (“too”, “toe”, κτλ), κάποιο άλλο για περιεχόμενα χωρίς στρογγιλεμένο στόμα, και πολλά άλλα για άλλα φωνητικά περιεχόμενα. Ο αριθμός των τριφθόγγων που απαιτούνται για την αγγλική γλώσσα είναι πολύ μεγαλύτερος από τον συνολικό αριθμό των φωνημάτων.



Σχήμα 5. Τρίφθογγοι.

Τα HMM τρίφθογγα μοντέλα αποθηκεύονται σε μία βάση δεδομένων. Κατά την ανάπτυξη του λεξιλογίου τα τρίφθογγα μοντέλα συνδέονται για να σχηματίσουν στοιχεία

εφαρμογών λεξιλογίου ή εφεδρικό λεξιλόγιο το οποίο είναι αποθηκευμένο σε ένα σύστημα λεξικού. Αντίθετα στην αναπαράσταση όπου χρησιμοποιούνται ίχνη και HMM σε επίπεδο λέξης, νέα στοιχεία λεξιλογίου μπορούν να προστεθούν σε μία εφαρμογή χωρίς την συλλογή επιπρόσθετων δειγμάτων για κάθε νέα λέξη. Οι νέες λέξεις εισάγονται από το πληκτρολόγιο. Η είσοδος αποτελείται από γραφημικά τμήματα υπολέξεων, τα αντίστοιχα τρίφθογγα μοντέλα προσπελαύνονται από την βάση δεδομένων, και τα μοντέλα συνδέονται για να σχηματίσουν το μοντέλο μίας νέας λέξης. Η διαδικασία ανάπτυξης ονομάζεται ***μοντελοποίηση υπολέξεων.***

> Αυτό που κάνουμε είναι να παίρνουμε πραγματικές εκφωνήσεις από εγχώριους ομιλητές. Τις διαχωρίζουμε σε κομμάτια από φωνήματα τα οποία οργανώνονται σε τρίφθογγους. Έπειτα τις ανακατασκευάζουμε έτσι ώστε να σχηματίσουμε τα μοντέλα. Είναι σαν να χρησιμοποιούμε στοχαστικά παιχνίδια. Είναι σαν να διαχωρίζω πράγματα σε κομμάτια και να τα ανακατασκευάζουμε ξανά, ακόμη και για λέξεις για τις οποίες δεν είχαμε ποτέ εκφωνήσεις.

Οι νέες λέξεις που δημιουργούνται αποθηκεύονται σε ένα σύστημα λεξιλογίου στο οποίο υπάρχουν δείκτες στους τρίφθογγους από τους οποίους έχουν κατασκευαστεί..

Τεχνικές περιγραφές της κατασκευής τριφθόγγων και της μοντελοποίησης υπολέξεων δίνονται από τον C-F Lee, και άλλους, (1990a και 1990b) και από τον Deng, και άλλους, (1990). Επίσης συνίσταται το βιβλίο του Kai-Fu Lee (1989) στο οποίο περιγράφονται τα φωνήματα, οι τρίφθογγοι, και η μοντελοποίηση λέξεων για το σύστημα SPHINX του πανεπιστημίου Carnegie Mellon.

Θέματα στην Ανάπτυξη Τριφθόγγων.

Με την ακουστοφωνητική μοντελοποίηση, η υλοποίηση τρίφθογγων μοντέλων σε ένα συγκεκριμένο σύστημα αναγνώρισης είναι συχνά μοναδική για τον συγκεκριμένο κατασκευαστή και για τμήμα του λογισμικού που ανήκει στον κατασκευαστή. Αυτό θα έχει ως αποτέλεσμα, τα τρίφθογγα μοντέλα του ενός κατασκευαστή να μην συνεργάζονται με το πρόγραμμα αναγνώρισης του άλλου κατασκευαστή.

Ένα άλλο θέμα στην ανάπτυξη των τριφθόγγων είναι η ποιότητα. Εφόσον βασίζονται σε στατιστική ανάλυση δεδομένων, η ποιότητα των τρίφθογγων μοντέλων είναι ένα υποπροϊόν της ποιότητας της εκμάθησης που χρησιμοποιείται για την δημιουργία αυτών των μοντέλων. Είναι συνήθως δύσκολο να αποκτηθούν αρκετά δείγματα για να καλύψουν όλα τα δυνατά φωνητικά περιεχόμενα, έτσι ώστε να δημιουργηθούν αξιόπιστα τρίφθογγα μοντέλα. Αυτό καλείται πρόβλημα ***αραιών δεδομένων***. Μερικοί κατασκευαστές επιλύουν αυτό το πρόβλημα ορίζοντας μονάδες λεξιλογίου οι οποίες αναπτύσονται μέσω της μοντελοποίησης υπολέξεων σαν αρχικά σημεία και καλούνται ***βασικές φόρμες***. Οι βασικές φόρμες σχεδιάζονται έτσι ώστε να προσαρμόζονται στον τρόπο ομιλίας κάθε χρήστη απέναντι στο σύστημα. Αυτή η προσέγγιση

καλείται ***προσαρμογή ομιλητή***. Οι εφαρμογές που αναπτύσονται χρησιμοποιώντας την μέθοδο της προσαρμογής ομιλιτή πρόκειται να χρησιμοποιηθούν επαναληπτικά από το ίδιο σύνολο χρηστών.

Μία δεύτερη μέθοδος, η οποία μπορεί να συνδιαστεί με την παραγωγή βασικών μορφών είναι η ανάπτυξη τρίφθογγων μοντέλων τα οποία μπορούν να χρησιμοποιηθούν σαν ***βάση δεδομένων αναφοράς***, (επίσης καλείται ***σώμα κειμένων ομιλούμενης γλώσσας***) περιλαμβάνοντας δείγματα από μεγάλο αριθμό ανθρώπων οι οποίοι διαβάζουν λέξεις ή προτάσεις. Μία από τις πιο ευρέως χρησιμοποιούμενες βάσεις αυτής της κατηγορίας είναι η TIMIT, ένα σώμα κειμένων που αναπτύχθηκε από την Texas Instruments και το MIT για να καλύψει ολόκληρο το σύστημα ήχων των Αμερικανικών Αγγλικών.

Ένας άλλος τρόπος ανάλυσης του προβλήματος των αραιών δεδομένων είναι η δημιουργία ενός πιο γενικού τρίφθογγου μοντέλου ή μοντέλου υπολέξης. Η ***γενικεύμενη τρίφθογγη προσέγγιση*** η οποία αναπτύχθηκε στο CMU είναι μία από τις πιο διάσημες προσεγγίσεις. Ένας γενικευμένος τρίφθογγος δημιουργείται συγκεντρώνοντας και συγχωνεύοντας ακουστικά δεδομένα από τα περιεχόμενα της συνάρθρωσης για ένα μόνο φώνημα. Αν και τα προκύπτοντα φωνήματα είναι πιο γενικευμένα, το πανεπιστήμο CMU κατέληξε ότι η ακρίβεια αναγνώρισης δεν ελαττώνεται όταν τα μοντέλα είναι σωστά καταρτισμένα. Ο K-F Lee, και άλλοι, (1990a) εξήγησαν τους γενικευμένους τρίφθογγους. Ένας αριθμός κατασκευαστών που χρησιμοποιούν υπολέξεις ενσωματώνουν τις παραλλαγές των γενικευμένων τριφθόγγων στα συστήματα τους με σκοπό να αυξήσουν την απόδοση χωρίς να εμφανίσουν επιπρόσθετη πολυπλοκότητα.

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΑΝΑΦΟΡΑΣ

Η εμφάνιση των ***βάσεων δεδομένων αναφοράς*** στα συστήματα αναγνώρισης ομιλίας έγινε στα μέσα της δεκαετίας του 1980 και αποτέλεσε ένα σημαντικό παράγοντα στην ανάπτυξη των συστημάτων αυτών. Οι πιο ευρέως αναφερόμενες στην βιβλιογραφία βάσεις δεδομένων αναφοράς είναι οι: DARPA Resources Management Database (DRMD) (Price, και άλλοι, 1988), TIMIT Acoustic Phonetic Database (Fisher, και άλλοι, 1986) και η Texas Instruments/National Bureau of Standards (TI/NBS) Database of Connected Digits (Leonard, 1984).

Η βάση δεδομένων DRMD είναι μία βάση χιλίων λέξεων η οποία περιέχει στοιχεία για την αναγνώριση ομιλίας εξαρτημένου-ομιλητή, ανεξάρτητου-ομιλητή, και προσαρμοζόμενου-ομιλητή. Βασίζεται σε 21000 εκφωνήσεις λέξεων της αγγλικής γλώσσας σχετικές με τον τομέα που αφορά την ναυτική διοίκηση, συγκεντρωμένες από 160 ομιλητές και από ένα σύνολο από διαλέκτους. Το υλικό της βάσης διαχωρίζεται σε σύνολα δεδομένων που έχουν εκπαιδευτικό και εξεταστικό χαρακτήρα. Η βάση DRMD δημοσιεύτηκε το 1988 και χρησιμοποιήθηκε ευρέως στον έλεγχο συστημάτων αναγνώρισης συνεχούς ομιλίας μεγάλου λεξιλογίου.

Η βάση δεδομένων TIMIT αποτελεί ένα άλλο πρόγραμμα υποστηριζόμενο από την DARPA. Η TIMIT είναι μία βάση δεδομένων αποτελούμενη από φωνητικά σύμβολα η εγγραφή των οποίων έχει γίνει με ψηφιακό τρόπο από την Texas Instruments Corporation (TI) και η μετεγγραφή της έγινε στο Massachusetts Institute of Technology (MIT). Το υλικό για την βάση συγκεντρώθηκε από το MIT και το TI και το Stanford Research Institute (SRI). Περιέχει τα δεδομένα για 4200 προτάσεις οι οποίες έχουν εκφωνηθεί από 630 ομιλητές διαφόρετικων διαλέκτων. Τα δεδομένα από 420 ομιλητές χρησιμοποιήθηκαν ως βάση με εκπαιδευτικό χαρακτήρα, ενώ τα άλλα δεδομένα έχουν εξεταστικό χαρακτήρα.

Τέλος η βάση TI/NBS είναι μία συλλογή από εκφωνήσεις ψηφίων για χρήση σε δοκιμές αναξάρτητων-ομιλητών. Τα δεδομένα περιλαμβάνουν την ομιλία 300 ανδρών, γυναικών και παιδιών, και η εγγραφή τους έχει πραγματοποιηθεί σε μη ενθόρυβο περιβάλλον. Το υλικό της βάσης περιλαμβάνει ακολουθίες από ψηφία το μήκος των οποίων κυμαίνεται από ένα έως επτά. Άξια αναφοράς είναι επίσης η πρώτη βάση δεδομένων που δημιουργήθηκε για βιομηχανικούς σκοπούς, η TI-46, η οποία κατασκευάστηκε από την TI. Ως δεδομένα περιείχε το αλφάβητο, τα ψηφία, και πολλές βασικές λέξεις. Χρησιμοποιείται ακόμη στον τομέα των νευρωνικών δικτύων και σε άλλες εφαρμογές.

ΠΡΑΓΜΑΤΟΠΟΙΩΝΤΑΣ ΑΝΑΓΝΩΡΙΣΗ ΟΜΙΛΙΑΣ ΜΕΓΑΛΟΥ ΛΕΞΙΛΟΓΙΟΥ

Το 1930 ο Tihamer Nemes, ένας Ούγγρος επιστήμονας, κατέθεσε μία ευρεσιτεχνία αυτόματης καταγραφής οπτικών τμημάτων ήχου από ταινίες. Η ευρεσιτεχνία απορρίφθηκε ως αβάσιμη, και έτσι δεν γνωρίζουμε πότε είχε αντιληφθεί το όνειρο της αναγνώρισης ομιλίας μεγάλου λεξιλογίου. Σαράντα χρόνια αργότερα μέσω του πρόγραμματος ARPA SUR επαναδραστηριοποιήθηκε το ενδιαφέρον για τον τομέα της αναγνώρισης ομιλίας με την χρήση μεγάλων λεξιλογίων. Σε αυτό το πρόγραμμα, όλοι οι συμμετέχοντες αξίωναν την ανάπτυξη συστημάτων ικανών για την επεξεργασία λεξιλογίων της τάξης των χιλίων λέξεων ή και περισσότερων. Δεν παρήχθησαν μόνο συστήματα αναγνώρισης ομιλίας, αξιοσημείωτο το ***Harpy,*** το οποίο ξεπέρασε αυτό το ελάχιστο όριο, αλλά επιβλήθηκε το όριο των χιλίων λέξεων σαν ορισμός του συστήματος αναγνώρισης μεγάλου λεξιλογίου. Ο ορισμός παρέμεινε σε ισχύ μέχρι το τέλος της δεκαετίας του 1980.

Το σύστημα ***Tangora*** της IBM ήταν μία από πρώτες ερευνητικές προσπάθειες που σχεδιάστηκαν για την παραγωγή ενός εμπορικού συστήματος ικανού να αναγνωρίζει λεξιλόγια των είκοσι και περισσοτέρων χιλιάδων λέξεων. Πήρε το όνομα του από τον Albert Tangora ο οποίος ήταν εγγεγραμμένος στο βιβλίο των ρεκόρ ***Guinness*** σαν ο γρηγορότερος δακτυλογράφος με 147 λέξεις το λεπτό. Ο πρώτος κατασκευαστής που διέθεσε ένα εμπορικό σύστημα με μία σημαντικότητα σε λεξιλόγιο μεγαλύτερη από χίλιες λέξεις ήταν η Speech Systems, Inc, (SSI). Το 1985 το PE100 της SSI περιείχε ένα λεξικό με είκοσι χιλιάδες λέξεις, αλλά απαιτούσε ένα σταθμό εργασίας που να διαθέτει λειτουργικό σύστημα UNIX.

Η άφιξη της τεχνολογίας των ισχυρών ηλεκτρονικών υπολογιστών άνοιξε το δρόμο για συστήματα μεγάλου λεξιλογίου βασιζόμενα σε ηλεκτρονικούς υπολογιστές. Αλγόριθμοι οι οποίοι θα οδηγούσαν τον 80286 σε μία κατάσταση αδύνατης λειτουργίας τρέχουν εύκολα σε περιβάλλον 386. Το τέλος της δεκαετίας 1980 είδε την άφιξη των συστημάτων υπαγόρευσης που βασίζονται σε ηλεκτρονικούς υπολογιστές με λεξιλόγια της τάξης των είκοσι χιλιάδων λέξεων ή και περισσότερων. Σήμερα ένα σύστημα το οποίο έχει χίλιες λέξεις θεωρείται σαν σύστημα με μικρό λεξιλόγιο, και εταιρίες τα προϊόντα των οποίων είχαν οριοθετηθεί σε έναν αριθμό λέξεων που είναι λιγότερες από χίλιες, κατά την δεκαετία του 1980, τώρα προσφέρουν συστήματα ικανά για την επεξεργασία πολλών χιλιάδων λέξεων. Από το 1994 υπάρχουν συστήματα με λεξιλόγια που ξεπερνούν τις εκατό χιλιάδες λέξεις.

ΣΧΕΔΙΑΣΜΟΣ ΜΕΓΑΛΟΥ ΛΕΞΙΛΟΓΙΟΥ.

Ένας από τους θεμελιώδεις στόχους που τίθεται στον τομέα αναγνώρισης ομιλίας είναι η ικανότητα επικοινωνίας του ομιλητή με ένα σύστημα αναγνώρισης χρησιμοποιώντας οποιοδήποτε λεξιλόγιο θεωρεί αναγκαίο για να επιτελέσει το στόχο του. Αν ο στόχος του είναι να υπαγορεύσει ένα γράμμα η αναγκαότητα ενός μεγάλου και εκτενούς λεξιλογίου είναι προφανής. Αν ο στόχος του είναι η αναζήτηση κάποιας πληροφορίας σε μία βάση δεδομένων ή η εισαγωγή δεδομένων, τότε η ανάγκη για χρήση μεγάλου λεξιλογίου δεν είναι άμεσα εμφανής γιατί πολλά συστήματα είναι σχεδιασμένα με τέτοιο τρόπο ώστε να κατευθύνουν τους χρήστες να προσαρμόζονται με το σύστημα. Καθώς ο αριθμός των τύπων των εφαρμογών αναγνώρισης ομιλίας αυξάνει και καθώς ο πληθυσμός των χρηστών καθίσταται πιο ανομοιογενής, η ανάγκη για συστήματα ικανά να εξυπηρετήσουν τις ανάγκες σε λεξιλόγιο μίας ευρύτερης περιοχής χρηστών θα αυξάνει. Τελικά τα συστήματα μεγάλου λεξιλογίου θα είναι ο κανόνας παρά η εξαίρεση.

Επεκτείνοντας το Λεξιλόγιο.

Τα συστήματα μεγάλου λεξιλογίου χρησιμοποιούν υπολέξεις για τον ορισμό των λέξεων στο λεξιλόγιο. Αυτό είναι αναγκαίο γιατί ο χρόνος, το κόστος, και η προσπάθεια για την ανάπτυξη μοντέλων της τάξης των δέκα χιλιάδων λέξεων ή και περισσοτέρων θα ήταν απαγορευτική. Χρησιμοποιώντας κατά τον σχεδιασμό υπολέξεις, η γραφημική ή η φωνητική αναπαράσταση μίας λέξης μπορεί να μετατραπεί σε τρίφθογγη αναπαράσταση σε διάστημα δευτερολέπτων.

Ένα περισσότερο ενδιαφέρον πρόβλημα που αντιμετωπίζουν οι κατασκευαστές συστημάτων μεγάλου λεξιλογίου είναι αυτό της συλλογής του λεξιλογίου που πρόκειται να εντάξουν στο σύστημα τους. Υπάρχουν πολλές συχνά χρησιμοποιούμενες λέξεις, όπως τα ψηφία ή οι μέρες της εβδομάδας, οι οποίες είναι προφανώς υποψήφιες σε ένα προς κατασκευή σύστημα, αλλά όταν κάποιος σχεδιάζει ένα λεξικό των πενήντα ή των εκατό χιλιάδων λέξεων για εφαρμογές σε μία ευρεία περιοχή από βιομηχανίες, η επιλογή του λεξιλογίου θα πρέπει να περιέχει εξέταση στοιχείων από τον τύπο και on-line εγγράφων στοιχείων που προέχονται από ένα μεγάλο αριθμό από ανόμοιες πηγές.

Πολλοί κατασκευαστές βασίζονται πάνω σε ένα on-line σώμα κειμένων (καλείται επίσης ***βάσεις δεδομένων αναφοράς***) για την παροχή των απαιτούμενων λεξιλογίων. Τα σώματα κειμένων περιέχουν μηχανικά-αναγνώσιμα λεξικά, καταλόγους λέξεων και δημοσιευμένο υλικό από ειδικά επαγγέλματα. Εφόσον πολλές από τις πηγές αυτές δεν είναι σχεδιασμένες για να χρησιμοποιηθούν στην ανάπτυξη του λεξιλογίου σε συστήματα αναγνώρισης, θα πρέπει να περικοπεί το λεξιλόγιο εκείνο που δεν εμφανίζεται συχνά, οι λαθεμένα προφερόμενες λέξεις,

κύρια ονόματα, μη προφερόμενες λέξεις, ξένες λέξεις, διαλεκτικές παραλλαγές και σύνθετες λέξεις. Το λεξιλόγιο που απομένει θα πρέπει να μετατραπεί από την αναπαράσταση κειμένου στην αναπαράσταση που απαιτεί το σύστημα αναγνώρισης. Επιπροσθέτως, θα πρέπει να ληφθεί μέριμνα στην αναπαράσταση σημαντικών παραλλαγών των λέξεων όπως του πληθυντικού αριθμού και του αορίστου χρόνου.

Όλα τα παραπάνω εμπεριέχουν μεγάλο φόρτο εργασίας, και παρόλα αυτά δεν εξασφαλίζουν ότι κάθε λέξη και κάθε παραλλαγή της λέξης που ο ομιλητής θέλει να χρησιμοποιήσει έχει συμπεριληφθεί. Επίσης, τα συστήματα μεγάλου λεξιλογίου στερούνται της δυνατότητας να εγκλιματίσουν τους χρήστες του συστήματος στις υπάρχουσες επιλογές λεξιλογίου.

> Στην πράξη στα συστήματα αναγνώρισης συνεχούς ομιλίας μεγάλου λεξιλογίου είναι σχεδόν αδύνατον για τον ομιλητή να θυμάται ποιές λέξεις συμπεριλαμβάνονται στο λεξιλόγιο. Η πιθανότητα ο ομιλητής να χρησιμοποιήσει λέξεις εκτός του λεξιλογίου είναι αρκετά υψηλή (Ayman Asadi, Northeastern University, Richard & John Makhoul, BBN Systems and Technologies, "Μοντέλο Αυτοματοποίησης για την Πρόσθεση Νέων Λέξεων σε Συστήματα Αναγνώρισης Συνεχούς Ομιλίας Μεγάλου Λεξιλογίου", 1991,p. 305).

Οι υποθέσεις αυτές κάνουν την πιθανότητα αυτόματης αύξησης του λεξιλογίου μέσω μηχανικής εκμάθησης εξαιρετικά ενδιαφέρουσα. Μία τέτοια διαδικασία συνεπάγεται την αναζήτηση μίας λέξης που ειπώθηκε και δεν αποτελεί στοιχείο του λεξιλογίου. Τα υπάρχοντα εμπορικά συστήματα μεγάλης κλίμακας λεξιλογίου ανιχνεύουν νέες λέξεις έμμεσα, όταν το σύστημα επιλέξει μία λανθασμένη λέξη ή όταν το διαθέσιμο λεξιλόγιο αποτυγχάνει να συναντήσει ένα κατώφλι αποδεκτικότητας. Στην τελευταία περίπτωση, δείχνεται στον ομιλητή ένα σύνολο από υποψήφιες λέξεις από τις οποίες θα πρέπει να επιλέξει. Αν η επιθυμητή λέξη δεν υπάρχει στον κατάλογο, ο χρήστης μπορεί να δημιουργήσει ένα μοντέλο λέξης για μία νέα λέξη.

Οι κατασκευαστές νευρωνικών δικτύων εργάζονται επάνω στην αυτόματη επέκταση του λεξιλογίου. Ένα δίκτυο εκπαιδευμένο πάνω σε 234 Ιαπωνικές λέξεις αναγνώρισε σωστά είκοσι λέξεις τις οποίες δεν είχε συναντήσει ξανά. Ένα άλλο πειραματικό σύστημα που κατασκευάστηκε στα εργαστήρια της AT&T Bell είχε την δυνατότητα να μαθαίνει νέο Αγγλικό λεξιλόγιο από εκφράσεις διακριτών λέξεων χρησιμοποιώντας σαν οδηγό μη δακτυλογραφημένη είσοδο.

Δύο άριστες πηγές πληροφοριών σχετικά με τα θέματα ανάπτυξης μεγάλου λεξιλογίου αποτελούν οι εργασίες των Seitz, (1990) και Jelinek (1985). Η εργασία του Jelinek στην IBM έθεσε την προκαταρτική εργασία για τον σχεδιασμό λεξιλογίου και γραμματικής στα περισσότερα εμπορικά συστήματα αναγνώρισης μεγάλου λεξιλογίου. Ο Gorin, και άλλοι, (1993)

περιέγραψαν το αυτόματο σύστημα εκμάθησης λεξιλογίου το οποίο αναπτύχθηκε στα εργαστήρια της AT&T Bell.

Αποτελεσματικότητα Αναζήτησης.

Όλα τα συστήματα αναγνώρισης ομιλίας θα πρέπει να προσδιορίζουν το θέμα της ακριβής αναγνώρισης της εισόδου σε ένα χρονικό πλαίσιο που να είναι αποδεκτό από τους χρήστες του συστήματος. Τα συστήματα μικρού λεξιλογίου επιτρέπουν γραμμική αναζήτηση δια μέσω του διεθέσιμου λεξιλογίου, αλλά αυτό δεν είναι δυνατό για συστήματα με λεξιλόγιο μεγαλύτερο από χίλιες λέξεις.

Μία λύση για αυτού του είδους το πρόβλημα είναι η χρήση ***γραμματικής*** για τον περιορισμό των επιλεγόντων λέξεων που είναι διαθέσιμες σε κάθε σημείο. Μία άλλη τεχνική είναι να οργανωθεί το λεξικό στην μορφή δέντρου ή στην μορφή πλέγματος.

> Η αναγνώριση ομιλίας μπορεί να θεωρηθεί σαν ένα πρόβλημα αναζήτησης σε δενδροειδές δίκτυο. Καθώς κάποιος κατευθύνεται από την ρίζα προς τα φύλλα, οι διακλαδώσεις που ξεκινούν από κάθε ένωση αναπαριστούν το σύνολο των λέξεων που θα μπορούσαν να προσαρτηθούν στην τρέχουσα μερική πρόταση (Douglas Paul, Βιβλιοθήκη Lincoln του MIT "Ενας Αποτελεσματικός A* Αλγόριθμος Αποκωδικοποίησης Στοίβας, για Αναγνώριση Συνεχούς Ομιλίας με ένα Στοχαστικό Γλωσσικό Μοντέλο," 1992, p. 25).

Αυτή η προσέγγιση έχει υιοθετηθεί από τα περισσότερα συστήματα μεγάλου λεξιλογίου. Μία τέτοια οργάνωση μπορεί να ελαττώσει την αναζήτηση κατά ένα παράγοντα της τάξης του εφτά ελαχιστοποιώντας έτσι τα τμήματα των μή πιθανών επιλογών.

Μία τρίτη μέθοδος (συχνά συνδιάζεται με την δενδροειδή δομή) είναι η εφαρμογή πολύπλοκων αλγόριθμων αναζήτησης. Δύο από τους πιο συχνά χρησιμοποιούμενους αλγόριθμους αναζήτησης είναι ο ***αλγόριθμος αναζήτησης δέσμης*** και ο ***αλγόριθμος αποκωδικοποίησης στοίβας***. Τα συστήματα μεγάλου λεξιλογίου απαιτούν σύνθετες μεθόδους αναζήτησης ικανές να επιτελέσουν λεπτομερή ακουστική ανάλυση. Όπως αναφέρεται και στο ιστορικό σημείωμα, που ακολουθεί, ένας αριθμός από ***τεχνικές γρήγορου ταιριάσματος*** και άλλων πολύπλοκης αναζήτησης έχουν αναπτυχθεί.

ΑΝΑΖΗΤΗΣΗ

Μία από τις ευρέως χρησιμοποιούμενες μεθόδους αναζήτησης είναι ο ***αλγόριθμος αναζήτησης δέσμης*** (καλείται επίσης και ***αλγόριθμος αναζήτησης Viterbi***). Η χρησιμότητα του αλγόριθμου αναζήτησης δέσμης στην αναγνώριση ομιλίας υπερτονίστηκε με την επιτυχία του συστήματος Harpy της ARPA SUR το οποίο είχε την δυνατότητα να αναγνωρίζει 1,011 λέξεις με ρυθμό λάθους πέντε τις εκατό (5%). Ο αλγόριθμος αναζήτησης δέσμης κινείται συστηματικά κατά μήκος ενός δικτύου ταιριάζοντας την είσοδο με τα ακουστικά πρότυπα του δικτύου. Οι πρώτοι ήχοι της εισόδου για παράδειγμα συγκρίνονται με τις αρχικές καταστάσεις του δικτύου λαμβάνοντας τα στοιχεία που ταιριάζουν καλύτερα. Όλα τα άλλα μονοπάτια περικόπτονται από τον εναπομείναντα λόγο. Τα μονοπάτια που παραμένουν συνθέτουν την δέσμη των υποθέσεων. Αυτή η διαδικασία παράγει ένα δέντρο απόφασης ή ένα πλέγμα. Η αναζήτηση δέσμης παραμένει μία από τις πίο συχνά χρησιμοποιούμενες μεθόδους τόσο σε εμπορικά όσο και σε ερευνητικά συστήματα.

Στην δεκαετία του 1980 νέοι αλγόριθμοι αναζήτησης αναπτύχθηκαν για να χειριστούν την πρόκληση των συστημάτων μεγάλου λεξιλογίου. Μία ευρέως χρησιμοποιούμενη τεχνική η οποία διαδόθηκε από την IBM, ήταν η ***τεχνική αποκωδικοποίησης στοίβας***. Καθώς ο χρήστης προβαίνει σε μία εκφώνηση, η αποκωδικοποίηση στοίβας παράγει ένα βαθμονομημένο κατάλογο με υποψήφια μονοπάτια όμοιο με την δέσμη υποθετικών μονοπατιών που συναντάται στον αλγόριθμο αναζήτησης δέσμης. Ο βαθμονομημένος κατάλογος καλείται στοίβα. Ποικίλες μορφές αποκωδικοποίησης στοίβας έχουν εφαρμοστεί σε ερευνητίκα και εμπορικά συστήματα μεγάλου λεξιλογίου. Άλλες προσεγγίσεις, όπως ο ***αλγόριθμος γρήγορου ταιριάσματος*** (επίσης καλείται ***αλγόριθμος γρήγορης αναζήτησης***) άρχισαν να εμφανίζονται προς το τέλος της δεκαετίας του 1980. Σκοπός τους ήταν να μειώσουν τον απαιτούμενο χρόνο για τον προσδιορισμό ενός μικρού συνόλου από υποψήφιες λέξεις με μεγάλη πιθανότητα. Αντικατέστησαν την λεπτομερή ακουστική αναζήτηση των καθιερωμένων αλγόριθμων αναζήτησης με μία τεχνική ελάττωσης της αναζήτησης, σύμφωνα με την οποία ομαδοποιούνται οι λέξεις και οι υπολέξεις με παρόμοια ακουστική μορφή.

Κατά την ίδια περίοδο, οι ερευνητές άρχισαν την ανάπτυξη πιο πολύπλοκων αλγόριθμων αναζήτησης για συστήματα κατανόησης της ομιλούμενης γλώσσας. Η προσέγγιση ***N-best*** αναπτύχθηκε από τους BBN και απέκτησε υψηλή δημοτικότητα. Η N-best περιέχει πολλές μεθοδολογίες αναζήτησης εφαρμόζωντας την απλούστερη και ταχύτερη αρχή, για να περικόψει μη πιθανές υποψήφιες λέξεις από το να ληφθούν υπόψη. Ο προκύπτων κατάλογος με τις N υποψήφιες λέξεις καλείται ***N-best κατάλογος***. Ο N-best κατάλογος στη συνέχεια στέλνεται σε άλλους αλγόριθμους λιγότερο γρήγορης αναζήτησης για περαιτέρω επεξεργασία.

ΤΙ ΕΙΝΑΙ ΜΙΑ ΛΕΞΗ;

Στο τμήμα που εστιάζει στην τεχνολογία μία λέξη ορίζεται σε όρους της ακουστικής της αναπαράστασης. Για την καλύτερη λειτουργικότητα μίας εφαρμογής, εξετάζονται δύο άλλες πλευρές της έννοιας της λέξης που είναι εξίσου σημαντικές με το ακουστικό πρότυπο. Είναι:

- Ο προσδιοριστής λέξης
- Η μετάφραση

Προσδιοριστές Λέξεων.

Ένας προσδιοριστής είναι μία μοναδική ετικέτα ή όνομα που αποδίδεται σε μία λέξη. Χρησιμοποιείται για να ξεχωρίσει την συγκεκριμένη λέξη από όλες τις άλλες λέξεις που υπάρχουν στο λεξικό. Για ευκολία στον σχεδιασμό και στην χρήση από τους κατασκευαστές εφαρμογών τα περισσότερα εμπορικά συστήματα χρησιμοποιούν την γραφημική ή την εκτυπωμένη αναπαράσταση της λέξης σαν προσδιοριστή της. Στα συστήματα αυτά η λέξη “one”, για παράδειγμα, μπορεί να προσδιοριστεί από τα γράμματα o n e ή από το ψηφίο 1. Η χρήση του συλλαβισμού επιτρέπει την διαφοροποίηση μεταξύ των ομόφωνων λέξεων σαν τις “one” και “won”. Μερικά συστήματα είναι ευαίσθητα. Σε αυτά τα συστήματα ο προσδιοριστής ONE αναφέρεται σε διαφορετική λέξη από ότι ο προσδιοριστής one, ακόμη και αν οι λέξεις είναι ταυτόσημες.

Σε πόλλα συστήματα αναγνώρισης, φράσεις ή μικρές προτάσεις που συμπεριφέρονται σαν μονάδες μπορούν να οριστούν σαν λέξεις αν οι ετικέτες τους δεν έχουν κενά διαστήματα. Οι εκφράσεις “end of application” και Where am I?” θα αποτελέσουν ετικέτες με την μορφή end-of-application ή Where-am-I.

Μεταφράσεις.

Η λέξη “μετάφραση” δεν είναι ένας τεχνικός όρος στην βιομηχανία αναγνώρισης ομιλίας, αλλά είναι μία συνιστώσα του σταδίου επικοινωνίας κατά την αναγνώριση ομιλίας. Η ***μετάφραση*** αναφέρεται στην μετατροπή της αναγνωρισμένης εισόδου σε μία μορφή όπου άλλες συνιστώσες της εφαρμογής μπορούν να χρησιμοποιήσουν. Η μορφή της και το περιεχόμενο της ορίζεται από το λογισμικό ή το υλικό που λαμβάνει η αναγνωρισμένη είσοδος. Εάν μία ομιλούμενη λέξη μετατρέπεται σε σειρά από πληκτρολογήσεις, τηλεφωνικούς παλμούς τόνου επαφής, μερικές παραγράφους κειμένου ή κάποια άλλη αναπαράσταση εξαρτάται εξ’ ολοκλήρου από τις απαιτήσεις της εφαρμογής.

Τα προϊόντα αναγνώρισης ομιλίας ποικίλουν ως προς τον τύπο και την περιοχή των μεταφράσεων που παράγουν. Αυτή είναι μία αποικόνιση του υλισμικού σχεδιασμού τους, του συνολικού σχεδιασμού, και τύπων των εφαρμογών για τις οποίες δημιουργήθηκαν. Η ικανότητα

της επικοινωνίας με το λογισμικό και το υλισμικό που εμπεριέχεται σε μία άλλη εφαρμογή είναι ένα θεμελιώδης κριτήριο αξιολόγησης προϊόντος.

Σχέση Μετάφρασης και Σημασίας.

Στα σύγχρονα εμπορικά συστήματα αναγνώρισης ομιλίας οι λέξεις δεν μεταφέρουν την έννοια που έχουν κατά την επικοινωνία μεταξύ των ανθρώπων. Η μετάφραση που αποδίδεται σε μία λέξη είναι απλά μία μετατροπή από την μία ψηφιακή αναπαραστασή στην άλλη, συνήθως από την αναπαράσταση με HMM, στην γραφημική.

Μερικοί κατασκευαστές συστημάτων αναγνωρίζουν τη δυνητική αξία της έννοιας της σημασίας σε επίπεδο λέξης (καλείται ***λεξιλογική σημασιολογία***) για την βελτίωση της ακρίβειας στην αναγνώριση. Η σημασία των λέξεων μπορεί για παράδειγμα να χρησιμοποιηθεί για την επιλογή της λέξης ανάμεσα στις υποψήφιες λέξεις με πανομοιότυπα ακουστικά πρότυπα. Μπορεί να χρησιμοποιηθεί για την μείωση του διαστήματος αναζήτησης, περιορίζοντας τον αριθμό των επιλεγομένων λέξεων. Μπορεί να βοηθήσει στην προφύλαξη από λάθη κατά την ομιλία, να ελέγχει τις διορθώσεις του ίδιου του ομιλητή και τέλος μπορεί να χρησιμοποιηθεί στο να κάνει ένα σύστημα να συμπεριφέρεται σε μεγαλύτερο βαθμό σαν ένα ανθρώπινο ον. Σημασιολογικές λεξιλογικές τεχνικές που εξετάζονται στα σημερινά συστήματα έρευνας κυμαίνονται από απλές κατηγορίες ταξινόμησης, όπως ΠΟΛΗ, ΟΝΟΜΑΤΑ, έως σε σύνθετες σημασιολογικές δομές που ορίζονται από γλωσσολογικές θεωρίες. Μερικές από τις απλούστερες μορφές θα ενσωματωθούν στα εμπορικά συστήματα αναγνώρισης στο προσεχές μέλλον.

Πολλαπλές Μεταφράσεις.

Δεν υπάρχουν “λογοπαίγνια” ή ασάφειες στην αναγνώριση ομιλίας. Ένα απλό πρότυπο ομιλίας μπορεί να έχει μία μόνο μετάφραση σε οποιοδήποτε σημείο στην εφαρμογή. Καμία ασάφεια δεν επιτρέπεται. Υπάρχει η ανάγκη να εξασφαλιστεί η γρήγορη και η ακριβής επεξεργασία.

Πολλά προϊόντα αναγνώρισης δεν επιτρέπουν περισσότερες από μία μεταφράσεις για την ίδια λέξη κατά την πορεία μίας εφαρμογής. Η λέξη “ένα” για παράδειγμα μπορεί να χρησιμοποιηθεί για να παράγει το ψηφίο 1 σε μερικά σημεία μίας εφαρμογής και τα γράμματα ***ε ν α*** σε κάποιο άλλο σημείο. Σε κάθε περίπτωση η μετάφραση που επιλέγεται θα πρέπει να είναι σαφής και ευκρινής.

Παραλλαγές Λέξεων

Σε ένα κανονικό λεξικό με την καταχώρηση για την λέξη "αναγνωρίζω" μπορεί να περιέχονται οι ακόλουθες πληροφορίες:

> Αναγνωρίζω ρήμα ΑΟΡΙΣΤΟΣ: αναγνώρισα, ΜΕΛΛΟΝΤΑΣ: θα αναγνωρίσω, ΤΡΙΤΟ ΠΡΟΣΩΠΟ: αναγνωρίζει, ΕΠΙΘΕΤΟ: ανγνωρίσιμο, ΕΠΙΡΗΜΑ: αναγνωρισμένα.

Ωστόσο, τα λεξικά αναγνώρισης ομιλίας απαιτούν ξεχωριστές καταχωρήσεις για κάθε μία από αυτές τις λέξεις. Αυτός ο περιορισμός είναι σύμφωνος με την ακουστική εστίαση των λεξιλογίων αναγνώρισης ομιλίας και έχει σχεδιαστεί για να ελαχιστοποιήσει την πολυπλοκότητα κατά την διαδικασία της αναγνώρισης. Επομένως, όταν ο κατασκευαστής ενός συστήματος αναγνώρισης λεξιλογίου δηλώνει ένα ***ολικό λεξιλόγιο*** σαράντα χιλιάδων λέξεων, σημαίνει ότι το σύστημα περιέχει σαράντα χιλιάδες μεμονωμένους αναγνωριστές λέξεων, μερικοί από τους οποίους μπορεί να είναι παραλλαγές λέξεων.

ΣΧΕΔΙΑΣΜΟΣ ΛΕΞΙΛΟΓΙΟΥ

Πριν το 1990 τα περισσότερα εμπορικά συστήματα αναγνώρισης πρόσφεραν μία από τις δύο δυνατότητες σχεδίασης λεξιλογίου:

- Λεξικά υλοποιημένα από τους κατασκευαστές
- Δημιουργία εφαρμογών σχεδιασμού λεξιλογίου σε επίπεδο λέξης

Από τότε οι επιλογές έχουν επεκταθεί ώστε να περιέχουν:

- Λεξικά υλοποιημένα από τους κατασκευαστές
- Δημιουργία εφαρμογών σχεδιασμού λεξιλογίου σε επίπεδο λέξης
- Δημιουργία εφαρμογών σχεδιασμού λεξιλογίου σε επίπεδο υπολέξης
- Δημιουργία λεξιλογίου από τελικούς χρήστες
- Αυτόματη εξαγωγή λεξιλογίου από το λεξιλόγιο
- Ένα συνδιασμό από τις παραπάνω εναλλακτικές προτάσεις.

Λεξικά Υλοποιημένα από τους Κατασκευαστές.

Τα λεξικά η υλοποίηση των οποίων έχει γίνει από τους κατασκευαστές, είναι συστήματα λεξιλογίου τα οποία έχουν κατασκευαστεί από τον ίδιο τον κατασκευαστή. Τα συστήματα μεγάλου λεξιλογίου, οι ετοιμοπαράδοτες εφαρμογές, τα συστήματα τα ενσωματωμένα σε κατάλληλο λογισμικό και τα συστήματα με χρήστες ανεξάρτητους ομιλητές συχνά περιέχουν

λεξικά φτιαγμένα από κατασκευαστές. Τα λεξικά τα οποία έχουν υλοποιηθεί από τους κατασκευαστές παίρνουν διάφορες μορφές, οι πιο συνήθεις από τις οποίες είναι:

- Τα λεξικά
- Εφαρμογές συγκεκριμένων λεξιλογίων

Λεξικά

Τα λεξικά συνήθως συναντόνται σε συστήματα μεγάλων λεξιλογίων και χρησιμεύουν ως εφεδρικοί πόροι σε εφαρμογές που αναπτύσονται χρησιμοποιώντας αυτά τα συστήματα. Το λεξικό ενός συστήματος αναγνώρισης μεγάλου λεξιλογίου μερικές φορές καλείται ***ολικό λεξιλόγιο*** του συστήματος. Είναι το τμήμα εκείνο όπου το σύστημα κωδικοποιεί λέξεις ή εκεί όπου βρίσκονται οι βασικές φόρμες. Τα πλήρη λεξικά μερικών εμπορικών συστημάτων περιέχουν περισσότερα από εκατό χιλιάδες μεμωνομένα στοιχεία λεξιλογίου καθένα από τα οποία αναπαρίσταται σαν μία ακολουθία από φωνήματα, τρίφθογγους, ή άλλες μονάδες που σχηματίζουν τις βασικές αρχές αναγνώρισης σε ένα σύστημα. Αναπαρίστανται σε ένα μη γραμμικό σχηματισμό, όπως αυτόν του πλέγματος.

Τα λεξικά θα πρέπει να σχεδιάζονται για επαρκή κάλυψη των εφαρμογών που πρόκειται να αναπτυχθούν από το σύστημα. Η αρχική επιλογή του λεξιλογίου για ένα λεξικό γενικά αποτελείται από τις λέξεις και τις φράσεις που θεωρούνται απαραίτητες για την λειτουργία της διάταξης αναγνώρισης και για την περιοχή των απαιτούμενων εφαρμογών. Μπορεί να περιέχουν συστήματα ελέγχου αναγνώρισης λέξεων και συνήθες λεξιλόγιο όπως τα ψηφία και οι ημέρες της εβδομάδας. Η συλλογή επιπρόσθετων μονάδων λεξιλογίου μπορεί να γίνει από βάσεις δεδομένων αναφοράς και άλλες on-line πηγές. Τέτοιες μονάδες μελετώνται για ορθογραφικά λάθη και εκτιμούνται με βάση την μετρική συχνότητα εμφάνισης.

Τα δεδομένα που χρησιμοποιούνται για την δημιουργία των μοντέλων των λέξεων πρέπει να αντανακλούν τον πληθυσμό των αναμενόμενων χρηστών. Για παράδειγμα τα συστήματα ανεξάρτητου ομιλητή τα οποία έχουν σχεδιαστεί για να χρησιμοποιηθούν από ομιλητές της αμερικανικής αγγλικής γλώσσας δεν θα έχουν ικανοποιητική απόδοση αν σχεδιαστούν χρησιμοποιώντας δεδομένα από ομιλητές της βρετανικής αγγλικής γλώσσας.

Μία άριστη πηγή για πρόσθετες πληροφορίες πάνω στον σχεδιασμό μεγάλου λεξιλογίου αποτελούν οι δημοσιεύσεις του Jelinel.

Λεξιλόγια για Συγκεκριμένες Εφαρμογές

Λεξιλόγια ετοιμοπαράδοτων εφαρμογών όπως η διεπαφή ομιλίας για ένα συγκεκριμένο προϊόν υπολογιστή, σαν το Excel, παρέχει λεξικά φτιαγμένα από κατασκευαστές προσαρμοσμένα σε αυτή την εφαρμογή. Στα ετοιμοπαράδοτα συστήματα με μικρό ή μεσαίο λεξιλόγιο, όπως είναι

η μετωπική ομιλία για την Microsoft Windows, το λεξιλόγιο που συνδέεται με την εφαρμογή αναπαριστά το ολικό λεξιλόγιο του συστήματος. Σαν αποτέλεσμα η διεπαφή ομιλίας είναι έτοιμη να χρησιμοποιηθεί σχεδόν αμέσως μετά την φόρτωση της. Τα ενσωματωμένα λεξιλόγια μειώνουν τον χρόνο ανάπτυξης της εφαρμογής διατηρώντας ένα υψηλό επίπεδο ποιότητας για τον σχεδιασμό του λεξιλογίου. Έχουν μεγάλη αποίχηση σε απλούς χρήστες υπολογιστών και σε ανθρώπους με αδυναμίες.

Στα συστήματα μεγάλου λεξιλογίου, με ενσωματωμένα λεξικά, μία συγκεκριμένη εφαρμογή λεξιλογίου αναπαριστά ένα υποσύνολο του συνολικού λεξιλογίου του συστήματος. Θα μπορούσε να ονομαστεί σαν το ***εσωτερικό λεξιλόγιο*** μίας εφαρμογής με σκοπό να διαφοροποιηθεί από το ολικό λεξιλόγιο του συστήματος που βρίσκεται στο λεξικό. Το εσωτερικό λεξιλόγιο γενικά φορτώνεται με την εφαρμογή. Το μέγεθος και η φύση του ενδημικού λεξιλογίου ποικίλλει ανάλογα με το σύστημα αναγνώρισης και την εφαρμογή. Μερικοί κατασκευαστές προσφέρουν αρκετές δυνατότητες επιλογής έτσι ώστε να ικανοποιήσουν διαφορετικές απαιτήσεις σε λεξιλόγιο (και σε κόστος) των πελατών τους.

Δημιουργία Λεξιλογίου από Κατασκευαστές Εφαρμογών.

Η δημιουργία λεξιλογίου από κατασκευαστές εφαρμογών απαιτεί ο κατασκευαστής της εφαρμογής να προσδιορίσει το αναγκαίο λεξιλόγιο για μία εφαρμογή και να το ορίσει για το σύστημα αναγνώρισης. Λίγο έως καθόλου λεξιλόγιο παρεχόμενο από τον πωλητή δίδεται. Αντί αυτού τα προϊόντα αναγνώρισης περιέχουν λεξιλόγιο και εργαλεία γαι την ανάπτυξη εφαρμογών. Τα περισσότερα συστήματα αναγνώρισης εξαρτημένου ομιλητή απαιτούν δημιουργία λεξιλογίου από σχεδιαστές εφαρμογών. Επίσης, και ένας αυξανόμενος αριθμός συστημάτων ανεξάρτητου ομιλήτη μικρού λεξιλογίου, παρέχουν εφαρμογές με εργαλεία ανάπτυξης λεξιλογίου.

Οι κατασκευαστές οι οποίοι χρησιμοποιούν μοντέλα υπολέξεων έχουν την τάση να δημιουργούν οι ίδιοι νέες μονάδες λεξιλογίου. Τα περισσότερα συστήματα επιτρέπουν στους κατασκευαστές εφαρμογών να απαιτούν νέο λεξιλόγιο, αλλά η ανάπτυξη υπολέξεων πραγματοποιείται από τους πωλητές. Τα συστήματα ομιλίας Inc (SSI) ήταν από τα πρώτα που έδωσαν την δυνατότητα στους κατασκευαστές εφαρμογών να προσθέτουν λέξεις στα δικά τους λεξικά. Παρείχαν μία λειτουργία μετατροπής της ορθογραφίας μίας νέας λέξης στην φωνητική της αναπαράσταση. Το ***HARK*** των BBN, ***Developer's Toolkit*** της Philips Dictation Systems και το ***Corona's Toolkit*** (Corona είναι το εμπορικά εκμεταλεύσιμο προϊόν της SRI) είναι παραδείγματα εμπορικών συστημάτων που βασίζονται σε μοντέλα υπολέξεων και επιτρέπουν στους κατασκεύαστες να παράγουν μία βασική φόρμα νέου λεξιλογίου δακτυλογραφώντας την φωνητική ορθογραφία των λέξεων. Εφόσον οι περιπλοκές στην φωνητική ορθογραφία μπορούν να προκαλέσουν συγχύσεις, οι κατακευαστές προέβησαν στην εφαρμογή άλλων μεθόδων για να επιτύχουν την προφορά των λέξεων, όπως να συνδέσουν τμήματα των υπάρχουσων υπολέξεων

που περιέχουν το ίδιο ηχητικό πρότυπο ή με το να βασίζουν την δημιουργία μίας λέξης στην ορθογραφία της.

Δημιουργία Λεξιλογίου από τους Τελικούς Χρήστες.

Η ανάπτυξη λεξιλογίου από τους τελικούς χρήστες είναι μία τεχνική για την προσαρμογή μίας εφαρμογής στις απαιτήσεις των πελατών (καλείται και ***εξατομίκευση της εφαρμογής***). Με τον συνυπολογισμό τέτοιων εργαλείων εκφράζεται η πεποίθηση ότι οι τελικοί χρήστες έχουν μοναδικές μεμονωμένες απαιτήσεις.

> Αρχικά επιτρεπόταν μόνο στους προγραμματιστές να εφαρμόσουν αυτή τη διαδικασία, μετέπειτα όμως έγινε αντιληπτό ότι είναι προτιμότερο οι χρήστες να προσθέτουν στοιχεία όπως νέα ονόματα (Elton Sherwin Manager of Speech Recognition Strategy & Market Development, Power Personal Systems, IBM [currently at Lexicus], personal communication, 1994).

Ένας δεύτερος λόγος για την ενσωμάτωση εργαλείων ανάπτυξης λεξιλογίου από τελικούς χρήστες είναι η τεράστια πλειονότητα των λαθών αναγνώρισης που γίνονταν από τα συστήματα αναγνώρισης μεγάλου λεξιλογίου, τα οποία είχαν σαν αποτέλεσμα την απώλεια λεξιλογίου. Εφόσον τα περισσότερα συστήματα δεν έχουν στο ενεργητικό τους την δυνατότητα ***εξωτερικής απόρριψης λεξιλογίου***, παράγουν μία αναγνώριση προσπαθώντας να βρουν ένα ταίριασμα μεταξύ των λέξεων στο υπάρχον λεξιλόγιο. Δεν είναι δυνατόν να προβλεφθούν όλες οι λέξεις που θα χρειαστούν σε κάθε χρήστη. Στην πραγματικότητα η εξωτερική είσοδος λεξιλογίου είναι ένα αρκετά συνήθες γεγονός σε συστήματα υπαγόρευσης, κάνοντας αναγκαία την δημιουργία εργαλείων λεξιλογίου για τελικούς χρήστες για συστήματα μεγάλου λεξιλογίου.

Ο τρόπος με τον οποίο οι τελικοί χρήστες μπορούν να προσθέσουν λεξιλόγιο σε ένα σύστημα μεγάλου λεξιλογίου ποικίλει ανάλογα με το σύστημα. Αν και υπάρχουν τα βασικά συστήματα υπολέξεων, το λεξιλόγιο των τελικών χρηστών εισάγεται σαν μοντέλο σε επίπεδο λέξης εξαρτημένου ομιλητή συνδέεται με το λεξιλόγιο του μεμονωμένου χρήστη που δημιούργησε την λέξη. Οι πωλητές καθιστούν ικανούς τους χρήστες να προσθέτουν λέξεις στο εσωτερικό λεξιλόγιο, σε πολλές περιπτώσεις όμως υπάρχουν ακόμη αναπαραστάσεις σε επίπεδο λέξης οι οποίες θα πρέπει να καταρτηστούν από κάθε χρήστη. Αυτές οι προσεγγίσεις αντικατοπτρίζουν τους περιορισμούς στις δυνατότητες αυτών των συστημάτων γιατί ο γενικός στόχος είναι να παραχθούν μοντέλα ανεξάρτητου ομιλητή για λεξιλόγιο που έχει προστεθεί από τον χρήστη.

Μερικά συστήματα μεγάλου λεξιλογίου σχεδιασμένα για ανάπτυξη εφαρμογών (όπως το Corona Toolkit, το HARK της BBN και το SpeechPro της Philips Dictation Systems) δίνουν στους κατασκευαστές την δυνατότητα πρόσθεσης νέων στοιχείων στο εσωτερικά αναπτυσόμενο λεξιλόγιο. Η αρχική έκδοση της εργαλειοθήκης της Philips είχε ένα ξεχωριστό σύνολο από

εργαλεία ανάπτυξης λεξιλογίου, κατασκευασμένα χρησιμοποιώντας την εργαλειοθήκη, για τους τελικούς χρήστες.

Ένας αριθμός μικρών αλλά σύνθετης δομής ετοιμοπαράδοτων εφαρμογών περιμένει από τους χρήστες να δημιουργήσουν τις νέες λέξεις σαν μέρος της λειτουργίας του συστήματος. Τα κυψελωτά τηλεφωνικά συστήματα κλίσεων είναι τα καλύτερα παραδείγματα τέτοιου είδους συστημάτων. Περιορισμένη κατασκευή λεξιλογίου τελικών χρηστών έχει αρχίσει να εμφανίζεται σε διεπαφές ομιλίας για Windows.

Οι κατασκευαστές αντιλαμβάνονται ότι η ικανότητα ικανοποίησης των αναγκών των τελικών χρηστών είναι υψηλής σημασίας για την τελική επιτυχία της τεχνολογίας αναγνώρισης ομιλίας. Συνεπώς κάθως τα συστήματα αναγνώρισης θα αυξάνουν σε ισχύ και ευλυγισία ο ρόλος των τελικών χρηστών στην ανάπτυξη λεξιλογίου θα αυξάνει.

Αυτόματη Αφαίρεση Λεξιλογίου.

Η αυτόματη αφαίρεση λεξιλογίου (μερικές φορές καλείται ***διερεύνηση εφαρμογής*** ή ***βελιστοποίηση λεξιλογίου***) συνεπάγεται την αυτόματη αφαίρεση λεξιλογίου από αρχεία ή on-line συστήματα. Αναπαριστά ένα πρώτο βήμα απέναντι στην αυτόματη ανάπτυξη λεξιλογίου.

Η αξία της αυτόματης αφαίρεσης λεξιλογίου για τους κατασκευαστές εφαρμογών είναι το ότι μεταφέρει την διαδικασία λεπτομερούς καθορισμού και ορισμού του λεξιλογίου της εφαρμογής από τον κατασκευαστή προς το σύστημα αναγνώρισης. Ο κατασκευαστής δεν χρειάζεται πλέον να ψάχνει για να καθορίσει το απαιτούμενο λεξιλόγιο της εφαρμογής και στην συνέχεια να ορίσει και τις μεταφράσεις του λεξιλογίου για το σύστημα. Για εφαρμογές μικρού λεξιλογίου η αυτόματη αφαίρεση λεξιλογίου απλοποιεί και καθιστά συντομότερη την διαδικασία ανάπτυξης. Αυτή η προσέγγιση μπορεί να χρησιμοποιηθεί μόνο για εφαρμογές που χρησιμοποιούν on-line συστήματα υπολογιστών, όπως οι διεπαφές ομιλίας για το λογισμικό συστημάτων υπολογιστών.

Οι κατασκευαστές συστημάτων αναγνώρισης μεγάλου λεξιλογίου στρέφονται στην αυτόματη αφαίρεση λεξιλογίου με σκοπό να παρέχουν ένα βαθμό εξατομίκευσης στο λεξικό. Προσφέρουν εργαλεία για την διερεύνηση του λεξιλογίου εγγράφων συγκρινόμενα με τα αντίστοιχα που παράγονται από χρήστες της εφαρμογής. Η συχνότητα εμφάνησης των στοιχείων του λεξιλογίου διαβαθμίζεται σύμφωνα με την συχνότητα εμφάνισης τους στα κείμενα, αλλά νέες λέξεις δεν προστίθενται στο λεξικό. Η Dragon Systems δημιούργησε το πρώτο σύστημα υπαγόρευσης το οποίο διαθέτει την τεχνική αυτόματης αφαίρεσης λεξιλογίου.

Η αυτόματη αφαίρεση λεξιλογίου αναπαριστά ένα άλλο τρόπο με τον οποίο οι κατασκευαστές συστημάτων αναγνώρισης προσπαθούν να απευθυνθούν στις ανάγκες των χρηστών τους. Είναι μία προσέγγιση με ολοένα αυξανόμενη σημασία και ζήτηση διάθεσης.

ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΛΕΞΙΛΟΓΙΟΥ

Το ***ενεργό λεξιλόγιο***, οι συγχεόμενες λέξεις, το αλφάβητο, και οι αριθμοί είναι ειδικά θέματα λεξιλογίου τα οποία θα συζητηθούν στις επόμενες ενότητες.

Ενεργό Λεξιλόγιο.

Το ενεργό λεξιλόγιο ενός συστήματος είναι το σύνολο των λέξεων τις οποίες η εφαρμογή επιτρέπει και αναμένει να ομιλούνται οποιαδήποτε στιγμή. Αποτελεί τις υποψήφιες λέξεις που το σύστημα θα αξιολογήσει κατά την διαδικασία της αναγνώρισης στην είσοδο. Μία εφαρμογή ελέγχου χρωμάτων μπορεί για παράδειγμα να επιτρέπει την ακόλουθη είσοδο:

Το χρώμα είναι [ξεφλουδισμένο γρατζουνισμένο μουντό ανομοιογενές]

όπου οποιαδήποτε από τις λέξεις ανάμεσα στις αγκύλες θα μπορούσε να ειπωθεί μετά την λέξη “είναι”. Οι λέξεις στις αγκύλες αναπαριστούν το ενεργό λεξιλόγιο για την περίπτωση αυτή. Το ενεργό λεξιλόγιο είναι σπανίως ισοδύναμο με το ολικό λεξιλόγιο και μερικές φορές μπορεί να είναι σχεδόν ισοδύναμο με το εσωτερικό λεξιλόγιο. Η έννοια του ενεργού λεξιλογίου συνδέεται με την χρήση της γραμματικής.

Συγχεόμενες Λέξεις.

Οι ***συγχεόμενες λέξεις*** είναι λέξεις που ηχούν πανομοιότυπα σε ένα σύστημα αναγνώρισης ομιλίας. Όταν χρησιμοποιούνται στο ίδιο ενεργό λεξιλόγιο, είναι πιθανή η αύξηση των λαθών κατά την διαδικασία της αναγνώρισης. Γενικά, οι μονοσύλλαβες λέξεις είναι περισσότερο συγχεόμενες από τις πολυσύλλαβες λέξεις γιατί περιέχουν λιγότερη ακουστική πληροφορία ώστε να βοηθηθεί το σύστημα. Όμως οι διακόσιες πιο συχνά χρησιμοποιούμενες λέξεις στην Αγγλική γλώσσα είναι μονοσύλλαβες λέξεις. Ανάμεσα σε αυτές υπάρχουν λειτουργικές λέξεις όπως “a” και “the” των οποίων η ακουστική ομοιότητα είναι υπεύθυνη για το πενήντα τις εκατό των λαθών στα συστήματα αναγνώρισης ομιλίας.

Μία συνήθης τεχνική για την βελτίωση στην ακρίβια αναγνώρισης δυναμικά συγχεόμενων λέξεων είναι να εισαχθεί ένα κατώφλι ομοιότητας για το ταίριασμα προτύπων ανάμεσα στις λέξεις που εισάγονται και στις αποθηκευμένες στο σύστημα. Αν το ταίριασμα που θα βρεθεί από το σύστημα αναγνώρισης αδυνατεί να προσεγγίσει το κατώφλι, τότε το σύστημα μπορεί να ζητήσει επιβεβαίωση ή επανάληψη από τον χρήστη. Μία άλλη τεχνική είναι η ***διεκμάθηση*** των συγχεόμενων λέξεων. Η τεχνική της διεκμάθησης είναι μία λειτουργία που παρέχεται από μερικά εμπορικά συστήματα αναγνώρισης. Περιλαμβάνει την αφαίρεση των ακουστικών προτύπων καθεμίας από τις λέξεις από την άλλη λέξη. Αυτή η διαδικασία αναδεικνύει τις διαφορές μεταξύ των λέξεων και ελαχιστοποιεί τις ομοιότητες. Όποτε είναι δυνατόν τα σύνολα των συγχεόμενων λέξεων θα πρέπει να αποφεύγονται. Αυτό δεν είναι πάντοτε

δυνατόν, ειδικά στα συστήματα μεγάλου λεξιλογίου, και όταν οι συγχεόμενες λέξεις αποτελούν μία φυσική συνάρθροιση λέξεων.

Παράλληλα με τις συγχεόμενες λέξεις στα συστήματα αναγνώρισης ομιλίας συναντάται και μία άλλη ομάδα λέξεων που ονομάζονται ***ακουστικά ασαφείς λέξεις*** και οι οποίες είναι δύσκολο να διακριθούν κατά την ομιλία. Παραδείγματα τέτοιων λέξεων είναι οι λέξεις "know" και "no" καθως και οι λέξεις "two", "to" και "too". Σε ακουστικό επίπεδο αυτές οι λέξεις είναι δύσκολο να προσδιοριστούν, εκτός και αν αναλυθούν με προσωδιακή λεπτότητα.

ΣΥΛΛΑΒΙΣΜΟΣ

Το 1989 η Voice Control Systems (VCS) ανέπτυξε και εξασφάλησε αποκλειστική συμμετοχή σε ένα εμπορικό αλγόριθμο για τον χειρισμό της προφερόμενης εισόδου. Σχεδιάστηκε για εφαρμογές που σχετίζονται με την τηλεφωνία. Το σύστημα ελαχιστοποίησε τα σφάλματα μέσω της χρήσης μίας βάσης δεδομένων από επιτρεπτές λέξεις οι οποίες περιόριζαν τον αριθμό των δυνατών ακολουθιών των γραμμάτων (βλέπε Foster & Schalk, 1993, κεφάλαιο 4). Ο αλγόριθμος συλλαβισμού περιείχε μερικές από τις τεχνικές διεκμάθησης διαθέσιμες στα VCS συστήματα εφαρμογών.

Από το 1993 και άλλοι κατασκευαστές άρχισαν να παρουσιάζουν ανάλογα κατασκευασμένα αλφαβητικά εργαλεία συλλαβισμού για μία περιοχή εφαρμογών. Πολλοί κατασκευαστές προσφέρουν το συλλαβισμό για εφαρμογές όπως το ταχυδρομείο φωνής, και τηλεφωνικές εφαρμογές. Η Amerigon χρησιμοποιεί τον αλγόριθμο αναγνώρισης αλφαβήτου των Lernout & Hauspie στο ελεγχόμενο μέσω της φωνής αυτοκίνητο. Αυτό το σύστημα καλείται AudioNav και βασίζεται στην συλλαβιστή είσοδο ονομάτων δρόμων για να σχηματίσει διαδρομές μεταξύ της τρέχουσας θέσης του χρήστη και του προορισμού του χρήστη. Μερικοί κατασκευαστές συστημάτων γραφείου όπως το SRI (και το Corona το εμπορικό του υποπροϊόν) και το BBN περιέχουν αλφαβητικά ονόματα γραμμάτων στα λεξικά τους σαν βασικά στοιχεία λεξιλογίου, αλλά οι περισσότεροι κατασκευαστές συστημάτων για εφαρμογές γραφείου δεν έχουν ολοκληρώσει τον αλφαβητικό σχηματισμό στα εμπορικά τους προϊόντα.

Το Αλφάβητο

Τα ονόματα των γραμμάτων σε ένα αλφάβητο αποτελούν το πιο επίπονο σύνολο συγχεόμενων λέξεων στην Αγγλική γλώσσα. Μία ομάδα ονομάτων γραμμάτων που είναι δύσκολη τόσο για τα συστήματα αναγνώρισης όσο και για τους ανθρώπους, έχει ένα ειδικό όνομα: λέγεται το ***Ε-σύνολο***. Το Ε-σύνολο αναφέρεται σε ονόματα γραμμάτων που έχουν σαν κατάληξη το φώνημα iy (συμπεριλαμβάνονται "b, c, d" και "g"). Οι δύο άλλες ομάδες με συγχεόμενα ονόματα γραμμάτων στην Αμερικανική Αγγλική γλώσσα είναι το ***α-σύνολο*** (συμπεριλαμβάνονται το "a" και το "h") και το ***eh-σύνολο*** το οποίο περιλαμβάνει τα "m, n, s" και "f".

Λόγω του ότι η δυνατότητα συλλαβισμού των λέξεων είναι μία σημαντική πλευρά πολλών εφαρμογών, μεγάλη προσοχή δόθηκε στην δημιουργία εργαλείων που να χειρίζονται τον συλλαβισμό. Το γεγονός ότι μόνο μερικά από τα εμπορικά συστήματα αναγνώρισης πραγματοποιούν αλφαβητικό συλλαβισμό δείχνει το επίπεδο δυσκολίας που χαρακτηρίζει την εφαρμογή αυτή. Μερικοί κατασκευαστές προτρέπουν τους χρήστες να το αντικαταστήσουν με το στρατιωτικό αλφάβητο ("alpha, bravo, charlie…") ή με ένα ανάλογο σύνολο από περισσότερο διακρίσημες λέξεις.

Οι κατασκευαστές εφαρμογών οι οποίοι εργάζονται με ένα προϊόν αναγνώρισης που δεν διαθέτει αλφαβητικό σχηματισμό μπορούν να χρησιμοποιούν μία τεχνική που αναπτύχθηκε από την Amerigon στο σύστημα ***AudioNav*** στα Media Laboratories του MIT, και από άλλους. Το σύστημα μειώνει την επίδραση του Ε-συνόλου χρησιμοποιώντας τόσο ακουστικά δεδομένα όσο και αποδεκτά πρότυπα συλλαβισμού. Όπως και οι περισσότερες άλλες υπάρχουσες προσεγγίσεις στον συλλαβισμό, έτσι και αυτή απαιτεί έναν κατάλογο αναφοράς. Η μέθοδος που χρησιμοποιήθηκε από τα Media Laboratories του MIT για τον διακριτό συλλαβισμό των γραμμάτων των κυρίων ονομάτων είναι βασισμένη σε έναν ***πίνακα σύγχυσης***, που δίνεται στο σχήμα 6. Ο πίνακας σύγχυσης αποικονίζει τα πρότυπα των συγχεόμενων γραμμάτων που βρέθηκαν κατά την δοκιμή του συστήματος αναγνώρισης στην συλλαβιζόμενη είσοδο. Το γράμμα "b" για παράδειγμα ανγνωρίστηκε σαν "a, b, d, e, p, v" και "z". Αν δεν υπάρχουν ονόματα στον κατάλογο αναφοράς που να αρχίζουν από "v" ή "z" τότε οι μόνες επιλογές που το σύστημα διατηρεί είναι "a, b, d, e" και "p". Κάθε γράμμα που ακολουθεί παρέχει ένα νέο σύνολο από εναλακτικές επιλογές που συνδιάζονται με τις προηγούμενες επιλογές και συγκρίνονται με τον κατάλογο των ονομάτων, εώς ότου να παραμείνει μία επιλογή. Εφόσον κάθε σύστημα αναγνώρισης και συσκευή εισόδου παρουσιάζει μοναδικά συγχεόμενα πρότυπα ο πίνακας που χρησιμοποιήθηκε από το MIT δεν μπορεί να θεωρηθεί ο πιο κατάλληλος. Κατά την δημιουργία των πινάκων και των αλγόριθμων είναι σημαντικό να υποβάλονται σε εκτεταμένες δοκιμές πριν τεθούν σε εφαρμογή.

Περισσότερα για τις εφαρμογές στα Media Laboratories του MIT μπορούν να βρεθούν στους Marx & Schmandt (1994). Μία άλλη εφαρμογή που χρησιμοποιεί νευρωνικά δίκτυα περιγράφεται από τον Fanty, και άλλους, (1992).

Πίνακας 1.

a→ ah	**h→ ah**	**n→ anrs**	**t→ dept**
b→ abdepvz	**i→ Iy**	**o→ lo**	**u→ qu**
c→ ctz	**j→ adgjktz**	**p→ cdepvz**	**v→ bdepvz**
d→ cdvz	**k→ adjkq**	**q→ qi**	**w→ fmnw**
e→ e	**l→ l**	**r→ iry**	**x→ sx**
f→ fx	**m→ mn**	**s→ fs**	**y→ y**
g→ gt			**z→ defnstvxz**

Σχήμα 6. Πίνακας σύγχυσης.

Αριθμοί

Πολλά συστήματα αναγνώρισης υποστηρίζουν ότι η έκφραση των αριθμών θα πρέπει να είναι μία ακολουθία από μεμονωμένα ψηφία. Με την χρήση τέτοιων συστημάτων ο χρήστης θα πρέπει να εισάγει τον αριθμό 1445 σαν " ένα τέσσερα τέσσερα πέντε". Αυτό είναι ένα επακόλουθο της ανάγκης των συστημάτων αναγνώρισης ομιλίας να περιορίσουν την επεξεργασία σε μία μετάφραση ανά λέξη. Αν η μετάφραση για το "χίλια" είναι "000" όταν ο ομιλητής λέει "μία χιλιάδα" το σύστημα να αποκριθεί με "1000". Εάν ο ομιλητής αναφέρει "μία χιλιάδα τέσσερα" τότε το σύστημα θα αποκριθεί με "10004".

Ο περιορισμός αυτός άρχισε να εγκαταλείπεται στις αρχές της δεκαετίας του 1990 όταν η αύξηση της ταχύτητας και της δύναμης του υλισμικού των υπολογιστών έκαναν ευκολότερη την χρήση μακρολέξεων για τον χειρισμό πιο σύνθετης δομικής ανάλυσης. Στα μέσα του 1994 οι φυσικοί αριθμοί αποτέλεσαν ένα στοιχείο πολλών εμπορικών συστημάτων, συμπεριλαμβανομένων της τεχνολογίας True Type της IBM, το Hark της BBN, την τεχνολογία των lernout & Hauspie και την εργαλειοθήκη της Corona.

Ένας κατασκευαστής εφαρμογών μπορεί να δημιουργήσει έναν αλγόριθμο φυσικών αριθμών ο οποίος να δέχεται αναγνωρισμένες σειρές από το προϊόν αναγνώρισης και να τις επαναναλύει. Η κατασκευή και η διαδικασία ελέγχου θα είναι παρόμοιες με εκείνες που απαιτούνται για την δημιουργία αλφαβητικής εισόδου.

ΑΠΟΤΙΜΩΝΤΑΣ ΤΙΣ ΑΠΑΙΤΗΣΕΙΣ ΣΕ ΛΕΞΙΛΟΓΙΟ ΣΕ ΜΙΑ ΕΦΑΡΜΟΓΗ

Η εκτίμηση του λεξιλογίου που χρειάζεται σε μία εφαρμογή θα πρέπει να περιέχει τις ακόλουθες θεωρήσεις:

- Μία ενημερωμένη εκτίμηση του μεγέθους του λεξιλογίου
- Προσεχτικό ταίριασμα μεταξύ του απαιτούμενου λεξιλογίου για την εφαρμογή και του προϊόντος
- Καθορισμός αναφορικά με το πως το λεξιλόγιο της εφαρμογής θα αυξηθεί.

Μέγεθος του Λεξιλογίου

Η απόδοση και η ταχύτητα ενός συστήματος αναγνώρισης μειώνεται με την αύξηση στο μέγεθος του λεξιλογίου μίας εφαρμογής. Μερικοί ερευνητές ομιλίας εκτιμούν ότι η δυσκολία του προβλήματος αναγνώρισης αυξάνει λογαριθμικά με την αύξηση του μεγέθους του λεξιλογίου, γεγονός που δεν οφείλεται στην αύξηση του αριθμού των λέξεων, αλλά στην αυξημένη πολυπλοκότητα που χαρακτηρίζει την διαδικασία αναγνώρισης, όταν γίνεται χρήση συστήματος μεγάλου λεξιλογίου.

Γενικά τα συστήματα ή οι αλγόριθμοι αναγνώρισης ομιλίας ταξινομούνται σαν συστήματα μικρού, μεσαίου, ή μεγάλου λεξιλογίου. Στην βιβλιογραφία υπάρχει κάποια ποσοτικοποίηση σε ότι αφορά τους όρους αυτούς, αλλά με βάση εμπειρικά δεδομένα τα μικρά λεξιλόγια είναι αυτά τα οποία το μέγεθος τους κυμαίνεται στην περιοχή των 1-99 λέξεων, τα μεσαία λεξιλόγια από 100-999, και τα μεγάλα λεξιλόγια από 1000 και περισσότερες λέξεις. Εφόσον έχουν σχεδιαστεί συστήματα αναγνώρισης τα οποία υποστηρίζουν λεξιλόγιο της τάξης των 200000 λέξεων, ένα λεξιλόγιο των χιλίων λέξεων θα μπορούσε να θεωρηθεί “μικρό” σε κάποιες περιπτώσεις. Έτσι θα πρέπει να είμαστε προσεκτικοί σε ότι αφορά την σημασία αυτών των όχι αυστηρά ορισμένων όρων ταξινόμησης για το μέγεθους του λεξιλογίου. Τα συστήματα μικρού λεξιλογίου, όπως ορίστηκαν παραπάνω, είναι ευρέως διαθέσιμα και χρησιμοποιούνται σε εφαρμογές που αφορούν πιστωτικές κάρτες, στην τηλεφωνική αναγνώριση κλίσεων και σε συστήματα ταξινόμησης ναυτιλιακών εφαρμογών. Η εφαρμογή των συστημάτων μεσαίου λεξιλογίου εστιάζεται σε πειραματικά εργαστηριακά συστήματα αναγνώρισης συνεχούς ομιλίας. Τα συστήματα μεγάλου λεξιλογίου χρησιμοποιούνται σε εμπορικά προϊόντα που αφορούν εφαρμογές όπως η αλληλογραφία μεταξύ γραφείων και η διόρθωση εγγράφων.

Παρά την ιδιαίτερη αίγλη που έχουν τα συστήματα μεγάλου λεξιλογίου, οι περισσότερες εφαρμογές περιέχουν λεξιλόγιο πολύ μικρότερο από ότι θα περίμενε κανείς. Για παράδειγμα, η κλίση ενός τηλεφώνου αυτοκινήτου απαιτεί λιγότερες από είκοσι λέξεις, και οι περισσότερες

εφαρμογές χρησιμοποιούν λιγότερες από εκατό λέξεις. Ακόμη και το λεξιλόγιο που χρησιμοποιείται σε εφαρμογές υπαγόρευσης δεν είναι απεριόριστο. Το αρχικό σύστημα παραγωγής ακτινολογικών αναφορών που κατασκευάστηκε από τον Kurzweil AI, για παράδειγμα, απαιτούσε μόνο πέντε χιλιάδες λέξεις.

Οι Απαιτήσεις σε Λεξιλόγιο σε Σχέση με ένα Προϊόν Αναγνώρισης

Τα συστήματα μεγάλου λεξιλογίου είναι πολύ ενδιαφέροντα, αλλά είναι σχεδιασμένα για εφαρμογές υπαγόρευσης και μπορεί να μην ανταποκρίνονται ικανοποιητικά στην είσοδο δεδομένων και σε εφαρμογές ελέγχου συστημάτων μικρού λεξιλογίου.

Αν το λεξιλόγιο μίας εφαρμογής είναι μεγάλο και σε μεγάλο βαθμό διαφοροποιημένο τότε θα πρέπει να υπάρχει ένας υψηλά συνδετικός παράγοντας. Σε αυτή την περίπτωση ένα σύστημα αναγνώρισης θα πρέπει να εκτιμηθεί από το πόσο ευαίσθητο είναι σε μεγάλους αριθμούς από επιλογές. Ισχυρισμοί σχετικά με το μέγεθος του λεξιλογίου είναι αβάσιμοι αν το σύστημα απαιτεί το μέγεθος του εσωτερικού και του ενεργού λεξιλογίου να είναι πολύ μικρό.

Αν η εφαρμογή απαιτεί μεταθέσεις ανάμεσα σε υπολεξιλόγια ή μεταξύ του βασικού και ενός ειδικού λεξιλογίου του συστήματος αναγνώρισης, ο χρόνος που απαιτείται για την μετάθεση από το ένα υπολεξιλόγιο στο άλλο θα πρέπει να εκτιμηθεί.

Αν ένα σύστημα με ενσωματωμένο λεξιλόγιο είναι κατάλληλο, είναι σημαντικό να ελεγθεί ότι το λεξιλόγιο του ταιριάζει με αυτό που χρειάζεται στην εφαρμογή. Αν οι διαφορές είναι εκτεταμένες πιθανόν το σύστημα αναγνώρισης να μην είναι κατάλληλο για την συγκεκριμένη εφαρμογή. Αν κανένα σύστημα αναγνωρίσης δεν καλύπτει τις απαιτήσεις σε λεξιλόγιο της εφαρμογής, ο κατασκευαστής θα πρέπει να δημιουργήσει νέο λεξιλογιο.

Τα δεδομένα που χρησιμοποιούνται για την δημιουργία των μοντέλων στα λεξικά θα πρέπει να αντικατοπτρίζουν τα χαρακτηριστικά ομιλίας του αναμενόμενου πληθυσμού χρηστών. Αυτό είναι ιδιαίτερα σημαντικό για μοντέλα ανεξάρτητου ομιλητή. Μοντέλα αναφοράς που δημιουργούνται χρησιμοποιώντας δείγματα από άνδρες δεν είναι κατάλληλα για γυναικίους πληθυσμούς χρηστών. Μοντέλα που βασίζονται πάνω σε δεδομένα από ομιλητές Αμερικανικών Αγγλικών δεν θα πρέπει να χρησιμοποιούνται από ομιλιτές των Αυστραλέζικων Αγγλικών. Ακόμη και μοντέλα που έχουν κατασκευαστεί από δείγματα που προέρχονται από κατοίκους της Βοστώνης μπορεί να μην ερμηνεύονται ικανοποιητικά από κατοίκους στην περιοχή του Μισισιπή. Κάθε φορά που προκύπτουν ερωτήσεις για την καταλληλότητα των μοντέλων αναφοράς για ένα νέο πληθυσμό από χρήστες, τα μοντέλα θα πρέπει να δοκιμάζονται ξανά, και εάν υπάρχει κάποιο πρόβλημα, ο κατασκευαστής θα πρέπει να σκιαγραφήσει τις μεθόδους που θα πρέπει να χρησιμοποιηθούν για να τροποποιηθούν τα μοντέλα.

Αναπτύσοντας το Λεξιλόγιο

Σε προηγούμενο τμήμα εξετάστηκαν ποικίλες μέθοδοι σχεδιασμού λεξιλογίου. Αυτές οι μέθοδοι εφαρμόζονται επίσης για να επεκτείνουν ένα ήδη υπάρχον λεξιλόγιο. Μια σημαντική ερώτηση για την ανάλυση του θέματος αυτού είναι πότε οι τελικοί χρήστες χρειάζεται να προσθέσουν λεξιλόγιο στην εφαρμογή. Αν είναι πιθανή η πρόσθεση λεξιλογίου ο κατασκευαστής θα πρέπει να καθορίσει πως αυτή θα επιτευχθεί. Αν η ανάπτυξη του λεξιλογίου γίνεται με την προσθήκη λεξιλογίου που παρέχεται από τον κατασκευαστή της εφαρμογής, οι χρήστες θα πρέπει να πληροφορηθούν για τα νέα στοιχεία που έχουν προστεθεί και μπορεί να είναι αναγκαία η προσθήκη δειγμάτων ομιλίας των νέων στοιχείων πριν αυτά χρησιμοποιηθούν.

Η ΓΡΑΜΜΑΤΙΚΗ ΤΗΣ ΓΛΩΣΣΑΣ

Στο τμήμα αυτό θα γίνει σε θεωρητική βάση παρουσίαση γλωσσικών μοντέλων που χρησιμοποιούνται στις διάφορες εφαρμογές αναγνώρισης ομιλίας. Η γλωσσική επεξεργασία έχει να κάνει με την αναγνώριση ενός μεγάλου προτύπου (πρότασης) μέσω της ανάλυσης του σε μικρά υποπρότυπα σύμφωνα με κάποιους κανόνες.

Η Έννοια της Γραμματικής

Η ***γραμματική*** μίας γλώσσας αφορά τον τρόπο με τον οποίο τα σύμβολα μίας γλώσσας σχετίζονται μεταξύ τους για να σχηματίσουν μονάδες μηνυμάτων. Η γραμματική περιλαμβάνει τους κανόνες εκείνους σύμφωνα με τους οποίους τα ***σύμβολα*** μίας γλώσσας, τα φωνήματα στην περίπτωση μας, συδέονται μεταξύ τους. Τα σύμβολα μίας γλώσσας είναι οι θεμελιώδης μονάδες από τις οποίες όλα τα μημύματα συντίθενται. Η έννοια της ***γλώσσας*** ορίζεται σαν το σύνολο όλων των ***τερματικών ακολουθιών*** που μπορούν να παραχθούν με βάση τους κανόνες που υπαγορεύει η γραμματική της γλώσσας.

Τυπικά η έννοια της γραμματικής μίας γλώσσας ορίζεται σαν μία συνάρτηση τεσσάρων συνόλων ως εξής:

$$\mathbf{G = (V_n, V_t P, S)}$$

όπου $\mathbf{V_n}$ και $\mathbf{V_t}$ είναι ***μη τερματικά*** και ***τερματικά λεξιλόγια*** (πεπερασμένα σύνολα), **P** είναι ένα πεπερασμένο σύνολο από ***κανόνες παραγωγής*** και **S** είναι το ***αρχικό σύμβολο*** για κάθε παραγόμενη ακολουθία συμβόλων. Τα σύνολα $\mathbf{V_n}$ και $\mathbf{V_t}$ έιναι ασύνδετα και η ένωση τους, έστω **V,** καλείται ***λεξιλόγιο*** της γλώσσας.

Περιπλοκή της Γλώσσας

Συγκρίνοντας την απόδοση των συστημάτων αναγνώρισης ομιλίας, είναι σημαντικό το να μπορούμε να ποσοτικοποιήσουμε την δυσκολία του προς αναγνώριση θέματος. Οι ***γλωσσικοί περιορισμοί*** έχουν σαν σκοπό να μειώσουν την αβεβαιότητα του περιεχομένου των προτάσεων

και να διευκολύνουν την διαδικασία της αναγνώρισης. Η έννοια των γλωσσικών περιορισμών έχει να κάνει με το πως οι θεμελιώδεις μονάδες, είτε αυτές θεωρούνται ότι είναι οι φθόγγοι, τα φωνήματα, οι συλλαβές ή οι λέξεις, συνδέονται μεταξύ τους, με ποια διάταξη, με ποιο περιεχόμενο, και με ποια προσδοκόμενη σημασία. Έτσι, για παράδειγμα, αν κατά μέσο όρο υπάρχουν πολύ λίγες λέξεις που ακολουθούν μία δοθείσα λέξη σε μία γλώσσα, τότε η διάταξη αναγνώρισης θα έχει λιγότερες περιπτώσεις να ελένξει, με αποτέλεσμα η απόδοση της διαδικασίας αναγνώρισης να είναι καλύτερη από την περίπτωση που θα υπήρχαν πολλές λέξεις. Το παράδειγμα αυτό δηλώνει ότι ένα κατάλληλο μέτρο δυσκολίας για κάθε γλώσσα, μπορεί να περιέχει κάποιο μέτρο του μέσου όρου των τερματικών που είναι δυνατό να ακολουθούν κάθε δοθέν τερματικό. Αν η γλώσσα θεωρηθεί σαν ένα διάγραμμα όπου τα τερματικά συσχετίζονται με τις μεταβάσεις, τότε αυτό το μέτρο θα σχετίζεται με το μέσο ***συνδετικό παράγοντα*** σε κάθε σημείο απόφασης στο γράφο. Προσεγγιστικά μιλώντας, αυτή η ποσότητα μετράται με το μέγεθος της ***"περιπλοκής"***.

Bottom-Up Ανάλυση σε Αντιδιαστολή με την Top-Down Ανάλυση

Σύμφωνα με τον ορισμό της έννοια της γλώσσας θεωρούμε ότι μία πρόταση είναι μία πλήρης έκφραση και ότι τα φωνήματα αποτελούν τα θεμελιώδη σύμβολα σε μία γλώσσα. Το σχήμα 7 αποικονίζει ένα παράδειγμα το οποίο υπόκειται στην παραπάνω θεώρηση. Από το σχήμα είναι δυνατή η εξαγωγή συμπαιρασμάτων για το πως από τα φωνήματα συντίθενται οι λέξεις, για το πως οι λέξεις ταξινομούνται σε μέρη του λόγου, πως τα μέρη του λόγου σχηματίζουν φράσεις, και πως οι φράσεις σχηματίζουν προτάσεις. Αντίστροφα μπορούμε να δούμε πως μία πρόταση μπορεί να αναλυθεί στα σύμβολα από τα οποία απαρτίζεται με βάση ένα σύνολο από κανόνες. Η διαδικασία με την καθορίζεται το αν υπάρχει ένα σύνολο από κανόνες σε μία γραμματική για την σύνθεση (σε τερματικά) ή για την αποσύνθεση (σε τερματικά) μίας πρότασης καλείται ***συντακτική ανάλυση της γλώσσας***. Ο αλγόριθμος ο οποίος χρησιμοποιεί τους γραμματικούς κανόνες στην κατεύθυνση:

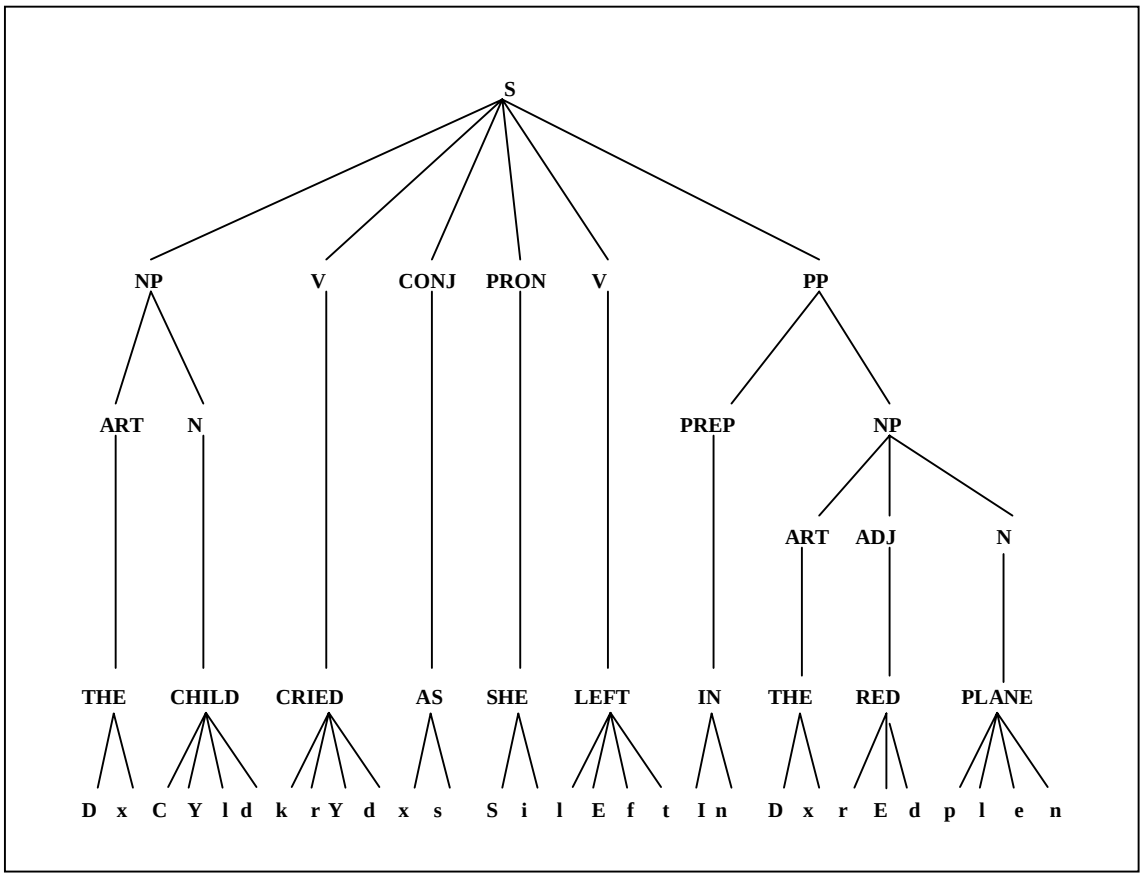
Τερματικό → Πρόταση

καλείται ***Bottom-Up συντακτικός αναλυτής***. Αντίθετα αν οι κανόνες χρησιμοποιούνται για αποσύνθεση,

Πρόταση → Τερματικό

ο αλγόριθμος καλείται ***Top-Down συντακτικός αναλυτής***.

Το τμήμα της διάταξης αναγνώρισης το οποίο μετατρέπει τα ακουστικά δεδομένα σε γλωσσικά σύμβολα καλείται ***ακουστικός αποκωδικοποιητής (AD).*** Ας υποθέσουμε ότι ο AD έχει επεξεργαστεί τη δοθείσα έκφραση και έχει αποφασίσει για το σύνολο των φθόγγων (σύμβολα ή τερματικά). Στην συνέχεια ο ***γλωσσικός αποκωδικοποιητής (LD)*** εφαρμόζει τους γλωσσικούς περιορισμούς. Δουλεύοντας από κάτω προς τα πάνω ο LD μπορεί να προσδιορίσει εάν το σύνολο των φθόγγων αντιστοιχεί σε μία κανονική πρόταση της γλώσσας. Το γεγονός αυτό μπορεί να αποτρέψει την αποδοχή λανθασμένων συνδιασμών από φθόγγους οι οποίοι έχουν θεωρηθεί από τον AD.



Σχήμα 7. Παραγωγή της έκφρασης "The child cried as left in the red plane" σύμφωνα με τους γραμματικούς κανόνες. Το μη τερματικό λεξιλόγιο για αυτό το παράδειγμα αντιστοιχεί σε φράσεις, μέρη του λόγου, και λέξεις: NP = ουσιαστική φράση, PP = προθετική φράση, V = ρήμα, CONJ = σύνδεσμος PRON = αντωνυμία, ART = άρθρο, PREP = πρόθεση, ADJ = επίθετο. Τα τερματικά είναι φωνητικά σύμβολα.

Κατά την Top-Down ανάλυση, οι γραμματικοί κανόνες αποτρέπουν την θεώρηση μη κανονικών ακολουθιών από σύμβολα, αλλά σε αυτή την περίπτωση κανένα συμβολό δεν έχει προέλθει υποθετικά. Στην περίπτωση αυτή, οι γραμματικοί κανόνες χρησιμοποιούνται για να περιορίσουν τον αριθμό των δυνατών συνδιασμών των συμβόλων που πρέπει να ληφθούν υπόψη από την διάταξη αναγνώρισης σε ακουστικό επίπεδο. Η διαδικασία ξεκινά από την κορυφή, όπου

ο LD παίρνει απόφαση για μία πρόταση που ανήκει στην γλώσσα. Οι γραμματικοί κανόνες χρησιμοποιούνται για να εξαχθεί ένα σύνολο από φθόγγους που να αντιστοιχούν στην πρόταση που αποφασίστηκε από τον LD. Η διαδικασία παράγει επίσης μία εκ των προτέρων πιθανότητα εμφάνησης των παραγόμενων φθόγγων σύμφωνα με την στατιστική δομή της γραμματικής της γλώσσας. Η συνολική πιθανότητα για την ακολουθία των φθόγγων υπολογίζεται στην συνέχεια από τον AD. Συμπαιρασματικά, ο LD χρησιμοποιεί την γραμματική μόνο για να εξάγει συμπεράσματα σχετικά με τους φθόγγους της πρότασης που επιλέγεται ως πιθανότερη και θα απορίψει αυτή την πρόταση στην περίπτωση που η πιθανότητα εμφάνισης της ακολουθίας των φθόγγων που προκύπτει είναι μικρή.

Το κύριο μειωνέκτημα της Bottom-Up ανάλυσης είναι ότι μία πρόταση δεν μπορεί να αναγνωριστεί παρα μόνο όταν για καθένα από τα σύμβολα της ο AD λάβει κάποια απόφαση. Κατά την Bottom-Up ανάλυση δεν γίνεται χρήση των γλωσσικών περιορισμών στην αποκωδικοποίηση του ακουστικού σήματος.

Άλλες Κανονικές Γραμματικές.

Γενικά κάθε τυπική γραμματική μπορεί να χρησιμοποιηθεί για την μοντελοποίηση της γλώσσας σε ένα γλωσσικό αποκωδικοποιητή LD. Σύμφωνα με τον Levinson (1985) κάθε πεπερασμένη γλώσσα μπορεί να παραχθεί από μία κανονική γραμματική, αλλά το κίνητρο για την χρήση κάποιας άλλης γραμματικής είναι η προσαρμογή του γλωσσικού μοντέλου σε ένα πιο συμβατικό μοντέλο της φυσικής γλώσσας. Για παράδειγμα οι φυσικοί γλωσσικοί κανόνες συχνά παρουσιάζονται σε ευαίσθητη προς το περιεχόμενο μορφή. Το μειωνέκτημα της χρήσης ανώτερης γραμματικής είναι η αυξημένη πολυπλοκότητα που συναντάται στους αντίστοιχους αλγόριθμους ανάλυσης. Έτσι ο αριθμός των πράξεων που απαιτείται κατά την αποκωδικοποίηση με χρήση του αλγόριθμού Viterbi για μία ακολουθία *w*, χρησιμοποιώντας κανονική γραμματική, είναι ανάλογος του $|V_n| \times |w|$, όπου $|V_n|$ είναι το μέγεθος του μη τερματικού λεξιλογίου και $|w|$ το μήκος της δοθείσας πρότασης.

Γενικά οι γλώσσες ελεύθερου περιεχομένου αναλύονται με την χρήση του αλγόριθμου Cocke-Younger-Kasami (CYK) ή με τον αλγόριθμο του Earley. Ο αλγόριθμος CYK αρχικά κατασκευάστηκε από τον Cocke, αλλά δημοσιεύτηκε αναξάρτητα από τον Kasami (1965) και τον Younger (1967). Η μέθοδος του Earley μερικές φορές καλείται αλγόριθμος ανάλυσης διαγράμματος και δημοσιεύτηκε το 1970. Ο CYK αλγόριθμος είναι μία προσέγγιση δυναμικού προγραμματισμού DP, ενώ αντίθετα ο αλγόριθμος Earley χρησιμοποιεί μία κύρια δομή δεδομένων που καλείται ***διάγραμμα*** για τον αποτελεσματικό συνδιασμό ενδιάμεσων υποαναλύσεων με σκοπό την ελάττωση των πλεονάζοντων υπολογισμών. Κάθε αλγόριθμος απαιτεί $O(|w^3|)$ πράξεις, ενώ η μέθοδος του Earley απαιτεί $O(|w^2|)$ πράξεις στην περίπτωση που

δεν υπάρχουν ασάφειες στην γραμματική. Πιο πρόσφατα ο Paeseler δημοσίευσε μία τροποποίηση της μεθόδου του Earley η οποία χρησιμοποιεί την διαδικασία αναζήτησης δέσμης με σκοπό την μείωση της πολυπλοκότητας, έτσι ώστε να είναι ανάλογη του μήκους της ακολουθίας εξόδου.

Η **ανάλυση από αριστερά προς τα δεξιά (LR)** είναι ένας αποτελεσματικός αλγόριθμος για την ανάλυση γλωσσών ελεύθερου περιεχομένου, ο οποίος αρχικά κατασκευάστηκε για γλώσσες προγραμματισμού. Ένας γενικευμένος αλγόριθμος ανάλυσης από αριστερά προς τα δεξιά εφαρμόστηκε στο πρόβλημα της αναγνώρισης συνεχούς ομιλίας. Το προκύπτον σύστημα καλείται HMM-LR γιατι βασίζεται σε ανάλυση των φθόγγων με χρήση HMM συνοδευόμενη με ανάλυση από αριστερά προς τα δεξιά με χρήση πρόβλεψης.

Για το πρόβλημα της αναγνώρισης συνεχούς ομιλίας έχουν επίσης χρησιμοποιηθεί ειδικοί τύποι γραμματικής. Η γραμματικές **επαυξημένου δικτύου μεταβάσεων (ATN)** είναι όμοιες με τις γραμματικές ελεύθερου περιεχομένου αλλά πιο αποτελεσματικές λόγω του ότι συγχωνεύουν κοινά μονοπάτια ανάλυσης. Οι γραμματικές αυτές κατασκευάστηκαν ειδικά για την επεξεργασία της φυσικής γλώσσας. Μία ATN γραμματική χρησιμοποιήθηκε στο σύστημα “HWIM” σε συνδιασμό με την μέθοδο “island-driven” στην οποία οι φθόγγοι, οι λέξεις ή οι φράσεις προσδιορίζονται πραγματοποιώντας μία αρχική διερεύνηση, και στην συνέχεια ενσωματώνονται χρησιμοποιώντας την αναζήτηση “middle-out”. Η μέθοδος αυτή αποτελεί μία πρωτότυπη απόκλιση από την συμβατική από τα αριστερά προς τα δεξια ανάλυση. Οι στοχαστικές ***ενοποιητικές γραμματικές*** αποτελούν γενικεύσεις των τυπικών γραμματικών στις οποίες τα χαρακτηριστικά προστίθενται στα στοιχεία του τυπικού λεξιλογίου. Χρησιμοποιήθηκαν στην επεξεργασία ομιλίας για να μοντελοποιήσουν την πληροφορία και να προσθέσουν τα χαρακτηριστικά της φυσικής γλώσσας στα μη τερματικά στοιχεία της γραμματικής. Ο συνυπολογισμός των χαρακτηριστικών της πληροφορίας στην γραμματική αποτελεί ένα βήμα προς την κατανόηση της ομιλίας σε ότι αφορά την παροχή γλωσσικής γνώσης πέραν της γραμματικής δομής. Στην δημοσίευση τους οι Hemphill και Picone παρουσιάζουν τον βασικό φορμαλισμό της ενοποιητικής γραμματικής και υποστηρίζουν ότι το να θεωρηθεί ότι η διαδικασία παραγωγής ομιλίας είναι βασισμένη στην γραμματική παρά σε ένα FSA έχει υπολογιστικά πλεονεκτήματα όταν ένας αλγόριθμος ανάλυσης διαγράμματος χρησιμοποιείται για να παράγει τις υποθέσεις.

ΕΠΙΣΚΟΠΙΣΗ ΤΩΝ ΣΥΣΤΗΜΑΤΩΝ ΑΝΑΓΝΩΡΙΣΗΣ ΟΜΙΛΙΑΣ.

Στο τμήμα αυτό πραγματοποιείται μία αναφορά στα συστήματα αναγνώρισης ομιλίας, σε εκείνα που διαδραμάτισαν κάποιο σημαντικό ρόλο στο παρελθόν αλλά και σε αυτά που πλαισιώνουν το παρόν. Σε τέτοια επισκόπηση δεν είναι δυνατόν να αναφερθεί το σύνολο των συστημάτων και των τεχνικών που έχουν παρουσιαστεί και υλοποιηθεί σε όλη την διάρκεια ανάπτυξης των συστημάτων αναγνώρισης ομιλίας. Σκοπός μας στο τμήμα αυτό είναι η

παρουσίαση μερικών συστημάτων που αντικατοπτρίζουν διαφορετικές προσεγγίσεις και ιδέες που έχουν παρουσιαστεί μέχρι σήμερα. Ακόμη θα παρουσιαστούν συγκριτικοί πίνακες της απόδοσης των διαφόρων συστημάτων. Τέλος να σημειωθεί ότι η όλη αναφορά γίνεται κυρίως σε ερευνητικά συστήματα και αποφεύγεται η συζήτηση πάνω σε εμπορικά συστήματα.

ARPA Πρόγραμμα Κατανόησης Ομιλίας. Στις Ηνωμένες Πολιτείες της Αμερικής η περίοδος των αυτόματων συστημάτων αναγνώρισης ομιλίας μεγάλης κλίμακας εισήχθει με το επονομαζόμενο πρόγραμμα **ARPA (Advanced Research Projects Agency)** από το τμήμα αμύνας, όταν το 1971, ανακοινώθηκε ένα πενταετούς διάρκειας πρόγραμμα με σκοπό την ανάπτυξη του τομέα κατανόησης ομιλίας. Οι στόχοι που ετέθησαν από το πρόγραμμα ARPA για την κατασκευή ενός προτώτυπου συστήματος, καθώς και τα χαρακτηριστικά του συστήματος HARPY του πανεπιστημίου Carnegie-Mellon, το μόνο σύστημα που ικανοποιούσε τους συγκεκριμένους στόχους δίδονται στον πίνακα 2.

Πίνακας 2.

Πενταετείς Στόχοι του Προγράμματος ARPA	**Χαρακτηριστικά του προγράμματος HARPY**
να δέχεται συνδεδεμένη ομιλία	ναι
από πολλούς	5 ομιλητές (3 γέννους αρσενικού, 2 θυληκού)
συνεργαζόμενους ομιλητές	Ναι
σε ένα μη θορυβώδες δωμάτιο	αίθουσα τερματικών υπολογιστών
χρησιμοποιώντας ένα καλό μικρόφωνο	κλειστό μικρόφωνο ομιλίας
με μικρού βαθμού ρύθμιση ανά ομιλητή	20 προτάσεις εκμάθησης ανά ομιλητή
δεχόμενο 1000 λέξεις	1011 λέξεις
χρησιμοποιώντας τεχνιτό συντακτικό	μέσος παράγοντας διακλάδωσης = 33
για ένα περιορισμένο θέμα	ανάκτηση εγγράφου
παρέχει λιγότερο από 10% σημασιολογικό λάθος	5% σημασιολογικό λάθος
σε λίγες φορές σε πραγματικό χρόνο	80 φορές σε πραγματικό χρόνο
με μία 100-MIPS συσκευή	0.4 MIPS PDP_KA10 χρησιμοποιώντας 256K 36-bit λέξεων με κόστος 5$ ανά επεξεργαζόμενη πρόταση

Σχήμα 8. Οι στόχοι του προγράμματος ARPA για το προτώτυπο σύστημα αναγνώρισης ομιλίας, σε συνδιασμό με τα χαρακτηριστικά του συστήματος HARPY του πανεπιστημίου Carnegie-Mellon.

Το 1977 ο Klatt έγραψε μία αναφορά για το πρόγραμμα ARPA η οποία συγκρίνει και αντιπαραβάλλει τις αρχιτεκτονικές και τις ιδιότητες λειτουργίας τεσσάρων συστημάτων που προέκυψαν από την μελέτη. Στον πίνακα 3 παρατίθενται η συνολική απόδοση των συστημάτων με βάση τα χαρακτηριστικά που δόθηκαν στον Πίνακα 2. Να σημειωθεί ότι η περιπλοκή χρησιμοποιείται σαν μέτρο της δυσκολίας αναγνώρισης ενός θέματος. Όπως αναφέρεται από τον Klatt με δεδομένο τον διαφορετικό παράγοντα διακλάδωσης μεταξύ των συστημάτων είναι δύσκολο να καθοριστούν απόλυτα οι διαφορές στην απόδοση μεταξύ των συστημάτων. Τα τέσσερα αυτά συστήματα παρέχουν διαφορετικές προσεγγίσεις, αλλά όλα υιοθετούν μία top-down μορφή επεξεργασίας.

Πίνακας 3

Σύστημα	**Κατανοηθήσες Προτάσεις (%)**	**Περιπλοκή**
CMU HARPY	95	33
CMU HEARSAY II	91, 74	33, 46
BBN HWIM	44	195
System Development Corp.	24	105

Σχήμα 9. Τα τέσσερα συστήματα που προέκυψαν από το πρόγραμμα ARPA 1971 και οι τιμές της συνολικής απόδοσης τους.

Τα συστήματα αυτά είναι τα εξής:

1. Το σύστημα **HARPY** του παναπιστημίου Carnegie-Mellon (CMU) (Lowerre και Reddy, 1980). Η βάση για το σύστημα HARPY είναι ένα δίκτυο 15.000 καταστάσεων που περιλαμβάνει λεξιλογικές αναπαραστάσεις, συντακτικό, και κανόνες λέξεων συντεταγμένους σε ένα απλό πλαίσιο. Το προκύπτον FSA αποκωδικοποιείται σε σχέση με ακουστικές μετρήσεις χρησιμοποιώντας δυναμικό προγραμματισμό DP και αναζήτηση δέσμης. Το σύστημα που προηγήθηκε το HARPY είναι το σύστημα DRAGON, το οποίο υλοποιήθηκε στο πανεπιστήμιο Carnegie-Mellon και περιλάμβανε αποκωδικοποίηση ενός FSA με την χρήση αναζήτησης breadth-first DP. Ο όρος αυτός σημαίνει ότι όλα τα μονοπάτια εκτείνονται παράλληλα αντί να εκτείνονται πρώτα τα μονοπάτια μέγιστης πιθανοφάνιας. Η προσθήκη της έννοιας αναζήτησης δέσμης στο σύστημα HARPY βελτίωσε σε μεγάλο βαθμό την υπολογιστική αποτελεσματικότητα του συστήματος. Η ακουστική επεξεργασία στο σύστημα HARPY αποτελείται από 14 εξαγόμενες παραμέτρους γραμμικής πρόβλεψης με διάρκεια στο πλαίσιο λόγου 10-msec. Στην συνέχεια τα πλαίσια συνδιάζονται (με άθροιση των πινάκων συσχέτισης) αν είναι όμοια σε ικανοποιητικό βαθμό, με σκοπό να μειώσουν τον απαιτούμενο χρόνο επεξεργασίας και να εξομαλύνουν την επίδραση του θορύβου. Τα

προκύπτοντα “Ακουστικά Τμήματα” ταξινομούνται σε μία από τις 98 υπάρχουσες ομάδες με κριτήριο την απόσταση Itakura.

2. Το σύστημα **HEARSEY II** του πανεπιστημίου Carnegie-Mellon. Το σύστημα HEARSEY εχεί τελείως διαφορετική αρχιτεκτονική από το σύστημα HARPY. Οι πληροφορίες από όλες τις ακουστικές και γλωσσικές πηγές ενσωματώνονται σε ένα “black-board” ο οποίος λειτουργεί σαν ελεκτής κατά την επεξεργασία. Οι επεξεργαστές πληροφορίας είναι τελείως τμηματοποιημένοι και σχετικά εύκολα τροποποιούνται. Μία συνιστώσα “ο επαληθευτής λέξης” έχει την μορφή ενός FSA του τύπου HARPY για την σύνθεση λέξεων από μονάδες υπολέξεων. Η αλληλεπίδραση με το ακουστικό σήμα πραγματοποιείται με τη μέθοδο island-driven κάτα την οποία αναζτητείται η λέξη με την μεγαλύτερη πιθανότητα με σκοπό την ολοκλήρωση της υποθετικής πρότασης. Ο αλγόριθμος των Cocke-Younger-Kasami (CYK) χρησιμοποιείται στην διαδικασία αυτή με σκοπό να επεξεργάζονται παράλληλα πολλές υποθετικές λέξεις. Η ακουστική επεξεργασία στο σύστημα HERSAYII συνίσταται στον από κορυφή σε κορυφή υπολογισμό του πλάτους και των σημείων όπου γίνεται διέλευση από το μήδεν, σε μη επικαλυπτόμενα πλαίσια εύρους 10-msec, τόσο σε προενισχυμένες κυματομορφές, όσο και σε εξομαλυμένες κυματομορφές ομιλίας. Πλαίσια με όμοιες τιμές ομαδοποιούνται και ταξινομούνται με βάση την άρθρωση κάνοντας χρήση ενός δέντρου απόφασης και μία σειρά από κατώφλια ελέγχου.

3. Το σύστημα **HWIM** (“Hear What I Mean”) χρησιμοποιεί και αυτό την μέθοδο island-driven κατά την οποία οι τελικές υποθέσεις δημιουργούνται από μεγάλης πιθανότητας λέξεις. Η λεξικολογική αποκωδικοποίηση επιτυγχάνεται χρησιμοποιώντας ένα πολύπλοκο δίκτυο από φωνολογικούς κανόνες που εφαρμόζονται σε ένα πλέγμα από τις δυνατές κατατμήσεις της κυματομορφής. Οι υποθετικές λέξεις κατευθύνονται με βάση τα παραπάνω από συντακτικό και σημασιολογικό τρόπο, ο οποίος μορφοποιείται από μία ATN γραμματική. Η τελική υπόθεση πραγματοπιείται χρησιμοποιώντας τη μέθοδο καλύτερης αναζήτησης.

4. Το σύστημα **Systems Development Corporation (SDC)** κατανόησης ομιλίας παράγει ένα σύνολο από εναλλακτικά φωνητικά αντίγραφα που προέρχονται από ακουστική και φωνητική επεξεργασία, τα οποία στην συνέχεια αποθηκεύονται σε ένα πίνακα όπου και γίνεται η επεξεργασία. Το σύστημα στη συνέχεια ακολουθεί τις μεθόδους αναζήτησης left-to-right και best-first χρησιμοποιώντας ένα φωνητικό χαρτογράφο ο οποίος συνδέει τις σημασιολογικές και συντακτικές πηγές πληροφορίας με τα υποθετικά φωνητικά αντίγραφα κατά την ακουστική επεξεργασία.

TANGORA Ταυτόχρονα με τα προγράμματα που σχεδιάζονταν από την ARPA, την ίδια εποχή βρίσκονταν σε εξέλιξη ανάλογη εργασία στην IBM που αφορούσε την εφαρμογή στατιστικών μεθόδων στην αυτόματη αναγνώριση ομιλίας. Συγκεκριμένα, οι πρώτες εργασίες σχετικά με τα

HMM δημοσιεύτηκαν από τον Jelinek (1976) και ανεξάρτητα κατά την ίδια περίοδο από τον Baker (1975) στο πανεπιστήμιο Carnegie-Mellon. Το 1983 οι ερευνητές της IBM παρουσίασαν μία δημοσίευση σύμφωνα με την οποία επιτυγχάνεται σημαντικά καλύτερη απόδοση στην περίπτωση συνεχούς αναγνώρισης ομιλίας εξαρτημένου ομιλητή περιορισμένου θέματος, σε σχέση με το σύστημα HARPY. Ένα άλλο σημαντικό αποτέλεσμα που προέκυψε από την έρευνα των ανθρώπων της IBM ήταν το εξής: Στις αρχές του 1980 η ερευνητική ομάδα της IBM εστίασε την προσοχή της στο πρόβλημα της υπαγόρευσης κείμενου στο χώρο εργασίας. Το αποτέλεσμα, το οποίο ανακοινώθηκε το 1984, ήταν ένα εξαρτημένου ομιλητή, αναγνώρισης μενομωμένων λέξεων, σχεδόν σε πραγματικό χρόνο, σύστημα αναγνώρισης μεγάλου λεξιλογίου (5000 λέξεων), το οποίο ήταν βασισμένο πάνω σε μία σειρά υπολογιστικών ευκολιών που περιλάμβαναν μεταξύ των άλλων ένα κεντρικό υπολογιστή και ένα σταθμό εργασίας. Το 1987, το σύστημα λειτούργησε σε πραγματικό χρόνο σε προσωπικό υπολογιστή με τέσσερις πίνακες επεξεργασίας σήματος και με επέκταση του λεξιλογίου από 5000 λέξεις στις 20000 λέξεις. Το σύστημα πήρε το όνομα του από τον Albert Tangora. Μερικά αποτελέσματα που αφορούν την απόδοση του συστήματος δόθηκαν από τον Picone (1990) και παρουσιάζονται στον πίνακα 4.

Πίνακας 4.

Θέμα Αναγνώρισης	**Ρυθμός Εσφαλμένων Λέξεων (%)**
Αλληλογραφία γραφείου 5000 λέξεων	2.9
Αλληλογραφία γραφείου 20000 λέξεων	5.4
Οι 2000 πιο συχνά χρησιμοποιούμενες λέξεις στην αλληλογραφία γραφείου που χρησιμοποιούν φωνητικές βασικές φόρμες	2.5
Οι 2000 πιο συχνά χρησιμοποιούμενες λέξεις στην αλληλογραφία γραφείου που χρησιμοποιούν fenonic βασικές φόρμες	0.7

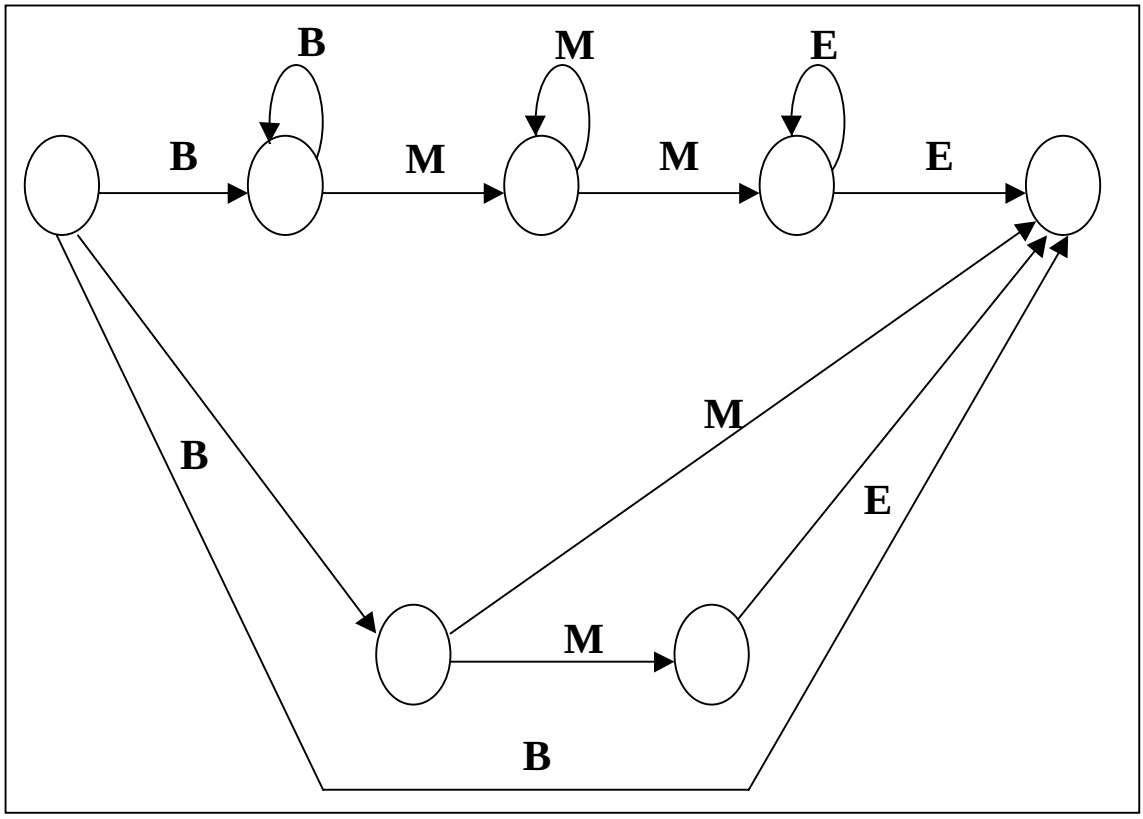
Σχήμα 10. Αποτελέσματα για την απόδοση του συστήματος TANGORA όπως παρουσιάστηκαν από τον Picone (1990).

BYBLOS. Στα επόμενα χρόνια η ARPA μετονομάσθηκε σε DARPA (Defense Advanced Projects Research Agency). Ένα από τα συστήματα που αναπτύχθηκαν στην περίοδο αυτή ήταν το σύστημα BYBLOS. Είναι ένα σύστημα συνεχούς αναγνώρισης ομιλίας εξαρτημένου ομιλητή και απευθύνεται σε εφαρμογές μεγάλου λεξιλογίου. Ένα από τα μοναδικά χαρακτηριστικά του συστήματος BYBLOS είναι ότι περιλαμβάνει μοντέλα φθόγγων ανεξέρτητα του περιεχομένου. Το γεγονός αυτό επιτρέπει την δυνατότητα να συμπεριληφθούν οι επιδράσεις της συνάρθρωσης. Το 1986, όταν το σύστημα παρουσιάστηκε υπήρχε μία ευρέως διαδεδομένη άποψη στην

κοινότητα επεξεργασίας ομιλίας ότι τα στοχαστικά συστήματα αναγνώρισης ομιλίας που βασίζονται σε φωνητικές μονάδες είναι μη πραγματοποιήσιμα.

Η ακουστική επεξεργασία στο σύστημα BYBLOS περιλαμβάνει τον υπολογισμό 14 mel-cepstral συντελεστών κάθε 10-msec χρησιμοποιώντας παράθυρο των 20-msec. Τα ακουστικά μοντέλα είναι HMM διακριτής παρατήρησης και βασίζονται σε διανυσματική κβάντιση (VQ), η οποία χρησιμοποιεί ένα κωδικό βιβλίο 256 συμβόλων. Το σύστημα BYBLOS έχει δοκιμαστεί κάτω από ποικίλες καταστάσεις και για διάφορα θέματα. Στο 1000 λέξεων DRMD χρησιμοποιήθηκε σαν υλικό δοκιμής. Τρία γραμματικά μοντέλα δημιουργήθηκαν για την κάλυψη των υποθετικών λέξεων. Το πρώτο είναι μία **κανονική γραμματική** (FSA) με περιπλοκή 9, το δεύτερο μία **γραμματική ζεύγους λέξης** περιπλοκής 60, και το τρίτο μία **κενή γραμματική** (απλά ίση με τον αριθμό των λέξεων).

Το σύστημα SPHINX. Το σύστημα SPHINX, το οποίο κατασκευάστηκε στο πανεπιστήμιο Carnegie-Mellon, αποτελεί ένα άλλο σύστημα που βασίζεται στην προσεκτική φωνητική μοντελοποίηση. Το σύστημα απευθύνεται σε αναγνώριση συνεχούς ομιλίας μεγάλου λεξιλογίου ανεξάρτητου ομιλητή.



Σχήμα 11. Φωνητικό μοντέλο που χρησιμοπιείται στο σύστημα SPHINX.

Το συστημα SPHINX βασίζεται σε ανεξάρτητα περιεχομένου, διακριτής παρατήρησης μοντέλα φθόγγων, τα οποία σε αυτή την προσέγγιση αναφέρονται ως τρίφθογγοι. Η βασική

HMM τοπολογία που χρησιμοποιείται δίνεται στο σχήμα 11. Χίλια τέτοιου είδους μοντέλα φθόγγων καταρτίστηκαν πάνω στους 1000 πιο συχνά εμφανιζόμενους φυσικούς τρίφθογγους στο DRMD, το οποίο χρησιμοποιήθηκε για να δοκιμαστεί το σύστημα. Τα HMM είναι διακριτές παρατηρήσεις, αλλά έχουν ένα ενδιαφέρον χαρακτηριστικό, στο ότι τρία κωδικά βιβλία των 256 συμβόλων χρησιμοποιούνται, cepstral, διαφορική cepstral, και ενεργειακά χαρακτηριστικά, τα οποία προέρχονται από LP ανάλυση και καθένα κωδικοποιείται ξεχωριστά. Κάθε μετάβαση στο FSA παράγει τρία μοναδικά χαρακτηριστικά και στην συνέχεια οι πιθανότητες τους συνδιάζονται.

Το σύστημα LINCOLN. Το σύστημα αυτό βασίστηκε στις προσπάθειες των ερευνητών στα εργαστήρια Lincoln να μοντελοποιήσουν την ομιλία κάτω από διαφορετικές καταστάσεις ταχύτητας, έντασης και συναισθημάτων. Το 1989 ο Paul επέκτεινε αυτή την έρευνα στον τομέα της αναγνώρισης συνεχούς ομιλίας σε μεγάλο λεξιλόγιο τόσο για την περίπτωση εξαρτημένου ομιλητή όσο και για την περίπτωση ανεξάρτητου ομιλητή. Το DRMD χρησιμοποιήθηκε με μία ***γραμματική ζεύγους λέξης***. Συνεχούς παρατήρησης, Gaussian's πυκνότητας HMM χρησιμοποιούνται για να μοντελοποιήσουν φωνήματα ευαίσθητα στο περιεχόμενο.

DECIPHER Το σύστημα Decipher, το οποίο κατασκευάστηκε από την SRI International, χαρακτηρίζεται από ιδιότητες παρόμοιες με αυτές των συστημάτων που παρουσιάστηκαν έως τώρα. Ένα αξιοσημείωτο χαρακτηριστικό αυτού του συστήματος είναι η μεγάλη προσοχή που δίνεται στην μοντελοποίηση των φωνολογικών λεπτομεριών όπως οι διαλεκτικές επιδράσεις συνάρθρωσης και η ιδιάζωντος ομιλητή φωνολογική προσαρμογή. Βασισμένο σε πειράματα της DRMD, συγκρίσεις με τα συστήματα SPHINX και BYBLOS δίνουν σαν αποτέλεσμα βελτίωση της απόδοσης σαν συνέπεια της φωνολογικής μοντελοποίησης.

CSELT Σαν μέρος της συνδιασμένης Ευρωπαϊκής προσπάθειας ESPRIT, οι ερευνητές στο Centro Studi e Laboratori Telecomunicazioni (CSELT) και στο πανεπιστήμιο του Σαλέρμο κατασκεύασαν ένα σύστημα συνεχούς αναγνώρισης ομιλίας 1000 λέξεων το οποίο περιέχει ένα μοναδικό σύστημα διατύπωσης υποθέσεων. Χρησιμοποιώντας ένα είδος της N-best προσέγγισης το σύστημα CSELT επιλέγει λέξεις με βάση μία τραχιά φωνητική περιγραφή και στην συνέχεια βελτιώνει την υπόθεση χρησιμοποιώντας λεπτομερές ταίριασμα. Σε μία γλώσσα με περιπλοκή 25, πειράματα στα οποία συμμετέχουν δύο ομιλητές οι οποίοι εκφωνούν 214 προτάσεις παράγεται μία ακρίβεια στις λέξεις της τάξης του 94.5% ενώ ο ρυθμός σωστών προτάσεων είναι 89.3%.

Connected-Digit Recognition with Language Models Σε πολλές περιπτώσεις οι ερευνητές των AT&T εργαστηρίων προσπάθησαν να επιλύσουν το πρόβλημα της αναγνώρισης συνεχών ακολουθιών ψηφίων από ανεξάρτητους ομιλητές. Οι προσπάθειες αυτές συνέβαλαν στην κατανόηση των ιδιοτήτων των HMM. Συγκεκριμένα, πολλά συμπεράσματα στην εφαρμογή των

HMM συνεχούς παρατήρησης αναπτύχθηκαν κατά την έρευνα αυτή. Επίσης αξιοσημειώτη είναι η εφαρμογή ενός LB αλγόριθμου που βασίζεται σε HMM συνεχούς παρατήρησης με μία γραμματική πεπερασμένων καταστάσεων εφαρμοσμένο σε ένα συστολικό επεξεργαστή που κατασκευάστηκε από την AT&T.

Very Large Vocabulary Systems Πολλές ερευνητικές ομάδες έχουν εργαστεί στο παρελθόν πάνω στο πρόβλημα της αναγνώρισης πολύ μεγάλων λεξιλογίων με την βοήθεια των διαφόρων μοντέλων γλωσσών. Στο INRS και Bell Northern του Καναδά οι ερευνητές εργάστηκαν πάνω σε σύστημα εξαρτημένου ομιλητή 75000 λέξεων με πολλά διαφορετικά μοντέλα γλώσσας. Η καλύτερη απόδοση επιτεύχθηκε με ένα τριγραμματικό μοντέλο, με το οποίο επιτεύχθηκε αναγνώριση της τάξης του 90%. Στην IBM, στο Παρίσι, τα πειράματα διεξήχθησαν με ένα λεξιλόγιο της τάξης των 200000 λέξεων στο οποίο ο τρόπος που γίνεται η είσοδος των δεδομένων είναι αυτός του εξαρτημένου ομιλητή με εκφωνήσεις που γίνονται συλλαβή προς συλλαβή.

ΒΙΒΛΙΟΓΡΑΦΙΑ

MARKOWITH A. JUDITH *Using Speech Recognition*. Upper Saddle River, New Jersey: Prentice Hall.

JOHN R. DELLER, JR., JOHN G. PROAKIS, JOHN H. L. HANSEN *Discrete-Time Processing of Speech Signals*. Upper Saddle River, New Jersey: Prentice Hall.

ΑΓΓΛΙΚΗ ΟΡΟΛΟΓΙΑ

A

Acoustic Decoder	Ακουστικός Απωκοδικοποιητής
Acoustic-phonetic	Ακουστοφωνητικός
Ambiguous word	Ασαφής Λέξη
Application Scanning	Διερεύνηση Εφαρμογής
Augmented Transition Network Grammars	Γραμματικές Επαυξημένου Δικτύου Μεταβάσεων

B

Baseform	Βασική Φόρμα
Beam Search Algorithm	Αλγόριθμος Αναζήτησης Δέσμης
Best Match	Καλύτερο Ταίριασμα
Best Path	Καλύτερο Μονοπάτι
Branching Factor	Συνδετικός Παράγοντας
Built-in Vocabulary	Ενσωματωμένο Λεξιλόγιο

C

Cepstral	Το Αποτέλεσμα του Μετασχηματισμού Fourier του Λογαριθμικού Φάσματος
Chart	Διάγραμμα
Coarticulation	Συνάρθρωση
Confusable Word	Συγχεόμενη Λέξη
Context-independent	Ανεξάρτητα του Περιεχομένου
Context-sensitive phone-like unit	Ευαίσθητη από το Περιεχόμενο Μονάδα που Μοιάζει με Φώνημα
Corpus	Σώμα Υλικού ή Κειμένου

Corpora	Σώμα Υλικού ή Κειμένου

D

Distance Metric	Μετρική Απόσταση
Dynamic Programming	Δυναμικός Προγραμματισμός
Dynamic Time Warping	Δυναμική Στρεύλωση Χρόνου

E

End-user	Τελικός Χρήστης
Ergotic	Εργοδικό

F

Fast matching	Γρήγορο Ταίριασμα
Finite State Automaton	Αυτόματο Πεπερασμένων Καταστάσεων

G

Grammar	Γραμματική
Graphemic	Γραφημικός
Generalized Triphone	Γενικευμένος Τρίφθογγος

H

Hidden Markov Model	Κρυφό Μοντέλο Markov

I

Inter-speaker	Μεταξύ των Ομιλητών
Intra-speaker	Του Ίδιου του Ομιλητή

K

Keystroke	Πληκτρολόγηση

L

Lexical Semantics	Λεξιλογική Σημασιολογία
Linguistic Constrains	Γλωσσικοί Περιορισμοί
Linguistic Decoder	Γλωσσικός Αποκωδικοποιητής

M

Maximum-Likelihood Algorithm	Αλγόριθμος Μέγιστης Πιθανοφάνιας

P

Parsing	Συντακτική Ανάλυση της Γλώσσας
Perplexity	Περιπλοκή
Personalizing the Application	Εξατομίκευση της Εφαρμογής
Phone	Φθόγγος
Phone-like Unit	Μονάδες που Μοιάζουν με Φωνήματα
Phoneme	Φώνημα
Phoneme in Context	Φώνημα στο Περιεχόμενο
Production Rules	Κανόνες Παραγωγής

R

Recursive Transition	Αναδρομική Μετάβαση
Reference Database	Βάση Δεδομένων Αναφοράς
Reference Model	Μοντέλο Αναφοράς
Resident Vocabulary	Εσωτερικό Λεξιλόγιο
Robustness	Ευρωστεία

S

Seed Model	Μοντέλο Προέλευσης
Sparse Data Problem	Πρόβλημα Αραιών Δεδομένων

Speaker Adaptation	Προσαρμογή Ομιλητή
Speaker-dependant	Εξαρτήμενος Ομιλητής
Speaker-independent	Ανεξάρτητος Ομιλητής
Spectrogram	Φασματογράφημμα
Speech Front-end	Μετωπική Ομιλία
Speech Recognition	Αναγνώριση Ομιλίας
Stack Decoding Algorithm	Αλγόριθμος Αποκωδικοποίησης Στοίβας
Stochastic	Στοχαστικός
Subword	Υπολέξη
Subword Modeling	Μοντελοποίηση Υπολέξεων
Symbol	Σύμβολο

T

Template	Ίχνος
Temporal Alignment	Χρονική Ευθυγράμμιση
Template Matching	Ταίριασμα Ιχνών
Terminal	Τερματικός
Token	Εκφώνηση
Total Vocabulary	Ολικό Λεξιλόγιο
Transition	Μετάβαση
Translation	Μετάφραση
Tree Network	Δενδροειδές Δίκτυο
Triphone	Τρίφθογγος
Turnkey Application	Ετοιμοπαράδοτη Εφαρμογή

U

Unification Grammars	Ενοποιημένες Γραμματικές

V

Variability	Μεταβλητότητα
Variation	Διακύμανση
Vector Quantization	Διανυσματική Κβάντιση
Vendor-built Lexicon	Λεξικά Φτιαγμένα από Κατασκευαστές
Vocal Tract	Φωνητική Οδός
Vocabulary Optimization	Βελτιστοποίηση Λεξιλογίου

W

Word-pair Grammar	Γραμματική Ζεύγους Λέξης
Word Spotting	Στόχευση Λέξης

ΠΙΝΑΚΑΣ ΣΥΝΤΜΗΣΕΩΝ

AD	Acoustic Decoder
ARPA	Advanced Research Projects Agency
ATN	Augmented Transition Network
BBN	Bolt Beranek and Newman
CMU	Carnegie-Mellon University
CS-PLU	Context-sensitive Phone-like Unit
DARPA	Defense Advanced Research Projects Agency
DRMD	DARPA Resources Management Database
FSA	Finite State Automata
HMM	Hidden Markov Model
LB	Level Building Algorithm
LD	Linguistic Decoder
MIT	Massachusetts Institute of Technology
NBS	National Bureau of Standards
PC	Personal Computer
PIC	Phoneme in Context
PLU	Phone-like Unit
SDC	Systems Development Corporation
SRI	Stanford Research Institute
SSI	Speech Systems Inc.
TI	Texas Instruments Corporation
VCS	Voice Control System
VQ	Vector Quantization