ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
Μεταπτυχιακό Πρόγραμμα Σπουδών

Ιωάννης Μπόγδος (Α.Μ. 97518)
Ιωάννα Παπαιωάννου (Α.Μ. 97524)
Ελένη Παπαιωάννου (Α.Μ. 97525)

Ο Ρόλος της Ομιλίας στην Επικοινωνία Ανθρώπου Μηχανής

Εργασία στο μάθημα: Επικοινωνία με Ομιλία
Διδάσκων: Γεώργιος Κουρουπέτρογλου

Αθήνα 1999

Περίληψη της εργασίας

Στην εργασία αυτή αρχικά μελετήθηκε η αναγκαιότητα της χρήσης μεθόδων οργάνωσης και δόμησης λεξιλογίων για εφαρμογές αναγνώρισης ομιλίας. Επίσης μελετήθηκαν οι κυριότερες μεθόδοι ως προς τα πλεονεκτήματα, τα μειονεκτήματα, την υλοποίησή τους και τις εφαρμογές για τις οποίες είναι κατάλληλες.

Με τη χρήση τεχνικών δόμησης λεξιλογίων επιτυγχάνεται *μείωση της περιπλοκής, αύξηση της ταχύτητας και της ακρίβειας της εφαρμογής* καθώς και *αύξηση της ευελιξίας του λεξιλογίου.*

Η πιο ευρέως χρησιμοποιούμενη τεχνική δόμησης λεξιλογίων είναι οι *γραμματικές πεπερασμένων καταστάσεων.* Με τη μέθοδο αυτή ορίζονται οι λέξεις που επιτρέπονται σε οποιοδήποτε σημείο εισόδου και επομένως επιτυγχάνονται οι παραπάνω στόχοι. Η μέθοδος όμως αυτή δεν είναι κατάλληλη για εγαρμογές που απαιτούν μεγάλη ευελιξία, αφού δεν μπορούν να αναγνωρίσουν ακολουθίες λέξεων και προτάσεων που δεν έχουν οριστεί και δεν μπορούν να κατανοήσουν συντακτικά πρότυπα. Υλοποιείται με ένα δίκτυο πεπερασμένων καταστάσεων.

Μια παραλλαγή της είναι η *γραμματική ζευγαρώματος λέξης.*

Μια άλλη μέθοδος είναι *οι γραμματικές που βασίζονται στη γλωσσολογία,* με την οποία όχι απλά αναγνωρίζεται αλλά και κατανοείται η ομιλούμενη γλώσσα. Η κυριότερη από αυτές είναι η *γραμματική ελεύθερου περιεχομένου*, η οποία είναι πολύ ευέλικτη, γιατί μπορεί να αναπαραστήσει μια ποικιλία από γλωσσολογικές προσεγγίσεις και υλοποιείται με το διάγραμμα σύνταξης.

Σε στοχαστικές προσεγγίσεις χρησιμοποιούνται πιθανοτικά μοντέλα τα κυριότερα από τα οποία είναι τα μοντέλα *Ν-γραμμάτων* και *Ν-κλάσεων.*

Στα μοντέλα *Ν-γραμμάτων* η τρέχουσα (άγνωστη) λέξη καθορίζεται, θεωρώντας οτι η ταυτότητά της εξαρτάται από τις Ν-1 προηγούμενες λέξεις συν την ακουστική πληροφορία στην άγνωστη λέξη. Η μέθοδος αυτή μειώνει το ενεργό λεξιλόγιο και είναι απλή, ευέλικτη, ταχύτατη και ακριβής. Παρ'όλα αυτά αδυνατεί να κατανοήσει τη γλώσσα. Υλοποιείται με τις τεχνικές της *εκπαίδευσης του συστήματος*, της *οπισθοχώρησης* και των *συντοποθετήσεων.*

Στα μοντέλα *Ν-κλάσεων* οι λέξεις των δεδομένων κατηγοριοποιούνται, οπότε υπολογίζεται η πιθανότητα Ν κατηγορίες να εμφανίζονται σε διαδοχή. Με τη μέθοδο αυτή παράγονται ακριβή και συμπαγή μοντέλα γλωσσών.

Μεγάλο ενδιαφέρον παρουσιάζουν και οι *γραμματικές με πολλαπλές πηγές γνώσεων*, οι οποίες συνδυάζουν συντακτική ανάλυση με σημασιολογία, στατιστικά και άλλες πηγές γνώσεων. Τα χαρακτηριστικά των πιο γνωστών γραμματικών αυτού του είδους, όπως το *Spoken Language System*, το *TINA*, το *JANUS,* και το *SUS* αναφέρονται στην εργασία.

Ο Εντοπισμός λέξεων (Word Spotting) σε συνεχή ομιλία είναι μια άλλη τεχνική που χρησιμοποιείται για την αναγνώριση ομιλίας με πρώτη εφαρμογή το σύστημα VRCP της AT&T , για το χειρισμό τηλεφωνικών κλήσεων μακρινών αποστάσεων στα τέλη της δεκαετίας του 80. Μια παραλλαγή του Word Spotting είναι το Gisting το οποίο επεκτείνει τον εντοπισμό λέξεων σε δομημένους διαλόγους ανάμεσα σε 2 άτομα με το σύστημα αναγνώρισης να ενεργεί σαν ένας μη αντιληπτός παρατηρητής.

Στη συνέχεια της εργασίας αναλύονται οι 4 βασικοί παράγοντες από τους οποίους εξαρτάται η σχεδίαση μιας επιτυχημένης γραμματικής , δηλ. η ανάλυση της εργασίας, ο χρόνος απόκρισης του συστήματος, οι εφαρμογές που αντιλαμβάνουν την ομιλία καθώς

επίσης και η ετοιμότητα του συστήματος να ανταπεξέλθει στη απρόσμενη συμπεριφορά των χρηστών.

Στο 2^{o} κεφάλαιο παρουσιάζονται οι μέθοδοι μοντελοποίησης του ομιλητή. Αυτοί είναι μοντελοποίηση εξαρτώμενη του ομιλητή, ανεξάρτητη, πολλαπλών ομιλητών και προσαρμογής του ομιλητή. Οι τρόποι υλοποίησής τους καθώς και τα πλεονεκτήματα, τα μειονεκτήματά τους και η εφαρμογές τους αναλύονται εκτενώς.

Στο παράρτημα που παρατίθεται στο τέλος της εργασίας δίδονται κάποιες εφαρμογές από τον χώρο της κινητής τηλεφωνίας.

1 ΠΡΟΣΘΕΤΙΚΕΣ ΔΟΜΕΣ

ΕΙΣΑΓΩΓΗ

Οι εφαρμογές αναγνώρισης ομιλίας δεν αποτελούνται από μία μη οργανωμένη συσσώρευση λέξεων. Εχουν την ίδια δομή και ακολουθία όπως και οι σκοποί με τους αποίους συνδέονται. Αυτή η δομή μπορεί να είναι τόσο συμπαγής όσο μία ακολουθία εντολών για τον έλεγχο μίας συσκευής ή τόσο ευέλικτο όσο η υπαγόρευση.

Οι μέθοδοι που χρησιμοποιούνται για την οργάνωση λεξιλογίων εφαρμογών έχουν μια ποικιλία ονομάτων. Πιο συχνά, καλούνται *γραμματικές, μοντέλα ή γραφές*. Ο όρος *γραμματική* είναι ο πιο κατάλληλος αφού αναφέρεται στη σχέση των λέξεων σε μία πρόταση. Δυστυχώς, απαντάται σε ένα φάσμα βασικών σχολικών ασκήσεων και τον αποφεύγουν μερικοί. Η λέξη *μοντέλο* εφαρμόζεται σε στοχαστικές/στατιστικές προσεγγίσεις, εκφράζοντας τη βασική διαδικασία αυτών των γραμματικών: στη δημιουργία μοντέλων για πρότυπα γλώσσας. Ο όρος *γραφή* συνήθως χρησιμοποιείται σε εφαρμογές που στηρίζονται στην τηλεφωνία. Εστιάζει στη δόμηση της αλληλεπίδρασης ανάμεσα σε αυτόν που κάνει την κλήση και στο σύστημα αναγνώρισης.

Οι πιό ευρέως χρησιμοποιούμενες μορφές δόμησης είναι:

- Γραμματικές πεπερασμένων καταστάσεων
- Πιθανοτικά μοντέλα
- Γραμματικές βασισμένες στη γλωσσολογία

Αυτές οι προσεγγίσεις διαφέρουν ως πρός

- Στόχους
- Υλοποίηση
- Ευελιξία
- Διαχείρηση μεγάλων λεξιλογίων
- Διαχείρηση σύνθετων γλωσσικών προτύπων

Αυτές οι διαφορές επηρεάζουν την καταλληλοτητά τους για τη χρησιμοποιησή τους σε συγκεκριμένους τύπους εφαρμογών.

Η Τεχνολογική Εστίαση ξεκινά με μια περισσότερο εκτεταμένη ερμηνεία του ρόλου των τεχνικών δόμησης. Αυτή ακολουθείται από την εξέταση των βασικών μορφών της γραμματικής που βρίσκονται σε εμπορικά και ερευνητικά συστήματα. Κάθε μορφή γραμματικής χαρακτηρίζεται ως προς την ευελιξία της, τους περιορισμούς της και την κανονική υλοποίησή της στα συστήματα αναγνώρισης ομιλίας. Η Εστίαση στις Εφαρμογές εξετάζει θέματα σχεδίασης εφαρμογών που συνδέονται με τις μορφές της γραμματικής που χρησιμοποιούνται σε εμπορικά συστήματα αναγνώρισης. Αυτά τα θέματα περιέχουν φυσικότητα και ακρίβεια.

Επιπλέον, το βιβλίο του Terry Winograd (1983) που δίνεται στις αναφορές του κεφαλαίου είναι μια πολύ καλή πηγή πληροφορίας για τις γραμματικές που χρησιμοποιούνται ευρέως σε συστήματα υπολογιστών.

ΠΡΩΙΜΕΣ ΓΡΑΜΜΑΤΙΚΕΣ

Όπως πολλά στοιχεία της υπάρχουσας τεχνολογίας αναγνώρισης ομιλίας, η χρησιμοποίηση των τεχνικών δόμησης σε εφαρμογές αναγνώρισης ομιλίας είναι ένα υπο - προιόν του προγράμματος ARPA SUR του 1970 (βλ. κεφάλαιο 1, παράγραφο 1.2). Τα ARPA SUR συστήματα ήταν απαραίτητα για την διαχείρηση λεξιλογίων με χίλιες ή περισσότερες λέξεις, αναγκάζοντας τους σχεδιαστές εφαρμογών να εστιάσουν σε τρόπους ελλάτωσης του ψαξίματος που έπρεπε να κάνει το σύστημα για να βρεί τον καλύτερο συνδυασμό. Η αξία της χρήσης γραμματικής πεπερασμένων καταστάσεων για δομημένες εφαρμογές αναγνώρισης ομιλίας παρουσιάστηκε από το Σύστημα Harpy του Πανεπιστημίου Carnegie Mellon, το οποίο ξεπέρασε τα άλλα συστήματα του προγράμματος.

Το σύστημα Hearsay - II του CMU παρουσίασε το *μοντέλο του πίνακα* στην αναγνώριση ομιλίας. Ένας πίνακας χρησιμοποιεί πολλαπλές πηγές γνώσης για την αναγνώριση και την κατανόηση της ομιλουμένης γλώσσας. Οι πηγές γνώσης στο Hearsay - II κυμαίνονται από φωνητικές παραμέτρους, όπως η φώνηση, έως δομές επιπέδου φράσης. Αυτές οι πηγές γνώσης αντανακλούν το εύρος της πληροφορίας που εφαρμόζεται για την κατανόηση της ομιλίας. Κατά τη διάρκεια της διαδικασίας ανάλυσης, οι πηγές γνώσης γεννούν υποθέσεις για τμήματα της εισόδου. Αφού οι πηγές γνώσης αποτελούν πολύ διαφορετικές μορφές πληροφορίας, γειτονικές πηγές γνώσης επικοινωνούν μεταξύ τους για τις υποθέσεις τους χρησιμοποιώντας ένα κέντρο μηνυμάτων, που καλείται *πίνακας*.

Το σύστημα *Hear What I Mean* (*HWIM*) των Bolt Beranek και Newman ήταν το πρώτο σύστημα αναγνώρισης ομιλίας που εφάρμοσε *αυξημένη γραμματική μεταβάσεων* (ATN) που περιλαμβάνει συντακτικά και σημασιολογικά στοιχεία.

Και οι τρείς αυτές προσεγγίσεις συνεχίζουν να χρησιμοποιούνται στην αναγνώριση ομιλίας, αν και μόνο οι γραμματικές πεπερασμένων καταστάσεων έχουν τύχει εμπορικής επιτυχίας. Αλλες πρώιμες προσεγγίσεις στη δόμηση, κυρίως γραμματικές βασισμένες στη γλωσσολογία, γνωρίζουν μία αναβίωση, σαν αποτέλεσμα της ώθησης για ανάπτυξη *συστημάτων κατανόησης ομιλουμένης γλώσσας* (SLU's). Τα συστήματα SLU δεν αναγνωρίζουν την είσοδο σαν ακολουθίες ακουστικών προτύπων. Προσπαθούν να καταλάβουν περισσότερα για την είσοδο, περιλαμβάνοντας το νοημά της και την σύνταξή της.

Για λεπτομερέστερες πληροφορίες για το πρόγραμμα ARPA SUR, βλ. Baker (1975b), Erman, et al. (1980) και Klatt (1977).

1.1 ΓΙΑΤΙ ΝΑ ΧΡΗΣΙΜΟΠΟΙΟΥΜΕ ΤΕΧΝΙΚΕΣ ΔΟΜΗΣΗΣ ;

Η χρήση δομής σε εμπορικά συστήματα αναγνώρισης ομιλίας κατευθύνεται προς τον καθορισμό επιτρεπτών και/ή πιθανών ακολουθιών από λέξεις. Αυτό μερικές φορές καλείται *σύνταξη*. Γραμματικές βασισμένες στη γλωσσολογία (που περιγράφονται στην παράγραφο 1.4) επεκτείνουν την κάλυψη της γραμματικής για να περιλάβει νόημα, κοινωνικούς κανόνες αλληλεπίδρασης, δομή λέξης και άλλες πλευρές της προφορικής επικοινωνίας.

Οι πρωταρχικοί στόχοι δόμησης κάθε είδους είναι να:

- Μειώσουν την περιπλοκή
- Αυξήσουν την ταχύτητα και την ακρίβεια
- Αυξήσουν την ευελιξία του λεξιλογίου

1.1.1 Μείωση της περιπλοκής

Το βήμα αναγνώρισης της διαδικασίας αναγνώρισης ομιλίας (βλ. κεφάλαιο 2, παράγραφος 2.3) περιλαμβάνει έρευνα μέσω του λεξιλογίου της εφαρμογής για να εντοπίσει την καλύτερη αντιστοιχία για την ομιλία εισόδου. Το σύνολο των στοιχείων το λεξιλογίου που πρέπει να εκτιμήσει το σύστημα αναγνώρισης για να ολοκληρώσει αυτόν τον στόχο καλείται *ενεργό λεξιλόγιο*. Ο αριθμός των στοιχείων στο ενεργό λεξιλόγιο σε ένα σημείο της διαδικασίας αναγνώρισης καλείται *παράγοντας διακλάδωσης* αυτού του σημείου στην εφαρμογή. Ο *μέσος παράγοντας διακλάδωσης* αναφέρεται στο ποσό της διακλάδωσης που υπάρχει στην όλη εφαρμογή. Ο όρος *περιπλοκή* συχνά χρησιμοποιείται σαν συνώνυμο για τον μέσο παράγοντα διακλάδωσης, αλλά έχει ένα ελαφρώς πιό περίπλοκο και ακριβή ορισμό (βλ. Jelinek 1968 και 1985 για λεπτομερέστερες συζητήσεις για τον ακριβή τεχνικό ορισμό του όρου *περιπλοκή*).

Οι εφαρμογές αναγνώρισης μπορούν να χαρακτηριστούν χρησιμοποιώντας την περιπλοκή. Μία εφαρμογή που αποτελείται αποκλειστικά από αποκρίσεις “ναι” και “οχι”, για παράδειγμα, έχει περιπλοκή δύο, αν και η υπολογισμένη περιπλοκή της αντιστοιχίας γενικών εργασιών έχει εκτιμηθεί σε κλίμακες από μία εκατοντάδα εως και μερικές εκατοντάδες.

Η μείωση της περιπλοκής αυτόματα μειώνει το ποσό της έρευνας που πρέπει να κάνει το σύστημα αναγνώρισης. Η μείωση της περιπλοκής αυξάνει επίσης τη ταχύτητα της εφαρμογής κατά τη διάρκεια της χρήσης και μπορεί επίσης να εφαρμοστεί για να ανεβάσει την ακρίβειά της. Οι γραμματικές μειώνουν την περιπλοκή περιορίζοντας τις λέξεις στο ενεργό λεξιλόγιο ή αναθέτοντας μία ιεραρχία πιθανοτήτων σε αυτά τα στοιχεία. Μία εφαρμογή με λεξιλόγιο συνόλου χιλίων λέξεων μπορεί, για παράδειγμα, να έχει ως είσοδο

Πάρε το δρόμο διοδίων για Milwaukee

Εάν αυτή η εφαρμογή δεν έχει γραμματική, ο παράγοντας διακλάδωσης σε κάθε σημείο της ακολουθίας είναι χίλια. Αυτό σημαίνει ότι το σύστημα αναγνώρισης πρέπει να εξετάσει κάθε μία από τις χίλιες λέξεις του λεξιλογίου έξι φορές για να βρεί τις σωστές ακολουθίες των στοιχείων εισόδου. Χωρίς γραμματική, η περιπλοκή αυτής της εφαρμογής είναι χίλια.

Εάν η εφαρμογή περιελάμβανε μία γραμματική με την επιτρεπόμενη πρόταση εισόδου:

Πάρε το ΤΥΠΟΣ δρόμου για ΜΕΡΟΣ

στην οποία ΤΥΠΟΣ και ΜΕΡΟΣ είναι μεταβλητά κενά που μπορούν να γεμίσουν με ένα από τα παρακάτω:

ΤΥΠΟΣ: εθνικός ή διοδίων ή μακρύς και με στροφές ή προς τα πίσω ή βραχώδης
ΤΟΠΟΣ: Milwaukee ή Kokomo ή πουθενά ή Αραβία ή Ρίο

η περιπλοκή της εφαρμογής είναι πολύ μικρότερη από χίλια. Ο παράγοντας διακλάδωσης για την πρώτη λέξη είναι ένα (το ενεργό της λεξιλόγιο είναι "πάρε"). Ο παράγοντας διακλάδωσης για τη δεύτερη, τέταρτη και πέμπτη λέξη είναι επίσης ένα. Ο παράγοντας διακλάδωσης για το τρίτο στοιχείο εισόδου είναι πέντε και το ενεργό του λεξιλόγιο αποτελείται από τα στοιχεία που μπορούν να γεμίσουν το μεταβλητό κενό ΤΥΠΟΣ (εθνικός ή διοδίων ή μακρύς και με στροφές ή προς τα πίσω ή βραχώδης). Η έκτη λέξη έχει επίσης παράγοντα διακλάδωσης πέντε. Σε κανένα σημείο στην έκφραση δεν θα απαιτείτο από το σύστημα να ψάξει όλο το λεξιλόγιο.

Τα προιόντα αναγνώρισης μπορούν να περιγραφούν με όρους της περιπλοκής που μπορούν να χειριστούν, κάνοντας την περιπλοκή ένα σημαντικό στοιχείο για την αξιολόγηση του συστήματος.

> Θεωρούμε οτι η μειώση της περιπλοκής ενός μοντέλου γλώσσας σε ένα κείμενο δοκιμής θα έχει ως αποτέλεσμα μικρότερο ρυθμό λαθών αναγνώρισης σε αυτό το κείμενο και επομένως εμείς χρησιμοποιούμε την περιπλοκή ως ένα μέσο της ποιότητας ενός μοντέλου γλώσσας (Frederick Jelinek, Robert Mercer και Salim Roukos, έμπειροι ερευνητές, IBM Watson Researcher Center, [ο Jelinek είναι τώρα στο Πανεπιστήμιο του Pittsbyrg]."Classifying words for improved statistical language models," 1990. p.621).

Περισσότερες πληροφορίες για την περιπλοκή υπάρχουν στον Jelinek (1985).

1.1.2 Αυξηση της ταχύτητας και της ακρίβειας

Όταν για ένα σύστημα αναγνώρισης υπάρχουν λιγότερα στοιχεία να υπολογίσει, ο χρόνος που απαιτείται για την αναγνώριση αυτόματα μειώνεται. Έαν μία εφαρμογή με λεξιλόγιο συνόλου χιλίων λέξεων χρειάζεται να ερευνήσει πέντε από αυτές τις λέξεις σε ένα μόνο σημείο, ο χρόνος που απαιτείται για αναγνώριση σε αυτό το σημείο γίνεται περίπου το 0.005 του χρόνου που θα απαιτείτο για ερευνήσει όλο το λεξιλόγιο. Η σημασία της μείωσης της περιττής έρευνας αυξάνει με το μέγεθος του λεξιλογίου της εφαρμογής. Η ακρίβεια αυξάνει γιατί υπάρχουν λιγότερες υποψήφιες λέξεις στο ενεργό λεξιλόγιο για να προκληθεί ένα λάθος.

1.1.3 Αύξηση της ευλιξίας του λεξιλογίου

Ο προσεκτικός σχεδιασμός των ενεργών λεξιλογίων αυξάνει την ακρίβεια όταν οι λέξεις μέσα σε κάθε ενεργό λεξιλόγιο διακρίνονται ακουστικά. Λέξεις που μπερδεύονται, όπως η "mine" και η "nine" και ομόφωνες, όπως η "to", "too" και η "two" μπορούν ακόμα να συμπεριληφθούν σε μία και μόνη εφαρμογή, αλλά τα λάθη αναγνώρισης που παράγουν ελαχιστοποιούνται τοποθετώντας αυτές σε διαφορετικά σύνολα ενεργών λεξιλογίων. Μία γραμματική μπορεί, για παράδειγμα, να καθορίσει οτι μόνο αριθμοί όπως το "two", θα αναγνωρίζονται σε ένα συγκεκριμένο σημείο, και επομένως να ελλατώσει πιθανά λάθη από τις ομόφωνες, όπως η "to" και η "too". Αφού φυσικά σύνολα λέξεων, μπορεί να περιέχουν λέξεις που μπερδεύονται, όπως γράμματα του Αγγλικού αλφαβήτου, δεν είναι δυνατό να ελλατωθούν οι λέξεις που μπερδεύονται από ένα και μόνο ενεργό λεξιλόγιο.

ΓΡΑΜΜΑΤΙΚΗ ΠΕΠΕΡΑΣΜΕΝΩΝ ΚΑΤΑΣΤΑΣΕΩΝ

Η γραμματική πεπερασμένων καταστάσεων κέρδισε έδαφος σαν ένα αποτελεσματικό εργαλείο για τη δόμηση εφαρμογών αναγνώρισης ομιλίας κατά τη διάρκεια του προγράμματος ARPA SUR της δεκαετίας του 70. Μια γραμματική πεπερασμένων καταστάσεων εχρησιμοποιείτο από το μόνο σύστημα που ικανοποιούσε τις απαιτήσεις ακρίβειας του ARPA SUR, Carnegie Mellon του πανεπιστημίου το *Harpy*. Στη δεκαετία του 80, η γραμματική πεπερασμένων καταστάσεων έγινε η κυρίαρχη τεχνολογία που εχρησιμοποιείτο σε εμπορικά συστήματα αναγνώρισης ομιλίας. Παραμένει η κύρια μορφή γραμματικής που συναντάται σε εμπορικά συστήματα με μικρό και μεσαίο μέγεθος λεξιλογίων. Χρησιμοποιείται επίσης και σε μεγαλύτερα συστήματα λεξιλογίων, περιλαμβάνοντας προιόντα Speech Systems Inc,'s (SSI) και σύστημα BBN's HARK.

Ο Winograd (1983, Κεφ. 2) περιλαμβάνει μία πολύ καλή τεχνική ανάλυση της γραμματικής πεπερασμένων καταστάσεων.

1.2 ΓΡΑΜΜΑΤΙΚΗ ΠΕΠΕΡΑΣΜΕΝΩΝ ΚΑΤΑΣΤΑΣΕΩΝ

> Η πιο ευρέως χρησιμοποιούμενη γραμματική σε αυτόματη αναγνώριση ομιλίας είναι η γραμματική πεπερασμένων καταστάσεων. Οι γραμματικές πεπερασμένων καταστάσεων είναι επιτυχημένες στη βελτίωση της ακρίβειας της αναγνώρισης σε περιορισμένης εμβέλειας συστημάτων αναγνώρισης. (L. Miller & S. Levinson, έμπειροι ερευνητές, AT&T Bell Laaboratories, "Syntactic analysis for large vocabulary speech recognition using a context - free covering grammar,"1988, p.270).

Μια *γραμματική παπερασμένων καταστάσεων* μειώνει την περιπλοκή σχεδιαγράφοντας τις λέξεις που επιτρέπονται σε οποιοδήποτε σημείο στην είσοδο. Αυτό είναι το είδος της γραμματικής που περιγράφεται στο τμήμα 1.1.1.

1.2.1 Η δύναμη της Γραμματικής Πεπερασμένων Καταστάσεων

Οι γραμματικές πεπερασμένων καταστάσεων αποτελούν μία από τις πιο ευθείς μεθόδους για την μείωση της περιπλοκής σε εφαρμογές αναγνώρισης ομιλίας. Είναι εύκολες στην κατανόηση και, ανάλογα με τις απαιτήσεις της εφαρμογής, μπορεί να είναι σχετικά απλές για να δημιουργηθούν.

Η χρήση μίας γραμματικής πεπερασμένων καταστάσεων απλοποιεί την ανάπτυξη της εφαρμογής αντικαθιστώντας λίστες από επιτρεπόμενες εκφράσεις με τυπικές περιγραφές. Για παράδειγμα, ένα σύστημα αναγνώρισης που χρησιμοποιείται σε ένα εργοστάσιο κατασκευών μπορεί να περιλαμβάνει το πρότυπο

ΠΡΟΤΑΣΗ 1 = αριθμός ανταλλακτικού ΑΡΙΘΜΟΣ $^{\text{πολλαπλάσιο}}$

το οποίο είναι το μόνο που απαιτείται για την αναγνώριση των εκφράσεων εισόδου

αριθμός ανταλλακτικού 1 2 4
αριθμός ανταλλακτικού 2 2 1 5
αριθμός ανταλλακτικού 9 2 5 4 7
αριθμός ανταλλακτικού 3 8 2 9 4 1
αριθμός ανταλλακτικού 8 2 4 2 6 1 7 5 9

και οποιεσδήποτε άλλες που αποτελούνται από τη λέξη "αριθμός ανταλλακτικού" και ακολουθείται από ένα ή περισσότερα ψηφία σε οποιαδήποτε σειρά. Ο δείκτης "πολλαπλάσιο" μετά τη μεταβλητή ΑΡΙΘΜΟΣ δείχνει ότι η γραμματική μπορεί να χειριστεί παραπάνω από ένα ψηφίο. Όμοια, εάν η μεταβλητή ΤΥΠΟΣ ορίζεται ως

ΤΥΠΟΣ = αριθμός ταυτότητος ή Σταθμός ή αριθμός Τμήματος

το πρότυπο γραμματικής

ΠΡΟΤΑΣΗ2 = ΤΥΠΟΣ ΑΡΙΘΜΟΣ $^{\text{πολλαπλάσιο}}$

μπορεί να χρησιμοποιηθεί για την αναγνώριση εκφράσεων όπως

Αριθμός ταυτότητας 1 2 4
Σταθμός 3 6
Αριθμός τμήματος 3

Δομές όπως ΠΡΟΤΑΣΗ1 και ΠΡΟΤΑΣΗ2 μπορούν να συμπεριληφθούν σε άλλες δομές για το σχηματισμό μεγαλύτερων προτάσεων, όπως

ΜΑΚΡΙΑ ΠΡΟΤΑΣΗ=ΠΡΟΤΑΣΗ2ΠΡΟΤΑΣΗ1 αναφορά επιθεώρησης

Χρησιμοποιώντας δομές όπως ΠΡΟΤΑΣΗ1, ΠΡΟΤΑΣΗ2 και ΜΑΚΡΙΑ ΠΡΟΤΑΣΗ σε μία γραμματική πεπερασμένων καταστάσεων μπορεί να αναπαραστήσει ένα μεγάλο εύρος εκφράσεων. Ελλατώνοντας όλες τις πληροφορίες που είναι διαφορετικές από τις ακουστικές του τρέχοντος ενεργού λεξιλογίου, η γραμματική πεπερασμένων καταστάσεων προωθεί την ταχύτητα και την ακρίβεια.

Όταν χρησιμοποιούνται σε εφαρμογές δόμησης, οι γραμματικές πεπερασμένων καταστάσεων είναι γρήγορες, αποτελεσματικές και εύκολες στην κατασκευή. Η απλότητα και στιβαρότητά τους είναι υπεύθυνες για την επιτυχία που έχουν φέρει στην εμπορική αναγνώριση ομιλίας.

1.2.2 Αδυναμία της Γραμματικής Πεπερασμένων Καταστάσεων

Η *αιτιοκρατική* φύση των γραμματικών πεπερασμένων καταστάσεων τις κάνει στερεές. Μόνο οι ακολουθίες λέξεων και προτάσεων που έχουν προγραμματιστεί στη γραμματική μπορούν να αναγνωριστούν. Οι χρήστες δεν μπορούν να αποκλίνουν από αυτά τα πρότυπα. Μόλις ο χρήστης αρχίσει να λέει μια ακολουθία, πρέπει να ολοκληρωθεί με ένα από τους τρόπους που καθορίζονται από τη γραμματική.

Οι γραμματικές πεπερασμένων καταστάσεων δεν καταλαβαίνουν συντακτικά πρότυπα που υπόκεινται στις ακολουθίες λέξεων που είναι προγραμματισμένες να δέχονται. Εξαιτίας αυτού, δεν μπορούν να παρακολουθήσουν την αναγνώρισή τους για να βεβαιώσουν ότι πρότυπα των ανθρώπινων γλωσσών που εξαρτώνται από τα συμφραζόμενα και εμφανίζονται συχνά, είναι σωστά αναγνωρισμένα. Ένα από αυτά τα πρότυπα είναι η συμφωνία υποκειμένου-ρήματος ("Αυτή πηγαίνει" αντί "Αυτοί πηγαίνουν"). Σε αυτό το πρότυπο το ενεργό λεξιλόγιο της δεύτερης λέξης συχνά περιλαμβάνει λέξεις που είναι παρόμοιες ακουστικά, και συχνά συγχέονται, όπως το "πηγαίνει" και "πηγαίνουν", κάνοντας την εφαρμογή ευάλωτη σε λάθη της μορφής "Αυτή πηγαίνουν" και "Αυτοί πηγαίνει". Επιπλέον,

> η γραμματική πεπερασμένων καταστάσεων δεν είναι αρκετά δυνατή για να αυξήσει ενεργά την αναγνώριση σε ευμετάβλητα και μεγάλα συστήματα λεξιλογίου αναγνώρισης ομιλίας. Πολλές κύριες δομές γλώσσας δεν μπορούν να χαρακτηριστούν σωστά στο πεδίο των πεπερασμένων καταστάσεων. Μια πιο δυναμική γραμματική είναι απαραίτητη (L. Miller & S. Levinson, έμπειροι ερευνητές, AT&T Bell Laboratories, "Syntactic analysis for large vocabulary speech recognition using a context - free covering grammar," 1988, p.270).

Συνεπώς, οι γραμματικές πεπερασμένων καταστάσεων δεν είναι κατάλληλες για εφαρμογές, όπως η υπαγόρευση, που απαιτούν μεγάλη ευελιξία.

Μία άλλη αποτυχία των γραμματικών πεπερασμένων καταστάσεων είναι η ανικανότητά τους να κατατάξουν το ενεργό λεξιλόγιο ως πρός τη πιθανότητα εμφάνισής του. Μία ακολουθία εισόδου, όπως,

ΕΠΙΠΛΟ - ΑΝΤΙΚΕΙΜΕΝΟ και ΕΠΙΠΛΟ - ΑΝΤΙΚΕΙΜΕΝΟ

όπου

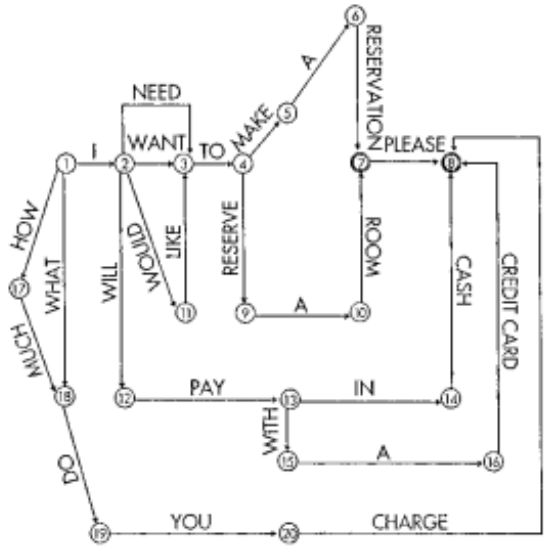
ΕΠΙΠΛΟ - ΑΝΤΙΚΕΙΜΕΝΟ = τραπέζι ή καρέκλα ή καναπές

η λέξη "τραπέζι" είναι πολύ πιθανό να ακολουθείται από τη λέξη "καρέκλα" και λιγότερο πιθανό από μία δεύτερη είσοδο της λέξης "τραπέζι". Οργανώνοντας την έρευνα με χρήση της πιθανότητας εμφάνισης θα βελτίωνε τη ταχύτητα έρευνας και την απόδοση.

Οι γραμματικές πεπερασμένων καταστάσεων δεν έχουν έμφυτο τρόπο για να κωδικοποιούν τέτοια προτεραιότητα μέσα στο ενεργό λεξιλόγιο. Όλα τα μέλη ενός ενεργού λεξιλογίου έχουν την ίδια πιθανότητα εμφάνισης. Μερικά εμπορικά προιόντα αναγνώρισης, όπως το *HARK* του BBN, περιέχουν εργαλεία γραμματικών πεπερασμένων καταστάσεων που έχουν προσαρμοστεί για να επιτρέπουν χειροκίνητη ανάθεση προτεραιοτήτων σε μεμονωμένες λέξεις μέσα στό ενεργό λεξιλόγιο. Αυτές οι αναθέσεις προτεραιότητας βοηθούν στην καθοδήγηση της διαδικασίας έρευνας κατά την αναγνώριση. Μερικά συστήματα έρευνας εφαρμόζουν μία πιό εκταμένη, πιθανοτική γραμματική πεπερασμένων καταστάσεων αυτόματης παραλλαγής.

1.2.3 Υλοποιώντας Γραμματική Πεπερασμένων Καταστάσεων

Οι γραμματικές πεπερασμένων καταστάσεων συνήθως αναπαριστάνονται χρησιμοποιώντας *δίκτυα πεπερασμένων καταστάσεων* (που επίσης ονομάζονται *μηχανές πεπερασμένων καταστάσεων* ή *δίκτυα μεταβάσεων*). Όπως τα "κρυμμένα" μοντέλα Markov, μία γραμματική πεπερασμένων καταστάσεων αποτελείται από καταστάσεις που συνδέονται με αριστερά-προς-δεξιά, κατευθυντικές μεταβάσεις, και αναδρομικές μεταβάσεις (βλ. καφάλαιο 3, παράγραφο 3.1.1). Το δίκτυο πεπερασμένων καταστάσεων που φαίνεται στο σχήμα 1.1 καθορίζει δεκαέξι προτάσεις μίας γραμματικής πεπερασμένων καταστάσεων για μία εφαρμογή κρατήσεων σε ξενοδοχείο με ολικό λεξιλόγιο δεκατεσσάρων λέξεων. Πιστοί στον αντικειμενικό σκοπό της μείωσης της περιπλοκής, η γραμματική του σχεδίου 1.1 έχει μέγιστο βαθμό διακλάδωσης τέσσερα (στην κατάσταση 2).



Σχήμα 1.1 Δίκτυο Πεπερασμένων Καταστάσεων για κρατήσεις ξενοδοχείου

Σε αντίθεση με τα κρυμμένα μοντέλα Markov, ούτε οι καταστάσεις ούτε οι μεταβάσεις ενός δικτύου πεπερασμένων καταστάσεων χαρακτηρίζονται από πιθανότητες. Αφού ένα δίκτυο μίας πεπερασμένης κατάστασης μπορεί να παραστήσει όλες τις επιτρεπτές εισόδους ενός στόχου, τα δίκτυα τείνουν να είναι πολύ μεγαλύτερα από μεμονωμένα κρυμμένα μοντέλα Markov.

Γενικά, οι επιτρεπόμενες ακολουθίες που παριστάνονται από μία γραμματική πεπερασμένων καταστάσεων αναγνωρίζονται, ορίζονται και κωδικοποιούνται από αυτόν που αναπτύσσει την εφαρμογή. Εμπορικά συστήματα έχουν ξεκινήσει να προσφέρουν εργαλεία για την εξαγωγή προτύπων δόμησης από συλλογές δειγμάτων προτάσεων. Μερικές εφαρμογές για interface γραφικών με το χρήστη προσφέρουν αυτόματη εξαγωγή μοντέλου ως ένα στοιχείο των εργαλείων τους για την εξαγωγή λεξιλογίου.

1.2.4 Γραμματική Ζευγαρώματος Λέξης

Η *γραμματική ζευγαρώματος λέξης* μπορεί να θεωρηθεί ως μια παραλλαγή της γραμματικής πεπερασμένων καταστάσεων. Κάθε λέξη στο λεξιλόγιο καθορίζει τις λέξεις που μπορούν να την ακολουθήσουν. Γενικά υλοποιείται ως ένα δίκτυο πεπερασμένων καταστάσεων και μπορεί να χρησιμοποιηθεί ως μοντέλο επιπέδου - λέξης ή υπολέξης. Σε μία γραμματική ζευγαρώματος λέξης για την εφαρμογή κρατήσεων σε ξενοδοχείο που παριστάνεται στο σχήμα 1.1, η λέξη "εγώ" (κόμβος 1) θα καθόριζε τις λέξεις "χρειάζομαι", "θέλω", " θα ήθελα" και "θα" σαν τη λίστα εκείνων που μπορούν να την ακολουθήσουν σε μία έκφραση εισόδου.

Όταν η γραμματική ζευγαρώματος λέξης υλοποιείται χρησιμοποιώντας μοντέλα υπολέξης, κάθε τελική υπολέξη συνδέεται με μία λίστα από πιθανές αρχικές υπολέξεις. Ο Pieraccini, et al. (1991), έδωσαν μία λεπτομερή περιγραφή της υλοποίησης μίας γραμματικής ζευγαρώματος λέξης χρησιμοποιώντας τριφωνικά μοντέλα.

1.3 ΣΤΑΤΙΣΤΙΚΑ ΜΟΝΤΕΛΑ

Τα στατιστικά μοντέλα χρησιμοποιούνται πολύ συχνά σε συστήματα υπαγόρευσης. Αντί να καθορίζουν τι είναι επιτρεπόμενο, τα στατιστικά/πιθανοτικά μοντέλα γλώσσας εξετάζουν τι είναι πιθανό. Αυτές οι προσεγγίσεις επιβεβαιώνουν την ευελιξία και τη μεταβολή της δομής που αποτελούν τη φυσική γλώσσα ομιλίας. Ψάχνουν να ισσοροπήσουν τη φυσικότητα ελέγχοντας την περιπλοκή και την απλότητα της υλοποίησης. Αφού η εστίαση αυτών των γραμματικών είναι η πρόβλεψη ακολουθιών λέξεων, συνήθως καλούνται *στοχαστικά μοντέλα γλώσσας*.

Υπάρχουν δύο βασικές μορφές στατιστικών μοντέλων που συναντώνται σε εμπορικά και ερευνητικά συστήματα:

- Μοντέλα Ν-γραμμάτων
- Μοντέλα Ν-κλάσεων

ΜΟΝΤΕΛΟΠΟΙΗΣΗ Ν-ΓΡΑΜΜΑΤΩΝ

Η προσέγγιση Ν-γραμμάτων υποστηρίχθηκε αρχικά για αναγνώριση ομιλίας στη δεκαετία του 70 από τον Frederick Jelinek της IBM, και αργότερα χρησιμοποιήθηκε για την ανάπτυξη του συστήματος αναγνώρισης *Tangora* της IBM. Κατά το τέλος της δεκαετίας του 80, τα συστήματα Dragon ενσωμάτωσαν δι-γράμματα, μία μορφή μοντελοποίησης Ν-γραμμάτων, στο Dragon Dictate. Από τότε έχει γίνει η κύρια μεθοδολογία γραμματικής σε συστήματα με μεγαλό λεξιλόγιο, όπου συνήθως υλοποιείται ως μοντέλο δι-γραμμάτων ή τρι-γραμμάτων.

1.3.1 Μοντέλα Ν-γραμμάτων

Η *μοντελοποίηση Ν-γραμμάτων* είναι η πιό κονή μορφή στατιστικής μοντελοποίησης που χρησιμοποιείται σε συστήματα ναναγνώρισης ομιλίας.

1.3.1.1 Η δύναμη των μοντέλων Ν-γραμμάτων.

Ενα *μοντέλο Ν-γραμμάτων* καθορίζει την τρέχουσα λέξη (την άγνωστη λέξη) θεωρώντας ότι η ταυτότητα αυτής της λέξης εξαρτάται από τις προηγούμενες Ν-1 λέξεις συν την ακουστική πληροφορία στην άγνωστη λέξη. Ένα μοντέλο τρι-γραμμάτων, όπου Ν=3, θα χρησιμοποιούσε τις δύο λέξεις που ειπώθηκαν πρίν την άγνωστη λέξη. Στην έκφραση

Αυτός είναι ο *εκτυπωτής μου* (άγνωστη λέξη)

η άγνωστη λέξη θα καθοριζόταν από τις δύο λέξεις, "εκτυπωτής μου", που ειπώθηκαν πρίν από αυτή. Το τι καθορίζεται από τη χρήση των λέξεων "εκτυπωτής μου" είναι μία λίστα από υποψήφιες λέξεις που κατατάσσονται με βάση τη πιθανότητα που έχει η υποψήφια λέξη να είναι η άγνωστη λέξη. Η μοντελοποίηση Ν-γραμμάτων αναγνωρίζει την άγνωστη λέξη συνδυάζοντας αυτές τις πιθανότητες με την ακουστική πληροφορία που δίνεται από την άγνωστη λέξη εισόδου. Δεν χρησιμοποιείται άλλη πληροφορία. Κατά την αναγνώριση, το μοντέλο Ν-γραμμάτων αποτελείται από ένα κινούμενο παράθυρο εύρους Ν-1 λέξεων που δημιουργεί λίστες από καταταγμένες υποψήφιες λέξεις. Αυτή η μέθοδος περιορισμού του σκοπού της ανάλυσης συνεισφέρει στην απλότητα και την ταχύτητα της Ν-γραμμάτων προσέγγισης.

Το ενεργό λεξιλόγιο ενός μοντέλου Ν-γραμμάτων μπορεί, θεωρητικά, να περιλαμβάνει ολόκληρο το λεξιλόγιο μίας εφαρμογής, αλλά στις περισσότερες περιπτώσεις περιορίζεται από τη χρήση των κατωφλίων πιθανότητας. Η έρευνα καθοδηγείται από ένα συνδυασμό ακουστικών μοντέλων της εισόδου και της κατάταξης των υποψήφιων λέξεων που βασίζεται στη πιθανότητα εμφάνισης. Αυτή η προσέγγιση παρέχει την ευελιξία που λείπει από τις γραμματικές πεπερασμένων καταστάσεων.

> Στατιστικά μοντέλα γλώσσας βασισμένα σε μοντέλα Ν-γραμμάτων έχει αποδειχτεί ότι ήταν χρήσιμα στην βελτίωση της ακρίβειας των συστημάτων αναγνώρισης ομιλίας. Μία πολύ χρήσιμη πλευρά τους είναι η η απλότητά τους, αφού μεγάλα ποσά από κείμενα για την εκπαίδευση δεδομένων μπορούν να χρησιμοποιηθούν για την εκτίμηση των παραμέτρων του μοντέλου (Marie Mateer, Renssalaer Polytechnic Instityte, & J. Robin Rohlicek, BBN Systems and Technologies, "Statistical language modeling combining N - gram and context - free grammars", 1993, ,p.37).

Η προσέγγιση Ν-γραμμάτων ταιριάζει καλά σε εφαρμογές υπαγόρευσης μεγάλων λεξιλογίων. Προσπαθεί να προσφέρει τη μέγιστη ευελιξία, ταχύτητα και την εύρεση αφθονίας χώρου με το ελάχιστο υπολογιστικό κόστος. Αυτές είναι οι συνεισφορές που υπογραμμίζουν την επιτυχία.

Οι Jelinek (1985) και Jelinek et al. (1975) είναι καλές πηγές επιπλέον τεχνικής πληροφορίας για τη μοντελοποίηση γλώσσας Ν-γραμμάτων.

1.3.1.2 Αδυναμία των μοντέλων Ν-γραμμάτων.

Μια κριτική δύο αιχμών έχει ασκηθεί στα μοντέλα Ν-γραμμάτων που βασίζεται στο περιορισμένο παράθυρο που χρησιμοποιείται για αναγνώριση. Οι παρακάτω δύο ακολουθίες από λέξεις:

> Αυτός είναι ένας εκτυπωτής (άγνωστη λέξη)
> Είναι αναγκαίο για σένα να εισάγεις έναν εκτυπωτή (άγνωστη λέξη)

θα εθεωρούντο ταυτόσημες σε ένα μοντέλο τρι-γραμμάτων γιατί οι δύο τελευταίες γνωστές λέξεις είναι ταυτόσημες. Και στις δύο περιπτώσεις, μόνο αυτές οι δύο λέξεις θα εχρησιμοποιούντο για τη δημιουργία μίας λίστας από υποψήφιες λέξεις για την άγνωστη λέξη. Οι κριτικοί ισχυρίζονται ότι τόσο στενή εστίαση δεν είναι αρκετή για να πιάσει πραγματικές διαφορές που υπάρχουν σε πολλά ιδεατά και γλωσσολογικά πρότυπα. Όταν εξετάζονται ως προς το περιεχόμενο των μακρύτερων δομών των οποίων είναι μέρος, μπορεί να μην είναι καθόλου ισοδύναμοι, όπως δείχνουν τα παρακάτω αποσπάσματα προτάσεων:

> Του κηπορού του αρέσει να *φυτεύει με* (άγνωστη λέξη)
> Το coleus είναι το *φυτό με* (άγνωστη λέξη)

Η κατανόηση της δομής ολόκληρου του αποσπάσματος της πρότασης θα περιόριζε απαράδεκτες επιλογές από τη λίστα των πιθανών επιλογών. Οι λέξεις "φτυάρι" και "ζήλος" μπορεί να είναι λογικές υποψήφιες για την άγνωστη λέξη στην πρώτη πρόταση, αλλά όχι και για τη δεύτερη.

Από τη αλλή πλευρά, πρότυπα δύο ή και τριών λέξεων που καταχωρούνται ως ξεχωριστά επειδή τελειώνουν με διαφορετικές λέξεις, μπορεί να είναι ισοδύναμα πρότυπα. Αν και περιέχουν διαφορετικές λέξεις, οι παρακάτω προτάσεις είναι συντακτικά και νοηματικά παράλληλες:

> Η Μαρία είναι μία καλή συνήγορος.
> Η Rose είναι μία έξοχη δικηγόρος.

Και οι δύο κριτικές είναι αντιδράσεις σε ένα είδος μυωπίας που μπορεί να οδηγήσει σε χωρίς έννοια λάθη αναγνώρισης, όπως η αποδοχή της μη-πρότασης

> Ποία από τις πτήσεις των αερογραμμών Delta;

ως έγκυρης ερώτησης, ή να θεωρήσει ως ένα κείμενο χωρίς νόημα την παρακάτω ακολουθία λέξεων:

> Εαν δεν χρειάζεται να είσαι μία καλή συμφωνία του κόσμου.
> Εγώ λέω.
> Αυτή ήταν μία καλή συμφωνία του κόσμου. Αλλά το γεγονός ότι η μόνη στο κόσμο. Όταν η πρώτη φορά στο κόσμο (Latin Bahl, Peter Brown, Peter deSouza & Robert Mercer, έμπειροι ερευνητές, IBM Watson Research Center, "A tree - based statistical language model for natural language speech recognition", 1989, p.1008).

Σε αυτά τα παραδείγματα, διαδοχικά πρότυπα τριών λέξεων είναι αποδεκτά, αλλά οι μακρύτερες ακολουθίες δεν έχουν νόημα.

> Ενώ τα μοντέλα Ν-γραμμάτων βασισμένα στη στατιστική έχουν αποδειχθεί ότι είναι αποτελεσματικά για την αναγνώριση ομιλίας, υπάρχει, γενικά, περισσότερη παρουσία δομής στη φυσική γλώσσα από

> ότι μπορούν να δώσουν τα μοντέλα Ν-γραμμάτων (Marie Mateer, Renssalaer Polytechnic Instityte, & J. Robin Rohlicek, BBN Systems and Technologies, "Statistical language modeling combining N-gram and context-free grammars", 1993, ,p.37).

Επομένως, εάν ο σκοπός της χρήσης αναγνώρισης ομιλίας είναι να αντιληφθεί και να αποδώσει τι θέλει να πεί ο ομιλητής, τα μοντέλα Ν-γραμμάτων είναι από μόνα τους ανεπαρκή. Εάν, παρ' όλα αυτά, ο στόχος της αναγνώρισης είναι να καθορίσει κατά λέξη τι λέγεται, μία προσέγγιση Ν-γραμμάτων δουλεύει πολύ καλά.

> Εάν ο στόχος είναι να αναγνωρίσει τι λέει το πρόσωπο, πρέπει να αναγνωρίσει κανείς αυτές τις μη γραμματικές προτάσεις. Είναι πιό σημαντικό να αποδεχτεί το σωστό παρά να απορίψει το λάθος. Η προσέγγιση Ν-γραμμάτων θα αποδεχθεί οποιαδήποτε πρόταση που έχει κάποιο είδος προσαρτημένης πιθανότητας. Αυτό δίνει στη προσέγγιση κάποιου είδους στιβαρότητα που δεν έχουν τα πιο υψηλά δομημένα, γλωσσολογικά μοντέλα (Richard Schwartz, έμπειρος ερευνητής, BBN Systems and Technologies, personal communication. 1994).

Η αδυναμία του μοντέλου Ν-γραμμάτων για την κατανόηση της γλώσσας γίνεται πλεονέκτημα όταν χρησιμοποιείται για υπαγόρευση και άλλες κατά λέξη εγγραφές ομιλούμενης γλώσσας, όπως οι αναφορές των δικαστηρίων.

1.3.1.3 Υλοποιώντας τα μοντέλα Ν-γραμμάτων.

Τα μοντέλα Ν-γραμμάτων εξάγουν πρότυπα ακολουθιών λέξεων απευθείας από μεγάλο όγκο δεδομένων γλώσσας. Όταν βρεθούν, αυτές οι ακολουθίες κωδικοποιούνται ως προς την πιθανότητα ότι η Ν-ιοστή λέξη της ακολουθίας θα ακολουθεί τις προηγούμενες Ν-1 λέξεις. Αυτή η διαδικασία καλείται *εκπαίδευση* του συστήματος.

Οι διαδικασίες εκπαίδευσης παράγουν ένα μοντέλο γλώσσας Ν-γραμμάτων. Αυτό το μοντέλο γενικά παριστάνεται ως ένα πλέγμα ή συνδεδεμένη λίστα της οποίας οι δεσμοί περιέχουν τις πιθανότητες που βρίσκονται κατά την εκπαίδευση. Κατά τη διάρκεια της λειτουργίας μίας εφαρμογής, αυτές οι πιθανότητες χρησιμοποιούνται για να επιλέξουν, να ελαττώσουν και να κατατάξουν τις υποψήφιες λέξεις.

Το ποσό των δεδομένων γλώσσας που απαιτείται για τη δημιουργία ενός καλού μοντέλου γλώσσας αυξάνει με το μέγεθος του λεξιλογίου της εφαρμογής και το μέγεθος του Ν. Ακόμα και ένα μικρό Ν, όπως το 3, απαιτεί πολλά δεδομένα γλώσσας. Ο Jelinek (1985) υπολόγισε ότι για να βρεί όλα τα πιθανά τρι-γράμματα (N=3) για ένα λεξιλόγιο πέντε χιλίαδων λέξεων θα απαιτείτο ένα σώμα 125 τρισεκατομμύρια λέξεων.

> Δυστυχώς, αυτές οι πιθανότητες δεν μπορούν να προσεγγιστούν απευθείας από τις σχετικές συχνότητες που αποκτούνται καταμετρώντας τρι-γράμματα που εμφανίζονται σε μεγάλα κείμανα, αφού η μεγάλη πλειοψηφία των πιθανών Αγγλικών λέξεων τρι- γραμμάτων δεν θα υπάρχει ακόμα και σε μεγάλες βάσεις δεδομένων (Frederick Jelinek, IBM [τώρα στο Πανεπιστήμιο του Pittsburgh], "The development of an experimental discrete dictation recognizer", 1985, p.589).

Οι πιθανότητες των δι-γραμμάτων (N=2) και του ενός-γράμματος (N=1) χρησιμοποιούνται για να συμπληρώσουν τα τρι-γράμματα που λείπουν. Για κάθε ακολουθία τρι-γραμμάτων

> λέξη1 λέξη2 λέξη3

η πρόβλεψη δι-γραμμάτων για τη λέξη2, δεδομένης της λέξης1, και η πρόβλεψη δι-γραμμάτων για τη λέξη3, δεδομένης της λέξης2, υπολογίζονται και χρησιμοποιούνται για να εξάγουν την πρόβλεψη τρι-γραμμάτων για τη λέξη3 στην ακολουθία. Έαν η ακολουθία λέξη1 λέξη2 δεν εμφανίζεται στο σώμα, η πρόβλεψη ενός-γράμματος (εμφάνιση μονής-λέξης) χρησιμοποιείται για να κάνει μία εκτίμηση δι-γράμματος. Αυτή η διαδικασία της μετακίνησης από τρι-γράμματα σε ένα-γράμμα καλείται *οπισθοχώρηση*.

Εναλλακτικές διαδικασίες στην οπισθοχώρηση εξερευνούνται σε συστήματα έρευνας. Κάποιες περιλαμβάνουν συλλογή επιπρόσθετων δεδομένων από τη δομή των λέξεων (που καλείται *μορφολογία*), μέρος της ομιλίας, ή σημασιολογική κλάση (όπως *ημερομηνία* ή *αριθμός*). Ο Jelinek (1990) πρότεινε να περιλαμβάνονται *συντοποθετήσεις* για τις λέξεις χαμηλής συχνότητας. Οι συντοποθετήσεις είναι λέξεις που χρησιμοποιούνται συχνά μαζί με μία λέξη στόχο αλλά δεν είναι απαραίτητα γειτονικές με αυτή. Η λέξη "δίσκος", για παράδειγμα, είναι συντοποθέτηση με τη λέξη "φλυτζάνι". Οι ερευνητές που ειδικεύονται στην επανάκτηση πληροφορίας και επεξεργασία φυσικής γλώσσας έχουν εξάγει ένα σημαντικό ποσό δεδομένων για συντοποθετήσεις από τυπωμένα υλικά. Και οι μεθοδολογίες και τα ευρήματα μπορούν να εισαχθούν στην αναγνώριση ομιλίας.

Λεπτομερής τεχνική πληροφορία για την υλοποίηση των μοντέλων N-γραμμάτων μπορεί να βρεθεί σε οποιαδήποτε δημοσίευση του Jelinek. Τεχνικές περιγραφές της οπισθοχώρησης και μερικές εναλλακτικές τεχνικές περιγράφονται από το Jelinek (1985 και 1990) και το Matlese & Mancini (1992). Περισσότερες πληροφορίες για τις συντοποθετήσεις και τις άλλες λεξικές συγγένειες μπορούν να βρεθούν στους Evens et al. (1980), Grishman & Sterling (1993) και Markowitz et al. (1936 και 1992).

1.3.1.4 Προσωποποιώντας τα μοντέλα N-γραμμάτων.

Οταν ένα σύστημα υπαγόρευσης επηρεάζει το στύλ και τις προτιμήσεις λεξιλογίου ενός χρήστη, είναι και πιο εύκολο στη χρήση και πιο αποδοτικό. Η διαμόρφωση ενός συστήματος με βάση τη συμπεριφορά του χρήστη καλείται *προσωποποίηση*. Υπάρχουν δύο βασικοί τρόποι για την προσωποποίηση ενός συστήματος. Ο ένας αναφέρεται στην ικανότητα να προσθέτει νέα τμήματα λεξιλογίου (βλ. κεφάλαιο 3, παράγραφος 1.4.3). Ο άλλος περιλαμβάνει τη χρήση μίας εσωτερικής γραμματικής που αντανακλά το στύλ και τα πρότυπα γλώσσας του χρήστη.

Όταν ένας ομιλητής δημιουργεί μία καινούργια λέξη, το σύστημα πρέπει να της αναθέσει μια πιθανότητα λέξης-ακολουθίας. Αφού δεν έχει πληροφορία εκπαίδευσης πάνω στην οποία να βασίσει την ανάθεση, πρέπει να βρεί άλλη μέθοδο. Η προσέγγιση που χρησιμοποιείται γενικά είναι να βασίσει την πιθανότητα σε γενικές εκτιμήσεις λέξης-συχνότητας. Αυτή η πρακτική συχνά έχει ως αποτέλεσμα χονδροειδώς μεγάλες ή μικρές πιθανότητες. Μία άλλη προσέγγιση. που χρησιμοποιείται από την IBM, καλείται *προσαρμοζόμενη κρύπτη*. Προσαρμόζει τις πιθανότητες των δι-γραμμάτων στο υπάρχον γλωσσικό μοντέλο για την αντανάκλαση καταστάσεων που περιλαμβάνουν πρόσφατα προσθετιμένες λέξεις.

Η διαδικασία της προσωποποίησης της γραμματικής μπορεί να περιγραφεί με παράδειγμα με δύο τεχνικές που χρησιμοποιούνται στο *Dragon Dictate* των Dragon Systems. Προγενέστερο της χρησιμοποίησης του συστήματος σε ένα οργανισμό, το σύστημα μπορεί να προσαρμόζει την εσωτερική του γραμματική ψάχνοντας πρότυπα λέξης-ακολουθίας που

βρίσκονται σε σχετικά κείμενα του οργανισμού. Κατά τη διάρκεια λειτουργίας του συστήματος, θα τροποποιήσει τις αναθέσεις πιθανότητας για να ανακλάσει καλύτερα αυτές των χρηστών του.

1.3.2 Μοντέλα N- κλάσεων

Τα μοντέλα N-κλάσεων επεκτείνουν τη σύλληψη της μοντελοποίησης N-γραμμάτων σε συντακτικές (ή σημασιολογικές) κατηγορίες. Η μοντελοποίηση δι-κλάσεων υπολογίζει την πιθανότητα ότι δύο κατηγορίες θα εμφανίζεται σε διαδοχή. Η μοντελοποίηση τρι-κλάσεων επεκτείνει τα πρότυπα συχνότητας σε τρείς διαδοχικές κλάσεις. Σε ένα σύστημα δι-κλάσεων για τα Αγγλικά, για παράδειγμα, οι λέξεις “a”, “an” και “the” μπορεί να ανήκουν στην ομάδα κλάση *άθρο* και οι λέξεις όπως “table”, “book” and “shoe” μπορεί να είναι μέλη της κλάσης *αριθμήσιμο-ουσιαστικό*. Η πιθανότητα εμφάνισης της δι-κλάσης *άρθρο αριθμήσιμο ουσιαστικό* υπολογίζεται βρίσκοντας:

1. Ολικές εμφανίσεις της κατηγορίας *άρθρο* στο κείμενο εκπαίδευσης (ολικά - άρθρα)
2. Ολικές εμφανίσεις της κατηγορίας *άρθρο* που ακολουθείται από την κατηγορία *αριθμήσιμο - ουσιαστικό* (ακολουθία)
3. ολικά - άρθρα / ακολουθία

Αυτή η διαδικασία απαιτεί όλες οι λέξεις των δεδομένων να είναι κατηγοριοποιημένες. Αν και κάποια κατηγοριοποίηση μπορεί να γίνει αυτόματα, η διαδικασία εκπαίδευσης της γλώσσας επιμηκύνεται.

Η μοντελοποίηση δι-κλάσεων και τρι-κλάσεων έχει υλοποιηθεί στην IBM και σε αρκετές Ευρωπαικά συστήματα αναγνώρισης. Μία Γαλλική υλοποίηση δημιούργησε μοντέλα δι-κλάσεων και τρι-κλάσεων για 150 κατηγορίες από ένα λεξικό 170000 λέξεων. Στις Η.Π.Α., η χρήση της μοντελοποίησης των N-κλάσεων μεγαλώνει ως μέσο για τη μείωση μερικών θεμάτων εκπαίδευσης μοντελοποίησης των N-γραμμάτων και ως εναλλακτική τεχνική αντί της οπισθοχώρησης.

Η μοντελοποίηση των N-κλάσεων έχει βρεθεί ότι παράγει ακριβή, συμπαγή μοντέλα γλώσσεων από σώματα πολύ μικρότερα από αυτά που απαιτούνται γιαι τα μοντέλα των N-γραμμάτων.

> Το πόσο των δεδομένων που απαιτούνται για την εκπαίδευση ενός μοντέλου τρι- γραμμάτων, θα είναι μεγάλο. Στην περίπτωση της χρήσης γραμματικών κατηγοριών [μοντέλα N-κλάσεων], το κείμενο πρέπει να έχει τίτλο, αλλά μπορεί να είναι μικρότερο. Επιπλέον, εάν μία καινούργια λέξη εισάγεται στο λεξιλόγιο μπορεί να κληρονομήσει τις πιθανότητες που υπολογίζονται για τις λέξεις που έχουν την ίδια γραμματική κατηγορία.

Αυτές είναι σημαντικές ιδιότητες για μία γραμματική γιατί διευκολύνουν την κατασκευή της γραμματικής και την χρήση συστημάτων μεγάλου λεξιλογίου σε νέα πεδία. Τα πρότυπα γλώσσας και τα λεξιλόγια που χρησιμοποιούνται σε επιχειρήσεις εξαγωγών είναι, για παράδειγμα, απίθανο να μοιάζουν με αυτά που χρησιμοποιούνται σε μεσητείες.

Ο Mariani (1993) δείνει μία τέλεια περιγραφή της μοντελοποίησης των δι-κλάσεων και των τρι-κλάσεων. Οι Pieraccini & Levin (1992) περιγράφουν μία παρόμοια προσέγγιση που αναπτύσεται στα Εργαστήρια Bell της AT&T, και οι Placeway et al. (1993)

περιφράφουν τη χρήση της μοντελοποίησης N-κλάσεων με μικρά και μεγάλα σώματα. Οι Jelinek et al. (1975), περιέχουν μία τεχνική περιγραφή μίας πρώιμης πρότασης N-κλάσεων και οι Jelinek, et al. (1990), περιγράφουν πως μία προσέγγιση N-κλάσεων μπορεί να επηρεάσει καλύτερη ολοκλήρωση νέων λέξεων σε ένα μοντέλο τρι-γραμμάτων.

1.4 ΓΡΑΜΜΑΤΙΚΕΣ ΒΑΣΙΣΜΕΝΕΣ ΣΤΗ ΓΛΩΣΣΟΛΟΓΙΑ

Γραμματικές σχεδιασμένες από γλωσσολόγους έχουν σκοπό να περιγράψουν όλες τις ιδιότητες της προφορικής επικοινωνίας που δίνεται στο σχήμα 1.2.

Ακουστικό σύστημα	→ Φωνητικά και Φωνολογία
Λέξεις και δομές λέξεων	→ Μορφολογία και Λεξικολογία
Φράσεις και δομές προτάσεων	→ Σύνταξη
Νόημα	→ Σημασιολογία
Κοινωνικοί κανόνες επικοινωνίας	→ Πραγματείες

Σχήμα 1.2 Περιεχόμενα μιας γραμματικής βασισμένης στη γλωσσολογία

Τέτοιες γραμματικές μπορεί να περιέχουν μία γραμματική πεπερασμένης κατάστασης και/ή ένα πιθανοτικό μοντέλο γλώσσας σαν στοιχείο, αλλά η εστίαση είναι στο σχεδιασμό ενός συστήματος που καταλαβαίνει την έννοια του τι είπε ο χρήστης καθώς επίσης (ή μερικές φορές, αντί αυτού) καθορίζει τις λέξεις που έχουν ειπωθεί. Αυτός ο στόχος μετασχηματίζει την διαδικασία αναγνώρισης ομιλίας που βασίζεται στην ακουστική σε γνωσιολογική-γλωσσολογική διαδικασία *της κατανόησης της ομιλούμενης γλώσσας*. Πρακτικά οι γραμματικές που βασίζονται στην γλωσσολογία για την κατανόηη της ομιλούμενης γλώσσας είναι συστήματα έρευνας.

ΓΛΩΣΣΟΛΟΓΙΚΕΣ ΓΡΑΜΜΑΤΙΚΕΣ

Οι γλωσσολογικές γραμματικές υπερίσχυσαν της προσπάθειας για την έρευνα ομιλίας στα μέσα της δεκαετίας του 70 αλλά ήταν περαγκωνισμένες από άλλες προσεγγίσεις πιο ικανές να μεταφερθούν από το εργαστήριο στη αγορά. Στη δεκαετία του 90 το ενδιαφέρον για τις γλωσσολογικές γραμματικές είναι ένα φυσικό υποπροιόν της πρόσφατης ανάπτυξης στο υπολογισμό της δύναμης, της μεγαλύτερης κατανόησης των γλωσσολογικών αρχών και της ωρίμανσης της αναγνώρισης ομιλίας. Όπως και τη δεκαετία του 70, η πρόσφατη έκρηξη της έρευνας για τη χρήση γλωσσολογικές γραμματικές έχει αναπτυχθεί και χρηματοδοτηθεί από την ARPA, άλλες ομοσπονδιακές εταιρείες και προγράμματα που χρηματοδοτούνται από το κράτος στην Ευρώπη και την Ασία. Ο σκοπός της χρηματοδότησης είναι να συνδέσει την έρευνα ομιλίας με τη διαδικασία της φυσικής γλώσσας ως ένα τρόπο για να δημιουργήσει συστήματα ικανά να κάνουν μετάφραση ομιλίας σε ομιλία ή ομιλίας σε κείμενο και συστήματα ικανά να συμετέχουν σε φυσική προφορική αλληλεπίδραση με ανθρώπους (που καλούνται ομιλία).

Υπάρχει μία ποικιλία από λόγους για τους οποίους η βιομηχανία αναγνώρισης ομιλίας θα ήθελε να καταλαβαίνει τη γλώσσα αντί να αναγνωρίζει απλώς την ομιλία. Ο ένας είναι ότι η χρήση επιπρόσθετης γνώσης, όπως είναι οι στόχοι του χρήστη και τα θέματα που έχουν ήδη καληφθεί στη συζήτηση, θα λειάνει το ενεργό λεξιλόγιο απορρίπτοντας μη λογικές ή απίθανες επιλογές.

> Κατάλληλη χρήση περιορισμών από την επεξεργασία της φυσικής γλώσσας μειώνει την περιπλοκή του καθήκοντος αναγνώρισης ομιλίας, αυξάνει την ακρίβεια αναγνώρισης λέξης (Hy Murveit & Robert Moore, SRI International, "Intergrating natyral language constrains into HMM-based speech recognition," 1990, p.573).

Ένας άλλος λόγος είναι οτι γνώση της γνωσιολογίας, της συμπεριφοράς και της γλωσσολογίας είναι φυσικά στοιχεία της ανθρώπινης αλληλεπίδρασης.

> Πολλές από τις εφαρμογές που είναι κατάλληλες για την αλληλεπίδραση ανθρώπου/μηχανής που χρησιμοποιούν ομιλία εμπλέκουν τυπικά ουν την επίλυση του προβλήματος αλληλεπίδρασης. Αυτό σημαίνει, επιπρόσθετα της μετατροπής του σήματος ομιλίας σε κείμενο, οτι ο υπολογιστής πρέπει επίσης να καταλαβαίνει τις απαιτήσεις του χρήστη, για να δημιουργεί μία απόκριση (Victor Zue, James Glass, David Goodine, Lynette Hirschman, Hong Leung, Micheal Phillips, Joseph Polifroni & Stephanie Seneff, MIT, "From speech recognition to spoken language understanding: The development of the MIT SUMMIT and VOYAGER systems," 1991, p.256).

Αναγνώριση ομιλίας που είναι φυσική, ευέλικτη, φιλική στο χρήστη, που μοιάζει στην ανθρώπινη, είναι ένα υποπροιόν που περιλαμβάνει μη ακουστική γνώση για την ανθρώπινη επικοινωνία.

Οι γραμματικές που είναι βασισμένες στη γλωσσολογία έχουν πάρει πολλές μορφές. Οι περισσότερες περιέχουν ξεχωριστά κομμάτια για κάθε μία από τις λειτουργίες που δίνονται στο σχήμα 1.2. Μερικές είναι βασισμένες σε αυστηρές θεωρίες γλωσσολογίας.

1.4.1 Γραμματική Ελεύθερου Περιεχομένου

Η γραμματική ελεύθερου περιεχομένου αναπαριστάνει την ευρύτερα χρησιμοποιούμενη προσέγγιση της συντακτικής ανάλυσης που βρίσκεται σε συστήματα αναγνώρισης ομιλίας βασισμένα στη γλωσσολογία. Μπορεί να υλοποιείται μόνη της ή σαν στοιχείο ενός μεγαλύτερου μοντέλου.

1.4.1.1 Η δύναμη της γραμματικής ελεύθερου περιεχομένου.

Οπως και οι γραμματικές πεπερασμένων καταστάσεων, οι γραμματικές ελεύθερου περιεχομένου είναι *αιτιοκρατικές*: ορίζουν επιτρεπόμενες δομές. Η δυνατότητα τους να αναπαριστάνουν και να ελέγχουν σχέσεις μέσα στην έκφραση τις κάνει πιο δυνατές από τις γραμματικές πεπερασμένων καταστάσεων. Μία γραμματική πεπερασμένων καταστάσεων μπορεί να χρησιμοποιηθεί για μία εφαρμογή που απαιτεί την κατασκευή ενός κώδικα αναγνώρισης που αποτελείται από ψηφία που ακολουθούνται από γράμματα. Μία γραμματική πεπερασμένων καταστάσεων δεν θα μπορούσε να χρησιμοποιηθεί για ένα κώδικα που αποτελείται από ένα μεταβλητό αριθμό ψηφίων που ακολουθείται από ένα ίσο

αριθμό γραμμάτων. Για αυτή την εφαρμογή θα χρειάζονταν μία γραμματική ελεύθερου περιεχομένου, όπως και για εφαρμογές που απαιτούν αριθμητική συμφωνία μεταξύ υποκείμενου και ρήματος μέσα στην πρόταση. Αυτό επιτυγχάνεται μέσω της χρήσης επανεγγραφής κανόνων και αναπαραστάσεων δένδρου, όπως φαίνεται στα σχήματα 1.3 και 1.4.

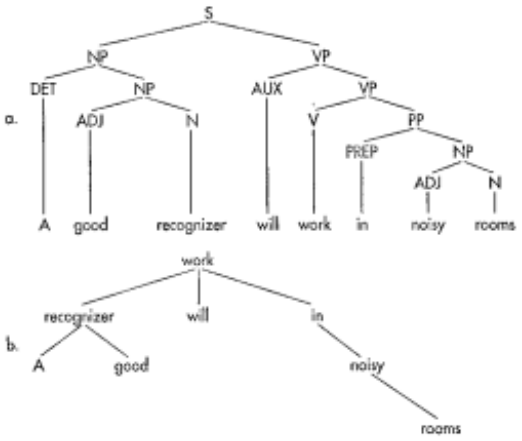
Κάθε κανόνας δείχνει μία κατηγορία, ή σύμβολο, στην αριστερή πλευρά του το οποίο μετετρέπεται (καλείται *επαναγεγραμμένο*) σε ένα ή περισσότερα άλλα σύμβολα. Η αριστερή κατηγορία εμπεριέχει τις κατηγορίες στα δεξιά. Ο πρώτος κανόνας στο σχήμα 1.3, για παράδειγμα, ξαναγράφει την κατηγορία *πρόταση* σαν μία ακολουθία δύο κατηγοριών: μία *φράση ουσιαστκού* που ακολουθείται από μία *φράση ρήματος*. Μία πλήρης γραμματική θα περιείχε πολλούς τέτοιους *κανόνες επανεγγραφής* για την κατηγορία *πρόταση* και τις άλλες γλωσσολογικές κατηγορίες που περιλαμβάνονται στη γραμματική. Οι κανόνες επανεγγραφής του σχήματος 1.3 παρουσιάζονται γραφικά σαν δένδρο στο σχήμα 1.4α.

	$S \rightarrow NP\ VP$	$VP \rightarrow AUX\ VP$
	$NP \rightarrow DET\ NP$	$VP \rightarrow V\ PP$
	$NP \rightarrow ADJ\ N$	$PP \rightarrow PREP\ NP$
όπου:	S= πρόταση	ADJ= αντικείμενο
	NP= φράση ουσιαστικού	N= ουσιαστικό
	VP= φράση ρήματος	AUX=βοηθητικό ρήμα (είναι, μπορεί, κ.α.)
	PP= φράση πρόθεσης	DET= άρθρο (οριστικό και αόριστο)
	V= ρήμα	PREP= πρόθεση (πάνω, πίσω, κ.α.)

Σχήμα 1.3 Κανόνες ελεύθεροι περιεχομένου

Μέρος της ευελιξίας των γραμματικών ελεύθερου περιεχομένου είναι ότι μπορούν να αναπαραστήσουν μία ποικιλία από γλωσσολογικές προσεγγίσεις. Οι κανόνες ελεύθερου περιεχομένου του σχήματος 1.3 και το αντίστοιχο δένδρο του σχήματος 1.4α, για παράδειγμα, απεικονίζουν μία απευθείας συντακτική γραμματική. Το δένδρο στο σχήμα 1.4β και οι κανόνες ελεύθερου περιεχομένου που ενυπάρχουν, αναπαριστάνουν μία θεωρία γραμματικής κεφαλής-τροποποιητή.

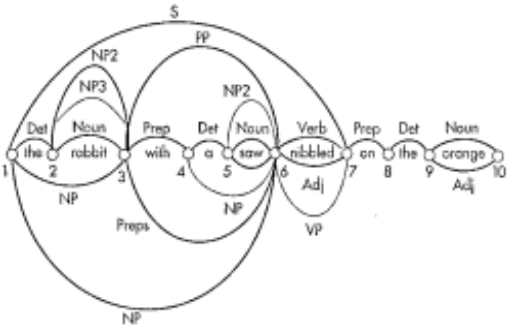
Η δύναμη, η ευελιξία και η απλή αναπαράστασή τους, έχουν κάνει τις γραμματικές ελεύθερου περιεχομένου μία δημοφιλή επιλογή για γλωσσολόγους, επιστήμονες υπολογιστών και ερευνητές συστημάτων αναγνώρισης ομιλίας. Ο Winograd (1983, κεφάλαιο 3) περιγράφει λεπτομερώς τις γραμματικές ελεύθερου περιεχομένου. Ο Ney (1991) παρέχει ένα παράδειγμα συνδυασμού των κανόνων ελεύθερου περιεχομένου με πιθανότητες.



Σχήμα 1.4 Αναπαράσταση δένδρου για την γραμματική ελεύθερου περιεχομένου του σχήματος 1.3

1.4.1.2 Υλοποιώντας τις γραμματικές ελεύθερου περιεχομένου.

Μία από τις πιο συχνές υπολογιστικές υλοποιήσεις των γραμματικών ελεύθερου πειεχομένου είναι το *διάγραμμα σύνταξης*. Το διάγραμμα σύνταξης αποτελεί μία δυναμική προσέγγιση της σύνταξης. Εφαρμόζει τους κανόνες ελεύθερου περιεχομένου της γραμματικής στην ομιλούμενη είσοδο και ακολουθεί τους κανόνες που ήταν επιτυχημένοι τοποθετώντας τους σε ένα διάγραμμα. Το σχήμα 1.5 δείχνει ένα διάγραμμα που περιέχει μία μερική σύνταξη της πρότασης "Ο λαγός με ένα κοπτήρα βυθισμένο σε ένα πορτοκάλι". Δείχνει ότι η δομή του διαγράμματος είναι παρόμοια με αυτή ενός δικτύου πεπερεσμένων καταστάσεων. Καθώς το διάγραμμα σύνταξης δουλεύει την ομιλία εκτείνει και απορρίπτει τις εναλλαγές μέχρι να έχει μία αναπαράσταση ολόκληρης της έκφρασης. Ο Winograd (1983, κεφάλαιο 3) παρέχει μία λεπτομερή περιγραφή των διαγραμμάτων σύνταξης.



Σχήμα 1.5 Γραμματικός Αναλυτής γραφικών όπου φαίνεται το γράγημα για την πρόταση " Ο λαγός με ένα κοπτήρα βυθισμένο σε ένα πορτοκάλι " (Winograd, Η γλώσσα ως μια γνωστική διαδικασία, Τόμος Ι Σύνταξη, Copyright 1983, Addison-Wesley)

1.4.2 Γραμματικές πολλαπλών πηγών γνώσης

Η έρευνα αναγνώρισης φωνής παράγει έναν υψηλά διαφοροποιημένο πίνακα γραμματικών που συνδυάζει συντακτική ανάλυση (συχνά γραμματική ανεξάρτητη περιεχομένου) με σημασιολογία, στατιστικά και άλλες πηγές γνώσης.

Το *Σύστημα Ομιλούμενης Γλώσσας* (*Spoken Language System*), αναπτυγμένο από την BBN, είναι ένα καλό παράδειγμα του πώς διαφορετικές προσεγγίσεις έχουν συνδυαστεί σε ένα μόνο σύστημα κατανόησης ομιλουμένης γλώσσας. Η *BYBLOS,* που είναι το στοιχείο αναγνώρισης φωνής του συστήματος, περιέχει δύο και τριών γραμμάτων μοντέλα γλώσσας υποστηριζόμενα από μία παραλλαγμένη μορφή backing-off. Έχει συγχωνευτεί με το *DELPHI,* το σύστημα κατανόησης γλώσσας του BBN. Το *DELPHI* περιέχει μια γραμματική ανεξάρτητη περιεχομένου εφαρμοσμένη ως αναλυτής γραφικών (chart parser) . Η ανεξάρτητη περιεχομένου γραμματική έχει επεκταθεί ώστε να ενσωματώνει γνώσεις σημασιολογίας. Το *DELPHI* χρησιμοποιεί στατιστικά για να κατατάξει τους ανεξάρτητους περιεχομένου κανόνες για να βοηθήσει στην επεξεργασία πρόχειρα σχηματισμένων ειπομένων δεδομένων (poorly-formed spoken input). Το σημείο επαφής ανάμεσα στη *BYBLOS* και στο *DELPHI* είναι μια λίστα των N καλύτερων υποψήφιων λέξεων.

Το σύστημα *TINA* του MIT συνδυάζει ανεξάρτητους περιεχομένου κανόνες με πιθανότητες. Όπως με το *Σύστημα Ομιλούμενης Γλώσσας*, οι στατιστικές πιθανότητες βοηθούν στη μείωση των αναγκών διεργασίας του γλωσσολογικού μοντέλου. Το TINA χρησιμοποιεί ένα επαυξημένο δίκτυο μετάβασης (ATN augmented transition network). Η δομή ενός ATN είναι παρόμοια με αυτή ενός δικτύου περιορισμένων καταστάσεων (finite state network). Αντίθετα προς τα δίκτυα περιορισμένων καταστάσεων, οι κόμβοι ενός ATN περιέχουν επιπρόσθετες πληροφορίες γραμματικής που χρησιμοποιούνται στην επιλογή της εξερχόμενης διαδρομής (outgoing path) από ένα κόμβο. Αυτές οι πληροφορίες μπορούν να αναπαραστήσουν ανεξάρτητους περιεχομένου κανόνες ή να περιέχουν πιο λεπτομερή

γλωσσολογικά δεδομένα. Οι καταστάσεις του ATN του *TINA* περιέχουν συντακτικές και σημασιολογικές πληροφορίες. Σε αντίθεση με τυπικά ATN και δίκτυα περιορισμένων καταστάσεων, οι σύνδεσμοι ενός ATN του *TINA* περιέχουν πιθανότητες που απορρέουν από στατιστική εκπαίδευση.

Σχεδιαστές συστημάτων κατανόησης ομιλούμενης γλώσσας συνειδητοποιούν ότι η πραγματική κατανόηση λόγου πρέπει να προχωράει πέρα από την πρόταση για να περιλαμβάνει ανάλυση της συνομιλητικής αλληλεπίδρασης με τον χρήστη (επικαλούμενη ως διαδικασία συνομιλίας, discourse processing). Το σύστημα της CMU, *MINDS,* είναι ένα παράδειγμα ενός συστήματος σχεδιασμένο να κάνει διαδικασία συνομιλίας. Περιέχει τριφωνικά, δύο γραμμάτων και τριών τάξεων μοντελοποίηση, δίκτυα περιορισμένων καταστάσεων, κανόνες για είδος λέξης και συναρτήσεις σημασιολογίας, πλαίσια (μια δομή αντικειμενοσταφή προγραμματισμού) που περιέχουν την ιστορία της συνομιλίας, γνώση του ομιλιτή και του θέματος ομιλίας, και επίπεδα *πραγματιστικών περιορισμών* (*pragmatic constraints*). Οι πραγματιστικοί περιορισμοί είναι πηγές γνώσεις χρησιμοποιούμενες για να προβλέψουν το περιεχόμενο της επόμενης λέξης του χρήστη. Προβλέψεις που παράγονται από τους περιορισμούς ταξινομούνται σύμφωνα με το πόσο έλεγχο επιβάλλουν στην απόκριση του χρήστη και στη πιθανότητά τους να συμβούν. Αυτά τα επίπεδα πραγματιστικών περιορισμών επιτρέπουν στο *MINDS* να κάνει ξανά και ξανά γραμματική ανάλυση των δεδομένων έως ότου ένα καλό ταίριασμα για αυτά βρεθεί. Συγκριτικός έλεγχος αναγνώρισης για χίλιες λέξεις , έδειξε πτώση στην πολυπλοκότητα από 279 (χωρίς περιορισμούς) σε 18 (με περιορισμούς) .

Το *JANUS* είναι ένα σύστημα μετάφρασης ομιλούμενης γλώσσας αναπτυγμένο ως μια προσπάθεια ένωσης καλούμενη C-STAR εμπλέκοντας τη CMU, το Πανεπιστήμιο της Καρλσρούης, την Siemens AG, και το Ινστιτούτο Έρευνας Προηγμένων Τηλεπικοινωνιών. Η διαδικασία αναγνώρισης ξεκινάει με ένα νευρωνικό δίκτυο (διάφορες παραλλαγές έχουν χρησιμοποιηθεί) για να αναλύσει ως δεδομένα εισόδου συνεχή ομιλία. Το δίκτυο στέλνει τα αποτελέσματά του σε μια αναζήτηση των N καλύτερων για να αναγνωρίσει και να ταξινομήσει από 6 έως 100 εφαρμόσιμες υποθέσεις σε επίπεδο πρότασης. Αυτές οι υποθέσεις φιλτράρονται και επαναταξινομούνται από ένα τριών γραμμάτων μοντέλο και στέλνονται σε έναν ειδικευμένο ανεξάρτητο περιεχομένου γραμματικό αναλυτή. Εάν υπάρχουν διαστρεβλωμένα δεδομένα εισόδου ή αν ο γραμματικός αναλυτής παρουσιάζει δυσκολία να αναγνωρίσει τα δεδομένα εισόδου, ένας γραμματικός αναλυτής νευρωνικού δικτύου , που ονομάζεται *PARSEC,* λειτουργεί ως εφεδρικό σύστημα (backup system). Η διαδικασία μετάφρασης περικλείει τα σημασιολογικά και πραγματιστικά συστατικά του συστήματος *MINDS*.

Ένα blackboard σύστημα ονομαζόμενο *SUS* σχεδιάστηκε από τον De Mori του Concordia Πανεπιστημίου και εφαρμόστηκε ως ένα αντικειμενοστραφές σύστημα. Οι πηγές γνώσεις του γεμίζουν τις σχισμές περιπτώσεων αντικειμένων με υποθέσεις και πληροφορίες. Οι περισσότεροι από τους γλωσσολογικούς ειδικούς της *SUS* απεικονίζουν τις γνώσεις τους ως σύνολα από κανόνες που λαμβάνουν υπόψιν σχετικούς περιορισμούς με την πηγή γνώσης.

Ο Bates, et al. (1993) περιγράφει το *Σύστημα Ομιλούμενης Γλώσσας* της BBN, ο Seneff (1992) περιγράφει το *TINA*, ο Young, et al. (1989) περιγράφει το *MINDS,* και ο Osterholz, et al. (1992) περιγράφει το *JANUS*. Ο Markowitz (1993e) παρέχει μια λεπτομερή μελέτη άλλων γλωσσολογικών και πολλαπλών πηγών γνώσης γραμματικών για αναγνώριση

ομιλίας. Επιπλέον πληροφορίες για το *SUS* blackboard σύστημα βρίσκονται στο De Mori (1983).

1.5 Ο ΕΝΤΟΠΙΣΜΟΣ ΛΕΞΕΩΝ (WORD SPOTTING)

Ο σκοπός του εντοπισμού λέξεων είναι να βρει και να ξεχωρίσει μια ή περισσότερες λέξεις κλειδιά ενσωματωμένες σε συνεχής ομιλία. Οι εφαρμογές με εντοπισμό λέξεων περιέχουν μοντέλα για κάθε μια από τις λέξεις κλειδιά της εφαρμογής καθώς επίσης μοντέλα για σιγή, θόρυβο, και ομιλία χωρίς λέξεις κλειδιά (τα ονομαζόμενα και μοντέλα σκουπιδιών – garbage or junk models). Αυτά τα μοντέλα πραγματοποιήθηκαν χρησιμοποιώντας ένα ευρύ φάσμα προσεγγίσεων, αλλά οι βασικές τεχνικές είναι κρυφά μοντέλα Markov (hidden Markov) και ίχνη (templates).

ΕΝΤΟΠΙΣΜΟΣ ΛΕΞΕΩΝ

Ενδιαφέρον στη σχεδίαση συστημάτων ικανών στην ανίχνευση ειδικών λέξεων σε συνεχή λόγο έχει εκδηλωθεί εδώ και 20 χρόνια τουλάχιστον. Έχει προσεγγιστεί χρησιμοποιώντας πρότυπα και HMM's. Το 1973, ο Bridle έκδοσε τις πρώτες προτάσεις χρησιμοποιώντας ίχνη και δυναμική χρονική στρέβλωση (dynamic time warping). Το σύστημά του υπολόγισε ένα ξεχωριστό αποτέλεσμα για κάθε λέξη κλειδί ίχνος ταιριασμένη με κάθε κομμάτι από τα δεδομένα εισόδου.Οι θέσεις στα δεδομένα που περιείχαν πιθανές λέξεις κλειδιά (ονομαζόμενες putative hits) συχνά επικαλύπτονταν και έπρεπε να κανονικοποιηθούν. Πιο σύγχρονες προτάσεις αντικατέστησαν τα επικαλυπτόμενα υποτιθέμενα hits με δύο τύπους ιχνών: λέξεις κλειδιά και συμπληρωτές (fillers). Οι συμπληρωτές αντιπροσώπευαν πλαίσια λόγου χωρίς λέξεις κλειδιά. Τα δεδομένα εισόδου θεωρήθηκαν ως μια ακολουθία πλαισίων από συμπληρωτές και λέξεις κλειδιά. Στη δεκαετία 1980, το HMM's άρχισε να αντικαθιστά πρότυπα, και οι συμπληρωτές εκτοπίστηκαν από πιο λεπτομερή ακουστική αναπαράσταση των λέξεων που δεν είναι κλειδιά, επικαλούμενα μοντέλα σκουπιδιών.

Μια από τις πρώτες εμπορικές εφαρμογές του εντοπισμού λέξεων ήταν το σύστημα *VRCP* (Voice Recognition Call Processing) της AT&T , το οποίο αναπτύχθηκε στα τέλη της δεκαετίας του 1980 για το χειρισμό τηλεφωνικών κλήσεων μακρινών αποστάσεων . Ο εντοπισμός λέξεων δεν εμφανίστηκε ως ένα χαρακτηριστικό εμπορικών προϊόντων για σχεδιαστές εφαρμογών μέχρις ότου το *Conversant* της AT&T το εισήγαγε στις αρχές της δεκαετίας του 1990. Από τα μέσα του 1994 είχε γίνει χαρακτηριστικό διάφορων άλλων εμπορικών προϊόντων που στόχευαν εφαρμογές βασιζόμενες στην τηλεφωνία.

Για περισσότερες ιστορικές πληροφορίες σχετικά με τον εντοπισμό λέξεων συμβουλευθείτε τον Bridle (1973) και τους Rose & Paul (1990). Ο Wilpon , (1985) παρέχει μια λεπτομερή συζήτηση σχετικά με τις ανακαλύψεις της AT&T σε ότι αφορά το *VRCP* σύστημα.

Εφόσον ο εντοπισμός λέξεων γενικά λειτουργεί σε περιβάλλοντα με θόρυβο με απρόβλεπτους χρήστες, οι εντοπιστές λέξεων μπορούν λανθασμένα να σημάνουν τον εντοπισμό μιας λέξεως κλειδί (ο επονομαζόμενος λάθος συναγερμός – false alarm) ή να αποτύχουν να ανιχνεύσουν μια λέξη κλειδί όταν εκστομίζεται (η επονομαζόμενη λάθος απόρριψη – false reject). Προσπάθειες γίνονται ώστε να αποφευχθούν όσο το δυνατό τέτοια λάθη. Μια τέτοια τεχνική είναι είναι η επιβεβαίωση οτι τα σύνολα των λέξεων κλειδιά δεν περιέχουν λέξεις που προκαλούν σύγχυση. Σύμφωνα με μια άλλη θεωρία , η SRI βρήκε οτι όσο περισσότερα μοντέλα λέξεων δίχως λέξεις κλειδιά ένα σύστημα περιέχει, τόσο καλύτερο εντοπισμό λέξεων κάνει. Σε παρόμοια συμπεράσματα κατέληξαν ερευνητές στα AT&T Bell εργαστήρια. Άλλες προσεγγίσεις εστιάζουν στο να διαχωρίσουν το σήμα της φωνής από το υπόβαθρο και το θόρυβο του καναλιού της φωνής ή ακόμα καλύτερα εντοπίζοντας την αρχή και το τέλος των λέξεων.

Ο εντοπισμός λέξεων έχει εσωτερική δομή. Αυτή η δομή είναι βασισμένη στα πρότυπα της σιγής, της εκστόμισης των λέξεων κλειδιών και επείσακτης (άσχετης) φωνής, τα οποία βρέθηκαν στην πρωταρχική έρευνα της AT&T.

Μια τυπική γραμματική εντοπισμού λέξεων έχει την ακόλουθη μορφή:

Εκστόμιση = σιγή – επείσακτη φωνή – λέξη κλειδί – επείσακτη φωνή – σιγή
(Utterance =silence – extraneous speech – keyword – extraneous speech – silence)

Αυτή η γραμματική είναι μέρος ενός συστήματος αναγνώρισης συνεχούς λόγου. Τα πρώτυπα μοντέλα για επείσακτη φωνή είχαν παρθεί από πραγματικά δείγματα φωνής από δεδομένα δοκιμών, όμως πιο πρόσφατα μοντέλα έχουν προκύψει από βάσεις δεδομένων φωνής μεγάλου λεξιλογίου βελτιωμένα μέσω της χρήσης λεπτομερών ακουστικών παραμέτρων.

1.5.1 Νευρωνικά δίκτυα για τον εντοπισμό λέξεων.

Οι περισσότεροι νευρωνικοί εντοπιστές λέξεων είναι υβριδικά συστήματα πού συνδυάζουν ένα δίκτυο με και δυναμική χρονική στρέβλωση (dynamic time warping). Δίκτυα που χρησιμοποιούνται με αυτό τον τρόπο περιλαμβάνουν το *MS-TDNN* της CMU, δίκτυα αναδρομής, και ένα εύρος από εμπροσθοτροφοδοτούμενα δίκτυα (feedforward networks).

Ένα υποσχόμενο δίκτυο είναι η γενικευμένης πιθανότητας πτώσης / ελάχιστη ταξινόμηση λάθους (GPD/MCE – Generalized Probability Descent / Minimum Classification Error) δίκτυο. Είναι μια παραλλαγή του εμπροσθοτροφοδοτούμενου learning vector quantization (LVQ) δικτύου. Προσπαθεί να ελαχιστοποιήσει τα λάθη ορίζοντας μια κλίση κατά μήκος της οποίας η ποιότητα της ιδιότητας μέλους σε μία ταξινόμηση αξιολογείται (γενικευμένη πιθανολογική πτώση). Το δίκτυο είναι έτσι φτιαγμένο ώστε να επιβάλλει ποινή στα λάθη αναγνώρισης που διαφέρουν κατά πολύ από την σωστή γραμματοσειρά (ταξινόμηση λάθους ως προς το ελάχιστο).

Περισσότερες πληροφορίες στα δίκτυα GPD/MCE για εντοπισμό λέξεων μπορούν να βρεθούν στο McDermott & Katagiri (1993). Ο Zeppenfeld (1993) αναπτύσει το θέμα της χρήσης των *MS - TDNN* για τον εντοπισμό λέξεων. Οι English & Boggess (1992) εξηγούν το όφελος από τη χρήση backpropagation training για δίκτυα που κάνουν εντοπισμό λέξεων.

1.5.2 Βρίσκοντας την ουσία ενός κειμένου (Gisting)

Το Gisting επεκτείνει τον εντοπισμό λέξεων σε δομημένους διαλόγους ανάμεσα σε 2 άτομα με το σύστημα αναγνώρισης να ενεργεί σαν ένας ανεμπόδιστος παρατηρητής. Είναι ένα προϊόν της καλυμμένης οικονομικά από την ARPA έρευνας να εξάγει πληροφορίες από μηνύματα και μεγάλες ποσότητες κειμένου. Το αντικείμενο του gisting είναι να εντοπίσει πληροφορίες που είναι γνωστό οτι υπάρχουν στην συνομιλία. Αντίθετα προς τους τυπικούς εντοπιστές λέξεων, οι ενδιαφέρον πληροφορίες καθορίζονται από τον τύπο και δεν περιορίζονται σε αντικείμενα ειδικού λεξιλογίου. Ένα σύστημα gisting θα μπορούσε για παράδειγμα να ανιχνεύσει το λόγο ενός ομιλητή ως μια γλώσσα που να υποδεικνύει ότι άτομο έχει εκστομίσει έναν προσωπικό κωδικό αναγνώρισης.

Ένα πρωτότυπο σύστημα για επικοινωνία ελέγχου εναέριας κυκλοφορίας (ATC –air traffic control) αναπτυγμένο από την BBN επεξηγεί τις δυνατότητες και τους περιορισμούς της τρέχουσας τεχνολογίας gisting. Το σύστημα αναγνωρίζει δεδομένα αναγνώρισης πτήσης και αποφασίζει εάν το αεροσκάφος προσγειώνεται ή απογειώνεται. Αποσπά αυτή την πληροφορία από ραδιομεταδόσεις ανάμεσα σε ελεγτές εναέριας κυκλοφορίας και πιλότες σε ρεαλιστικό χρόνο. Κάθε ελεγτής διατηρεί πολλαπλούς παρεμβαλλόμενους διαλόγους με πιλότους πάνω σε ένα θορυβώδες κανάλι επικοινωνίας. Η ομιλία είναι απότομη, ελεύθερα ρεούμενη και ιδιοματική. Επίσης γίνεται από άτομα τα οποία δεν συνεργάζονται ενεργά με το gisting σύστημα. Εφόσον η λειτουργία της ATC επικοινωνίας είναι να καθοδηγεί τους πιλότες, οι ενδοσυνεννοήσεις είναι σύντομες και η πληροφορία που απορρέει από το gisting σύστημα είναι καθαρά εμφανιζόμενη σε κάποια μορφή.

Για την αναγνώριση σχετικών πληροφοριών ενσωματωμένων στην επικοινωνία, το gisting απαιτεί τη χρήση και τεχνικής νοημοσύνης και τεχνικών στατιστικής. Βασίζεται σε αντικειμενοστραφή αναπαράσταση, μεθόδους συστημάτων βασισμένων στη γνώση και εργαλεία για περιορισμένη κατανόηση φυσικής γλώσσας. Πολύ καλός εντοπισμός λέξεων και καλές τεχνικές μοντέλου ομιλητή δίνουν τη δυνατότητα στο σύστημα να ξεχωρίσει τη φωνή του ελεγτή από τους πιλότες και να διαφοροποιήσει ανάμεσα στις φωνές των πιλοτών. Ο διάλογος πρέπει να είναι καλά ορισμένος και προσεκτικά κατανοημένος για να υποστηρίξει διαδικασία σε ρεαλιστικό χρόνο.

Χρηματοδοτήσεις από κρατικές υπηρεσίες ενθαρύνουν τη σχετική έρευνα με το gisting που θα το κάνει πιο ισχυρό και θα επεκτείνει την εφαρμογή του σε λιγότερα δομημένες συνομιλίες. Ένα ευκαταφρόνητο ποσοστό αυτής της χρηματοδότησης προέρχεται από τις κοινότητες υψηλής νοημοσύνης και στρατού. Στις Ηνωμένες Πολιτείες , για παράδειγμα, μια από τις πιο ενεργές πηγές χρηματοδότησης για αυτή τη δουλειά είναι το Εργαστήριο της Ρώμης των Εναέριων Δυνάμεων των Ηνωμένων Πολιτειών (Rome Laboratory of the United States Air Force). Οι δυναμικές εφαρμογές του gisting εκτείνονται στην απόσπαση πληροφορίας από μεγάλες ποσότητες ομιλίας ή συζήτησης που μπορούν να πάρουν μια επικεφαλίδα και να γίνει η περίληψή τους.

Περισσότερες πληροφορίες για το πρωτότυπο gisting σύστημα του BBN παρέχονται από τον Dennenberg (1993). Ο Sundheim (1989 και 1993) περιγράφει τεχνικές που έχουν αναπτυχθεί ώστε να εξάγουν πληροφορίες από κείμενο.

Ουσιαστικά όλα τα εμπορικά συστήματα αναγνώρισης περιέχουν μια ή περισσότερες από τις ακόλουθες γραμματικές:

- Γραμματική περιορισμένων καταστάσεων
- Ν-γραμμάτων μοντέλο
- Εντοπισμός λέξεων

1.6 ΓΡΑΜΜΑΤΙΚΗ ΠΕΡΙΟΡΙΣΜΕΝΩΝ ΚΑΤΑΣΤΑΣΕΩΝ ΣΕ ΕΜΠΟΡΙΚΑ ΣΥΣΤΗΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ

Γραμματικές περιορισμένων καταστάσεων είναι πιο αποτελεσματικές σε δομημένες εφαρμογές, όπως είσοδος δεδομένων και διαταγή φωνής και έλεγχος εξοπλισμού. Αυξάνουν την ταχύτητα και την ακρίβεια περιορίζοντας το μέγεθος του ενεργού λεξιλογίου. Όταν είναι καλά σχεδιασμένες, είναι γρήγορες, ικανές και ακριβείς.

Οι περισσότερες γραμματικές περιορισμένων καταστάσεων είναι κατασκευασμένες από ένα σχεδιαστή εφαρμογών έτσι ώστε να ικανοποιούν τις απαιτήσεις μιας συγκεκριμένης εργασίας. Γενικά, οι γραμματικές περιορισμένων καταστάσεων είναι εύκολες στην κατανόηση και στην κατασκευή. Τα πιο δύσκολα σημεία της σχεδίασης μιας γραμματικής είναι η εντόπιση και ο καθορισμός των γλωσσολογικών και δομικών αναγκών της εργασίας. Η πρόσθετη εμφάνιση της αυτόματης αφαίρεσης λεξιλογίου, που είναι τώρα διαθέσιμη σε κάποια προϊόντα με GUI (graphical user interface) εφαρμογή, κάνει τη διαδικασία της σχεδίασης ακόμα πιο εύκολη.

Σημαντικό για εφαρμογές που περιέχουν γραμματικές περιορισμένων καταστάσεων είναι η επιβεβαίωση ότι οι άνθρωποι που χρησιμοποιούν το σύστημα μιλούν μόνο τις επιτρεπόμενες λέξεις και ακολουθίες λέξεων. Εάν οι χρήστες είναι μία μικρή και καλά καθορισμένη ομάδα ανθρώπων που θα απασχολούν το σύστημα τακτικά, μπορούν να βοηθηθούν με ένα εκπαιδευτικό πρόγραμμα. Επίσης η εφαρμογή θα μπορούσε να περιέχει και ένα σύστημα βοήθειας. Αυτές οι τεχνικές είναι λιγότερο χρήσιμες σε εφαρμογές που προβλέπουν μεγάλο αριθμό χρηστών της μίας φοράς. Σε τέτοιες περιπτώσεις, μπορεί να χρησιμοποιηθούν γραμματικές περιορισμένων καταστάσεων για να δημιουργήσουν δομές λόγου που φαίνονται φυσικές στους χρήστες του συστήματος. Αυτό απαιτεί πλήρη κατανόηση των τρόπων με τους οποίους οι χρήστες εκτελούν την εργασία , της γλώσσας που πιθανώς χρησιμοποιούν, της επίδρασης των διαφορετικων συμπεριφορών των χρηστών και της καλής διόρθωσης λαθών.

1.7 Ν-ΓΡΑΜΜΑΤΩΝ ΜΟΝΤΕΛΑ ΣΕ ΕΜΠΟΡΙΚΑ ΣΥΣΤΗΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ

Τα περισσότερα μεγάλα συστήματα λεξιλογίου χρησιμοποιούν δύο και/ή τριών γραμάτων μοντέλα γλώσσας. Αυτά τα μοντέλα αυξάνουν την ταχύτητα και την ακρίβεια ταξινομώντας τα στοιχεία στο ενεργές λεξιλόγιο. Παρέχουν φυσικότητα επιτρέποντας στους χρήστες μεγάλη ευλιγησία κινήσεων.

Είναι δύσκολη η κατασκευή των καλών Ν-γραμμάτων μοντέλων γιατί απαιτούν τεράστιες ποσότητες γλωσσικών δεδομένων και επιτηδευμένη στατιστική ανάλυση. Εξαιτίας της δυσκολίας στην απόκτηση καλών πηγών δεδομένων, τα περισσότερα Ν- γραμμάτων

μοντέλα αναπτύσσονται από πωλητές της τεχνολογίας. Κάποιοι πωλητές παρέχουν εργαλεία σε σχεδιαστές εφαρμογών για να τροποποιήσουν το γλωσσικό μοντέλο και κάποιοι έχουν αυτοματοποιήσει τη διαδικασία προσαρμογής του μοντέλου. Τέτοια εργαλεία είναι σημαντικά γιατί ένα Ν- γραμμάτων μοντέλο γλώσσας απεικονίζει τα πρότυπα ακολουθιών λέξεων που βρίσκονται στα δεδομένα και χρησιμοποιούνται για να το καταρτίσουν. Τα πρότυπα γλώσσας μιας βιομηχανίας μπορεί να είναι αρκετά διαφορετικά από εκείνα άλλων βιομηχανιών. Μοντέλα κατασκευασμένα από εκδόσεις του *Wall Street Journal,* για παράδειγμα, μπορεί να μην δουλεύουν καλά για τη βιομηχανία μεταφοράς, ή μπορεί ακόμα να είναι ακατάλληλα για διαφορετικούς οργανισμούς εντός της οικονομικής βιομηχανίας. Συνεπώς, τα μοντέλα γλώσσας ενός προϊόντος υπαγόρευσης πρέπει να δοκιμαστούν σε ένα αντιπροσωπευτικό δείγμα των αναμενόμενων μορφών της υπαγόρευσης.

Τα κείμενα ενός οργανισμού ή ενός ατόμου αποτελούν το καλύτερο υλικό για την ανάπτυξη ενός πιθανολογικού μοντέλου γλώσσας. Επιπρόσθετη μοντελοποίηση υπολέξεων μπορεί να χρησιμοποιηθεί για να συμπληρώσει το μόνιμο λεξιλόγιο μιας εφαρμογής έτσι ώστε να ικανοποιούνται οι ανάγκες μιας ιδιαίτερης οργάνωσης. Μερικά προϊόντα προσαρμόζουν το μοντέλο γλώσσας για να εκφράζει πρότυπα ακολουθιών λέξεων που βρίσκονται στα κείμενα. Αυτά τα εργαλεία μοντελοποίησης γλώσσας διαφέρουν στην ικανότητά τους να ενσωματώνουν στατιστικά για κάθε νέα λέξη που βρίσκεται στα κείμενα.

Ορισμένα προϊόντα περιέχουν μοντέλα γλώσσας που προσαρμόζονται κατά τη χρήση έτσι ώστε να συνταιριάξουν τα πρότυπα ακολουθιών λέξεων του ανθρώπου που χρησιμοποιεί το σύστημα. Αυτό μπορεί ή όχι να περιλαμβάνει την ένταξη καινούργιου λεξιλογίου στα μοντέλα γλώσσας. Άλλοι πωλητές βασίζονται σε γενικές εκτιμήσεις συχνοτήτων άγνωστων λέξεων αλλά δεν αλλάζουν τα υπάρχοντα μοντέλα γλώσσας τους , όταν καινούργιες λέξεις προστίθενται.

1.8 ΕΝΤΟΠΙΣΜΟΣ ΛΕΞΕΩΝ ΣΕ ΕΜΠΟΡΙΚΑ ΣΥΣΤΗΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ

Ο εντοπισμός λέξεων τυπικά χρησιμοποιείται για μικρές, υψηλά δομημένες εφαρμογές που εμπλέκουν μεγάλο και διαφορετικό πληθυσμό μιας φοράς χρηστών. Ο αριθμός των δυναμικών χρηστών, η έλλειψη δέσμευσης από την πλευρά των χρηστών σε ένα σύστημα εντοπισμού λέξεων και η προσωρινή φύση της επαφής τους με αυτό ενέπνευσε τον David Pallett του Εθνικού Ινστιτούτου Προτύπων και Δοκιμών (National Institute of Standards and Testing - NIST) να αναφερθεί σε τέτοιες αλληλεπιδράσεις ως αμέθοδη - αδιάκριτη ομιλία (promiscuous speech).

Η μη παρεμβολή / ενόχληση του εντοπισμού λέξεων επιτρέπει σε ομιλητές που δεν είναι οικείοι με συστήματα αναγνώρισης ομιλίας να επικοινωνούν κανονικά. Για αυτό το λόγο ο εντοπισμός λέξεων γίνεται αυξανόμενα δημοφιλής σε τηλεφωνικές εφαρμογές. Το αθέατο του εντοπισμού των λέξεων είναι ένα παραπροϊόν προσεκτικής σχεδίασης εφαρμογών και εκτενών δοκιμών. Η επιλογή και η κωδικοποίηση των λέξεων κλειδιών , η σχεδίαση οδηγιών και υποβολών (prompts) ομιλίας, και η δομή της αλληλεπίδρασης , όλα πρέπει να ταιριάζουν στην πιθανή συμπεριφορά των χρηστών. Όλοι αυτοί οι παράγοντες συνεισφέρουν στη φυσικότητα, αποτελεσματικότητα και ακρίβεια της εφαρμογής.

1.9 ΧΡΗΣΙΜΟΠΟΙΩΝΤΑΣ ΓΡΑΜΜΑΤΙΚΗ

Η σχεδίαση μιας επιτυχημένης γραμματικής εξαρτάται από την ανάλυση της εργασίας, το χρόνο απόκρισης, τις εφαρμογές που αντιλαμβάνουν την ομιλία και από την αναμονή του απρόσμενου. Όλα αυτά αναλύονται στις ακόλουθες παραγράφους.

1.9.1 Ανάλυση της εργασίας

Το πιο προκλητικό κομμάτι στη σχεδίαση μιας γραμματικής είναι η ανάλυση μιας εργασίας. Απαιτεί προσεκτική κατανόηση των εξής:

- της οργάνωσης της εργασίας και των υποεργασιών της
- του λεξιλογίου που απαιτείται για την εργασία
- των γλωσσικών προτύπων της εργασίας
- των προτύπων ελέγχου εφαρμογών
- των αναμενόμενων διαφορετικών συμπεριφορών των ανθρώπων που εκτελούν την εργασία

Τα πρότυπα ελέγχου εφαρμοργών περιλαμβάνουν φαινομενικά φυσικές μεθόδους που ενεργοποιούν ή όχι την αναγνώριση φωνής , διορθώνουν λάθη, γυρίζουν σε προηγούμενη κατάσταση της εφαρμογής, δέχονται βοήθεια και δίνουν προφορικές οδηγίες και βοήθεια στους χρήστες.

Για τα ελεύθερης μορφής συστήματα υπαγόρευσης, το πιο σημαντικό είναι η εξασφάλιση ότι οι σημαντικές λέξεις βρίσκονται στο μόνιμο λεξιλόγιο. Για όλους τους άλλους τύπους εφαρμογών, συμπεριλαμβανομένης και της δημιουργίας δομημένων αναφορών, η ανάλυση της εργασίας και του χρήστη είναι ένα κεντρικό χαρακτηριστικό της αποδεκτικότητας της εφαρμογής και της φυσικότητας του προσαρμοστικού ομιλίας. Αυτόματη αφαίρεση λεξιλογίου είναι χρήσιμη, εάν ο λόγος προστίθεται σε ένα υπολογιστικοποιημένο σύστημα. Αυτά τα εργαλεία αφαιρούν λεξιλόγιο και δομή από εφαρμογές υπολογιστών. Σαν αποτέλεσμα, είναι χρήσιμα για την κατανόηση της εκτέλεσης της εργασίας. Οι αποφάσεις αυτόματης αφαίρεσης λεξιλογίου δεν θα παραλληλίζουν απαραίτητα και τη φυσική ανθρώπινη συμπεριφορά και συνεπώς, θα πρέπει να εκτιμηθούν όσο αφορά στην αναμενόμενη συμπεριφορά των χρηστών , όταν εκτελούν την εργασία. Τα menu μιας εφαρμογής των Windows, για παράδειγμα, είναι πιθανό να υποστηρίζουν το πρότυπο “File open” και “File close” επειδή αυτές οι ακολουθίες λέξεων αντανακλούν τις απαιτήσεις εφαρμογών Windows βασισμένες σε menu. Σε αντίθεση, ένας άνθρωπος είναι πιο πιθανό να πει “Open the file” και “Close the file“. Όταν χρησιμοποιείται με αυτό τον τρόπο, η ομιλία είναι πραγματικά ένα πληκτρολόγιο ή υποκατάστατο ποντικιού, αλλά αποτυγχαίνει το στόχο της να είναι μια πιο φυσική , βασισμένη στον άνθρωπο συσκευή εισόδου δεδομένων.

Ένας τρόπος να μεγιστοποιηθεί η φυσικότητα είναι η εμπλοκή των χρηστών του συστήματος στη σχεδίαση και στις δοκιμές του συστήματος. Μία άλλη τεχνική είναι η δοκιμή συγκεκριμένων χαρακτηριστικών της εφαρμογής ως οδηγοί ή πρωτότυπα. Μία τρίτη προσέγγιση είναι η χρήση επαναληπτικής σχεδίασης και δοκιμών.

1.9.2 Χρόνος απόκρισης

Ένας λόγος αποδοχής ενός χρήστη είναι ο λογικός χρόνος απόκρισης. Οι χρήστες επιβάλλουν όρια στο χρόνο που είναι πρόθυμοι να περιμένουν ένα σύστημα αναγνώρισης να εκτελέσει την εργασία του. Αυτό το όριο μπορεί να είναι εξαιρετικά μικρό και είναι απίθανο να επεκταθεί με το μέγεθος του λεξιλογίου μιας εφαρμογής.

Η σχεδίαση μιας γραμματικής περιορισμένων καταστάσεων θα επηρεάσει τον χρόνο απόκρισης. Ένα συστατικό είναι το μέγεθος του ενεργού λεξιλογίου. Η οργάνωση της γραμματικής και η αλληλεπίδραση ανάμεσα στις δομές της γραμματικής περιορισμένων καταστάσεων μπορούν επίσης να έχουν επίδραση στην ταχύτητα του σύστηματος. Απλότερες δομές που περιέχουν λίγη αναδρομή τρέχουν γρηγορότερα από μεγαλύτερες και πολύπλοκες δομές.

1.9.3. Εφαρμογές αντίληψης ομιλίας (Speech aware applications)

Τα περισσότερα συστήματα αναγνώρισης είναι σχεδιασμένα έτσι ώστε να λειτουργούν παράλληλα με μια εφαρμογή. Οι δομές τους μιμούνται την οργάνωση της εργασίας, αλλά είναι περιορισμένες στην εργασία. Σαν αποτέλεσμα, μπορούν να βγούν εκτός συγχρονισμού (out-of-sync) με την εφαρμογή.

Η καλύτερη λύση σ' αυτό το πρόβλημα είναι η σχεδίαση μίας speech-aware (γνωστή και ως speech-enabled) εφαρμογής. Στις εφαρμογές αντίληψης ομιλίας, το πρόγραμμα της εφαρμογής αναγνωρίζει το λεξιλόγιο και τη δομή που απαιτούνται για την εκτέλεση μιας συγκεκριμένης εργασίας και μεταδίδει αυτή την πληροφορία στο σύστημα ομιλίας. Προγράμματα εφαρμογών αναπτυγμένα εντός ενός οργανισμού ή φτιαγμένα λαμβάνοντας υπόψην και τη σχεδίαση των μετωπικών αναγνώρισης ομιλίας (speech recognition front-end) είναι από τα πιο εύκολα για να κατασκευάσουν συστήματα αντίληψης ομιλίας. Τα περισσότερα εμπορικά προγράμματα δεν επιτρέπουν ακόμα αυτόν τον τύπο σύνδεσης για ομιλία.

1.9.4. Αναμένοντας το απρόσμενο

Άσχετα από το πόσο καλά μια εφαρμογή είναι σχεδιασμένη, και πόσο πολύ είναι δοκιμασμένη, οι καλοί σχεδιαστές αναμένουν το απρόσμενο. Μερικά απροσδόκητα γεγονότα μπορεί να προκύψουν από πλατφόρμας, καναλιού ή επικοινωνίας εξοπλισμού πλοβλήματα, αλλά η πλειοψηφία θα δημιουργηθεί από χρήστες. Αυτό είναι ιδιαίτερα αλήθεια για εφαρμογές σχεδιασμένες για χρήστες μιας φοράς. Οι παράμετροι του συστήματος , όπως το κατώφλι λήξης χρόνου ή έντασης εισόδου ίσως χρειάζονται προσαρμογή. Οι υποβολές (prompts) ίσως χρειαστεί να αλλάξουν και μερικά χαρακτηριστικά μπορεί να χρειαστεί να ξαναμπούν σε μια σειρά.

Ακόμα και επιτηδευμένοι χρήστες μπορεί να μην προσέξουν, να μπερδευτούν ή να ξεχαστούν. Εάν το κόστος ενός λάθους αναγνώρισης είναι υψηλό, το σύστημα πρέπει να σχεδιαστεί με συχνές επιβεβαιώσεις αναγνωρισμένης είσοδου. Εάν οι χρήστες μπορεί να

είναι αφηρημένοι, ο σχεδιαστής πρέπει να προσθέσει επαναληπτικές υπενθυμήσεις και/ή χαρακτηριστικά που να υπενθυμίζουν στους χρήστες που βρίσκονται στην εφαρμογή.

Η καλύτερη προστασία από το απρόσμενο είναι η ανάπτυξη ενός μικρού πρωτότυπου ή πιλότου για δοκιμές. Είναι πολύ ευκολότερη η διόρθωση προβλημάτων σε ένα πρωτότυπο σύστημα από την εκ νέου σχεδίαση μιας μεγάλης εφαρμογής.

> *Εάν κάνεις κάτι τόσο εύκολο όσο αυτό θα έχεις καλύτερες εφαρμογές και θα μπορείς ακόμα να αποφασίσεις εάν η αναγνώριση φωνής θα δουλέψει στην εφαρμογή ή όχι. Η AT&T, παρόλη την εμπειρία μας, συνήθως κάνει δοκιμές αυτού του τύπου (Judith Tschirgi, Διευθυντής στο τμήμα Τεχνολογία Υπηρεσιών και Φωνής, AT&T Συστήματα Δικτύου, προσωπική επικοινωνία, 1994).*

Μία άλλη χρήσιμη τεχνική ονομάζεται η προσέγγιση του Μάγου του Οζ (*the Wizard of Oz approach*). Αυτή περιέχει προσωμοίωση της εφαρμογής αναγνώρισης. Οι χρήστες του συστήματος πιστεύουν ότι είναι σε επικοινωνία με ένα σύστημα αναγνώρισης, αλλά στην πραγματικότητα αλληλεπιδρούν με έναν άνθρωπο. Οι δοκιμές στην προσέγγιση του Μάγου του Οζ είναι καταπληκτικές στην εξακρίβωση μεταβλητών συμπεριφοράς και σημείων σύγχυσης.

1.10 ΚΑΤΑΝΟΗΣΗ ΟΜΙΛΟΥΜΕΝΗΣ ΓΛΩΣΣΑΣ

Η κατανόηση της ομιλουμένης είναι μια περιοχή ενεργής έρευνας που είναι ισχυρά υποστηριζόμενη από επιδοτήσεις του ARPA (Αμερικάνικο Συμβούλιο Ερευνών) και κυβερνητικούς παράγοντες άλλων κρατών. Η κατανόηση της γλώσσας είναι μια πολύπλοκη και πολύπλευρη εργασία, και είναι απίθανο ότι εμπορικά συστήματα κατανόησης της ομιλουμένης θα προκύψουν στο εγγύς μέλλον. Στοιχεία κατανόησης μιας γλώσσας έχουν όμως ήδη αρχίσει να εμφανίζονται σε εμπορικά στραμμένα πρωτότυπα.

Το *Speak2Directions*, για παράδειγμα, είναι ένα αρχέτυπο σχεδιασμένο από την Pure Speech για την υψηλά δομημένη και καλά ορισμένη επικοινωνία για να λαμβάνει κατευθύνσεις μέσω τηλεφώνου. Χρησιμοποιεί ένα σημασιολογικό πλαίσιο με δύο αντικείμενα: ένα αντικείμενο ομιλίας (discourse object) και ένα αντικείμενο ιστορίας (history object). Το αντικείμενο ομιλίας αποθηκεύει το στόχο της αλληλεπίδρασης μαζί με σημαντικές πληροφορίες σχετικά με το στόχο στις σχισμές του αντικειμένου, object's slots (επίσης ονομαζόμενες και ιδιότητες - attributes). Στο *Speak2Directions* πρωτότυπο ο στόχος είναι να ληφθούν οι κατευθύνσεις, επιβάλλοντας σχισμές για να αναπαραστήσουν την τρέχουσα θέση και τον προορισμό του προσώπου. Το αντικείμενο της ιστορίας παρακολουθεί τις συγκεντρωμένες πληροφορίες στην πορεία του διαλόγου. Χρησιμοποιείται για να ταυτοποιήσει την αναφορά αντωνυμιών και λέξεων όπως "εδώ". Εάν ένας καλών επιθυμεί να πάει στο Cambridge Marriott, το *Speak2Directions* θα συμβουλευτεί το αντικείμενο ομιλίας του για να αποφασίσει εάν του έχει ειπωθεί η τρέχουσα θέση αυτού που καλεί. Εάν όχι, θα ζητήσει αυτή την πληροφορία. Αργότερα στην συνομιλία, το αντικείμενο ιστορίας μπορεί να κληθεί να μεταφράσει τη λέξη "εδώ" ως η τρέχουσα θέση του καλούντος.

1.11 ΣΥΣΤΗΜΑΤΑ ΧΩΡΙΣ ΓΡΑΜΜΑΤΙΚΗ

Μερικοί πωλητές υποστηρίζουν ότι τα προϊόντα τους δεν έχουν γραμματική. Γενικά, αυτό σημαίνει ότι δεν υπάρχει γραμματική περιορισμένων καταστάσεων. Τα περισσότερα προϊόντα που περιγράφονται με αυτό τον τρόπο χρησιμοποιούν ένα στατιστικό μοντέλο

γλώσσας , συνήθως ένα δύο ή ένα τριών γραμμάτων. Εάν ο πωλητής επιμένει ότι δεν υπάρχει κανενός είδους γραμματική στο προϊόν, θα πρέπει να είναι έτοιμοι να εξηγήσουν (και να παρουσιάσουν) πώς το προϊόν χειρίζεται υψηλή περιπλοκή.

2. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΟΜΙΛΗΤΗ

ΕΙΣΑΓΩΓΗ

Ένα σύστημα αναγνώρισης φωνής πρέπει να καταλαβαίνει την ομιλία του οποιουδήποτε που χρειάζεται να το χρησιμοποιήσει. Αυτό μπορεί να περιλαμβάνει μια μικρή, καλά ορισμένη ομάδα τακτικών χρηστών του συστήματος ή μια μεγάλη ομάδα ποικίλου πληθυσμού μιας φοράς χρηστών. Δίνοντας αυτήν την απαίτηση, η εστίαση της μοντελοποίησης ενός ομιλητή γίνεται στην επινόηση τρόπων παρουσίασης της παραλλακτικότητας του ομιλητή που δίνουν τη δυνατότητα στα συστήματα αναγνώρισης να λειτουργούν καλά με την ομιλία όλων των δυνατών χρηστών του συστήματος. Οι διαφορες προσεγγίσεις που έχουν δημιουργηθεί για την κατάκτηση αυτού του στόχου μπορούν να ομαδοποηθούν ως εξής:

- Μοντελοποίηση εξαρτώμενη από τον ομιλητή
- Μοντελοποίηση πολλαπλών ομιλητών
- Μοντελοποίηση ανεξάρτητη από τον ομιλητή
- Προσαρμογή του ομιλητή

Αυτό το κεφάλαιο εξετάζει τις τεχνολογίες που απαιτούνται για τη δημιουργία διαφόρων μορφών μοντελοποίησης ομιλητή και τα αποτελέσματα εφαρμογών συσχετισμένα σε κάθε προσέγγιση μοντελοποίησης ομιλητή. Το επίκεντρο της τεχνολογίας ξεκινάει με ορισμούς καθεμίας από τις προσεγγίσεις στην μοντελοποίηση του ομιλητή και συζητήσεις γύρω από τεχνολογικές προκλήσεις σε σχέση με αυτές. Ειδική προσοχή δίνεται στις δύο κυρίαρχες μορφές μοντελοποίησης : στην ανεξάρτητη από τον ομιλητή και σ΄αυτή της προσαρμογής του ομιλητή.

Το επίκεντρο της εφαρμογής χαρακτηρίζει τις τέσσερις προσεγγίσεις της μοντελοποίησης του ομιλητή ως σημεία κατά μήκος μιας συνεχούς σειράς μοντελοποίησης ομιλητή. Περαιτέρω συζήτηση μεμονωμένων προσεγγίσεων μοντελοποίησης ομιλητή εστιάζει στα δυνατά και στα αδύνατα σημεία κάθε προσέγγισης, την ακρίβειά της, και την ευκολία με την οποία νέες λέξεις προστίθενται σε εφαρμογές. Τέλος, κάποιες γενικές εκβάσεις από την μοντελοποίηση ομιλητή, όπως η επίδραση του άγχους στην ομιλία αναπτύσσονται.

Παρακάτω περιγράφονται οι κύριες τεχνολογίες που χρησιμοποιούνται στην αναπαράσταση των λέξεων:

- Ίχνη
- Κρυφά Markov μοντέλα
- Νευρωνικά δίκτυα

Κάθε τεχνολογία μπορεί να χρησιμοποιηθεί για να δημιουργήσει μια ψηφοποιημένη αναπαράσταση για κάθε αντικείμενο του λεξιλογίου σε μία εφαρμογή ή ένα σύστημα. Τα πρότυπα που αποθηκεύονται ως κομμάτι μίας εφαρμογής καλούνται πρότυπα αναφοράς. Η αναγνώριση επιτυγχάνεται όταν ο εισερχόμενος λόγος συγκρίνεται με αυτά τα αποθηκευμένα πρότυπα αναφοράς και ένα καλό ταίριασμα εντοπίζεται. Η ακουστική πληροφορία κωδικοποιημένη στα εσωτερικά πρώτυπα αναφοράς που έχουν δημιουργηθεί από αυτές τις τεχνολογίες διαφέρει ανάλογα με το ποιά από τις τέσσερις προσεγγίσεις μοντελοποίησης ομιλητή χρησιμοποιείται.

Για περαιτέρω πληροφορίες σχετικά με μοντελοποίηση ομιλητή, συμβουλευτείτε τους Markowitz (1990 και 1993a). Rabiner & Juang (1993), και Huang &Lee (1993).

2.1 ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΕΞΑΡΤΩΜΕΝΗ ΑΠΟ ΤΟΝ ΟΜΙΛΗΤΗ

Ένα σύστημα αναγνώρισης που χρησιμοποιεί μοντελοποίηση εξαρτώμενη από τον ομιλητή δημιουργεί ένα ειδικού ομιλητή μοντέλο για κάθε ένα από τα άτομα που θα χρησιμοποιεί το σύστημα. Ένα μοντέλο ομιλητή περιέχει τις ιδιότητες της ομιλίας ενός μεμονωμένου εξουσιοδοτημένου ομιλητή. Τα πιο πολλά εμπορικά συστήματα αποθηκεύουν το μοντέλο του κάθε ομιλητή σε ένα ξεχωριστό αρχείο. Όταν ένας εξουσιοδοτημένος χρήστης επιθυμεί πρόσβαση στο σύστημα αναγνώρισης, αυτός/αυτή εισέρχεται σε έναν προσωπικό προσδιοριστή ταυτότητας που προκαλεί στο σύστημα να φορτώσει αυτού του ατόμου το αρχείο ομιλητή.

Η δημιουργία ενός οποιουδήποτε μοντέλου εξαρτώμενου από τον ομιλητή εμπεριέχει τρία βήματα:

- Συλλογή δεδομένων
- Υπολογισμός
- Κατασκευή μοντέλου

Η διαδικασία της συλλογής δεδομένων λέγεται παρασκευή ή καταγραφή. Συνίσταται στην απόσπαση ενός ή περισσότερων δειγμάτων ομιλίας (που ονομάζονται tokens) από κάθε λέξη στην εφαρμογή από τον ομιλητή. Τα βήματα του επακόλουθου υπολογισμού και της κατασκευής του μοντέλου εξαρτώνται από το εάν το μοντέλο είναι δημιουργημένο χρησιμοποιώντας ίχνη (templates) ή HMM's.

Επειδή η εξαρτώμενη από τον ομιλητή αναπαράσταση προσπαθεί να συλλάβει τα ακουστικά πρότυπα ενός μόνο ατόμου, μπορεί να παράγει καταπληκτικά μοντέλα.

> *Εξαιτίας των παραλλαγών ανάμεσα σε ομιλητές, καλά καταρτισμένα συστήματα αναγνώρισης ομιλίας εξαρτώμενα από τον ομιλητή λειτουργούν καλά και ως συστήματα ανεξάρτητα από τον ομιλητή με ποσοστό δύο στα τρία (Xue-dong Huang & Kai_Fu Lee, Carnegie Mellon University, " On speaker-independent, speaker-dependent, and speaker-adaptive speech recognition", 1993, p.150).*

2.1.1 Μοντέλα ίχνους

Ο υπολογισμός για τη δημιουργία ίχνους αποτελείται από τον υπολογισμό κανόνων για ακουστικές παραμέτρους για τα δείγματα ομιλίας (tokens). Ένα ή περισσότερα ίχνη κατασκευάζονται χρησιμοποιώντας τα αποτελέσματα της διαδικασίας υπολογισμού.

Τα πρώτα συστήματα ίχνους δημιούργησαν ένα ίχνος για κάθε δείγμα ομιλίας. Κάθε ίχνος έγινε ένα ξεχωριστό αντιπροσωπευτικό πρότυπο για το αντικείμενο του λεξιλογίου του. Εξαιτίας της κακής του χρησιμοποίησης σε μνήμη και του υψηλού επιπέδου υπολογισμών απαιτούμενων για τη σύγκριση της εισερχόμενης ομιλίας με κάθε ίχνος, εφαρμογές σχεδιασμένες χρησιμοποιώντας αυτή τη μεθοδολογία κατασκευής ίχνους χαρακτηρίζονταν από μικρά λεξιλόγια. Επιπλέον, αυτή η μεθοδολογία δεν μπορούσε ούτε τις παραλλαγές του ίδιου του ομιλητή να χειριστεί αποτελεσματικά ούτε να ελαχιστοποιήσει τα αποτελέσματα λαθών που γίνονταν κατά τη διάρκεια της καταγραφής.

Βελτιώσεις στην τεχνολογία ίχνους έκαναν δυνατή την κατασκευή ιχνών μοντέλων αναφοράς από δύο ή περισσότερα δείγματα ομιλίας. Κάθε πλαίσιο του αντιπροσωπευτικού ίχνους αναπαριστά το στατιστικό μέσο όρο των ακουστικών δεδομένων που περιέχονται στα δείγματα που χρησιμοποιούνται για να παράγουν το πλαίσιο. Αυτή η προσέγγιση που λέγεται *εύρωστη κατάρτιση* (*robust training*), αποφεύγει πολλά από τα προβλήματα που χαρακτήριζαν νωρίτερα συστήματα ιχνών.

2.1.2 Κρυφά Markov μοντέλα

Ο υπολογισμός για κρυφά Markov μοντέλα (HMM's) χρησιμοποιεί πιο εξειδικευμένα στατιστικά από τους μέσους υπολογισμούς που χρησιμοποιούνται για παραγωγή ίχνους. Τα στατιστικά είναι σχεδιασμένα έτσι ώστε να συλλαμβάνουν τα πρότυπα μεταβλητότητας του ομιλητή τόσο καλά όσο κανόνες. Τα αποτελέσματα των υπολογισμών είναι ενσωματωμένα στις καταστάσεις και μεταβάσεις του HMM μοντέλου λέξεων. Η πιο κοινώς χρησιμοποιούμενη τεχνική δημιουργίας HMM μοντέλου, ο Baum-Welch αλγόριθμος, ενημερώνει και ξαναυπολογίζει τις ακουστικές παραμέτρους ενός HMM βασιζόμενος στα δείγματα που του παρέχονται. Αυτή η διαδικασία ονομάζεται *επανεκτίμηση (reestimation)*.

Καλές τεχνικές αναλύσεις μοντελοποίησης εξαρτώμενης από τον ομιλητή παρέχονται από τους Rabiner & Juang (1993) και από τον Moore (1992).

2.2 ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΠΟΛΛΑΠΛΩΝ ΟΜΙΛΗΤΩΝ

Η μοντελοποίηση πολλαπλών ομιλητών είναι μια επέκταση της εξαρτώμενης μοντελοποίησης από τον ομιλητή, και συστήματα αναγνώρισης εξαρτώμενα από τον ομιλητή γενικώς χρησιμοποιούνται για να τα δημιουργήσουν. Τα μοντέλα είναι σχεδιασμένα ώστε να αναπαραστούν τις ιδιότητες ομιλίας μίας ομάδας ατόμων. Η δημιουργία ακριβών και υψηλής ποιότητας μοντέλα πολλαπών χρηστών απαιτεί έναν γνωστό και καλά ορισμένο πληθυσμό εξουσιοδοτημένων ομιλητών. Όλοι οι χρήστες προμηθεύουν ένα ή περισσότερα δείγματα από κάθε αντικείμενο του λεξιλογίου που πρόκειται να χρησιμοποιήσουν.

Επειδή η μοντελοποίηση πολλαπλών ομιλητών προσπαθεί να αναπαραστήσει τα πρότυπα ομιλίας περισσοτέρων από έναν ομιλητή χρησιμοποιώντας τεχνολογία εξαρτώμενη από τον ομιλητή, είναι λιγότερο ακριβής από την εξαρτώμενη μοντελοποίηση από τον ομιλητή. Η ακρίβειά της μειώνεται όσο τα μέλη της ομάδας τών ομιλητών διαφοροποιούνται.

2.3 ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΑΝΕΞΑΡΤΗΤΗ ΤΟΥ ΟΜΙΛΗΤΗ

Εφαρμογές με μοντέλα ανεξάρτητα του ομιλητή σχεδιάζονται για να χρησιμοποιηθούν από ανθρώπους που δεν έχουν εγγραφεί ως μέλη. Τα μοντέλα αναφοράς για πολλές από αυτές τις εφαρμογές πρέπει να αναγνωρίζουν μεγάλους και ετερογενείς πληθυσμούς ομιλητών. Δοθέντος το εύρος μεμονωμένων διαφορών, διαλέκτων, ξενικών προφορών και δυσχερειών στην ομιλία από τους ομιλητές, αυτό είναι ένας δύσκολος στόχος για να επιτευχθεί.

Τα μοντέλα αναφοράς σχεδιασμένα για αναγνώριση ανεξάρτητη του ομιλητή απεικονίζουν τις αναμενόμενες ακουστικές παραμέτρους ολόκληρου του πληθυσμού των ομιλητών. Όλοι οι ομιλητές χρησιμοποιούν το ίδιο σύνολο μοντέλων αναφοράς. Συνεπώς, τα μοντέλα αναφοράς σχεδιασμένα για εφαρμογές ανεξάρτητες του ομιλητή είναι πιο πολύπλοκα και δύσκολα να κατασκευαστούν από αυτά που δημιουργούνται για αναγνώριση εξαρτώμενη από τον ομιλητή.

> *Με την εξαρτώμενη από τον ομιλητή αναγνώριση δεν χρειάζεται να ανησυχείς με την μεταβλητότητα όπως κάνεις με την ανεξάρτητη από τον ομιλητή αναγνώριση. Η ομιλία ενός μόνου προσώπου διαφέρει από μέρα σε μέρα και από το πρωϊ και τη νύχτα, αλλά το εύρος της μεταβλητότητας είναι πολύ λιγότερο από αυτό που απαιτείται από ένα μοντέλο ανεξάρτητο του ομιλητή για ολόκληρη τη χώρα (Jeff Hill, Αντιπρόεδρος ανάπτυξης νέων προϊόντων, Voice Processing Corp, personal communication, 1993) .*

Για να ελαχιστοποιήσουν τα λάθη, τα ανεξάρτητα του ομιλητή συστήματα έχουν κατά παράδοση μικρά λεξιλόγια και έχουν αποφύγει αντικείμενα λεξιλογίου που προκαλούν σύγχυση. Η εμφάνιση γρηγορότερων και δυναμικότερων PC συστημάτων συνδυαζόμενα με όχι ακριβά τσιπς μνήμης έχει ωθήσει την επορευματοποίηση πιο πολύπλοκων τεχνικών μοντελοποίησης και αναγνώρισης. Τα ως αποτέλεσμα μοντέλα είναι πιο ακριβή και ικανά να κάνουν διάκριση ανάμεσα σε μεγάλους αριθμούς υποψήφιων λέξεων.

Οι τρεις βασικές μεθοδολογίες που χρησιμοποιούνται για να δημιουργήσουν μοντέλα αναφοράς ανεξάρτητα του ομιλητή είναι :

- Η δειγματοληψία
- Η μοντελοποίηση υπολέξεων
- Τα νευρωνικά δίκτυα

2.3.1 Δειγματοληψία

Η δειγματοληψία παράγει μοντέλα υψηλής ποιότητας που είναι προσαρμοσμένα στον πληθυσμό που θα απευθυνθούν και το περιβάλλον ομιλίας της εφαρμογής. Ακολουθεί την ίδια διαδικασία τριών βημάτων που εφαρμόζεται στην εξαρτώμενη από τον ομιλητή μοντελοποίηση.

Η συλλογή δεδομένων είναι πιο εκτεταμένη και περίπλοκη από αυτήν που απαιτείται για τη δημιουργία μοντέλου εξαρτώμενου από τον ομιλητή. Ο αριθμός δειγμάτων που πρέπει να συγκεντρωθούν για κάθε αντικείμενο του λεξιλογίου εξαρτάται από μια πληθώρα παραγόντων , συμπεριλαμβανομένου

- Μέγεθος και ανομοιότητα του πληθυσμού
- Τύπος του λεξιλογίου
- Ροή της ομιλίας
- Ομοιότητα των λέξεων
- Περιβάλλον ομιλίας

Η διεγματοληψία για μεγάλους και διαφορετικούς πληθυσμούς απαιτεί περισσότερα δείγματα από τη δειγματοληψία για ομογενείς πληθυσμούς ομιλητών. Ο πληθυσμός χωρίζεται σε υποομάδες βάσει δημογραφικών πληροφοριών σχετικά με τα εύρη ηλικίας, το φύλλο, πρότυπα διαλέκτων, και άλλα χαρακτηριστικά που είναι πιθανόν να επηρεάσουν την ομιλία. Η συλλογή δεδομένων από υποομάδες απεικονίζει το ποσοστό πληθυσμού αυτής της ομάδας.

> *Για ανεξαρτησία από τον ομιλητή πρέπει να πάρεις μια καλή μίξη από άντρες και γυναίκες. Θέλεις να καθρεφτίσεις το είδος των ανθρώπων που θα χρησιμοποιούν το σύστημα. Το ίδιο ισχύει και για διαλέκτους (Ed Tagg, Αντιπρόεδρος Εφαρμοσμένης Μηχανικής, Voice Control Systems [πρόσφατα με την ConnectWare],personal communication , 1993).*

Η εργασία "*Η φωνή διαμέσου της Αμερικής*" (*Voice Across America*) της Texas Instruments, για παράδειγμα, συνίσταται από τη συλλογή μεγάλων αριθμών δειγμάτων από Αμερικανούς ομιλητές από όλα τα μέρη των Ηνωμένων Πολιτειών. Τα δείγματα είναι για ομιλία διαμέσου του τηλεφώνου. Τα μοντέλα που προκύπτουν από τα αποτελέσματα μπορούν να χρησιμοποιηθούν για την ανάπτυξη λεξιλογίου βασισμένου σε δείγματα για εφαρμογές και βάσεις δεδομένων ομιλιών για μοντελοποίηση υπολέξεων.

Η ανάγκη του εφοδιασμού με ένα ευρύτερο φάσμα μεταβλητότητας απαιτεί τη χρήση πιο σύνθετων υπολογισμών από αυτούς που χρειάζονται για τη δημιουργία μοντέλων εξαρτώμενων από τον ομιλητή. Αλλά οι υπολογισμοί μοιάζουν με αυτούς που χρησιμοποιούνται για τη δημιουργία μοντέλων εξαρτώμενων από τον ομιλητή. Τα συστήματα με ταίριασμα ιχνών ψάχνουν για κοινά πρότυπα ταξινομώντας τα ακουστικά δεδομένα για κάθε λέξη σε ομάδες παρόμοιων προτύπων. Το κέντρο κάθε ομάδας, που λέγεται centroid, γίνεται το πρότυπο αναφοράς για τον αντιπροσωπευτικό τύπο της προφοράς αυτής της ομάδας. Στην δημιουργία HMM, ακουστικά δεδομένα από όλα τα δείγματα για μία λέξη συγχωνεύονται σαν να είχαν δημιουργηθεί από έναν μόνο ομιλητή. Υπολογισμοί γίνονται σε ολόκληρο το σύνολο των δεδομένων χρησιμοποιώντας τον Baum-Welch και/ή άλλους αλγορίθμους για την κατασκευή ενός μόνου HMM. Η SSI (Speech Systems Inc.) χρησιμοποιεί μία παρόμοια προσέγγιση αλλά αυξάνει τα μοντέλα με δέντρα απόφασης.

Το 1991, η BBN (Bolt Beranek & Newman) πρότεινε μία άλλη μεθοδολογία δειγματοληψίας σχεδιασμένη έτσι ώστε να μειώνει το χρόνο και το κόστος της δημιουργίας ενός μοντέλου ανεξάρτητο του ομιλητή. Περικλείει τη συλλογή μιας μεγάλης ποσότητας δεδομένων από έναν μικρό αριθμό ομιλητών. Ο υπολογισμός και η δημιουργία του μοντέλου γίνονται σε τρείς φάσεις. Η πρώτη φάση ταυτίζεται με την παραδοσιακή προσέγγιση: τα συγχωνευμένα δεδομένα για μια λέξη χρησιμοποιούνται για τη δημιουργία ενός μόνο μοντέλου. Αυτό το μοντέλο γίνεται μια κωδικολέξη σ' ένα κωδικοβιβλίο ανυσματικής κβάντισης (vector-quantization codebook) ανεξάρτητης του ομιλητή. Στη δεύτερη φάση ένα μοντέλο εξαρτώμενο του ομιλητή δημιουργείται για κάθε ομιλητή. Τελικά, ένα ή περισσότερα ανεξάρτητα του ομιλητή μοντέλα δημιουργούνται από τον μέσο όρο των

ακουστικών παραμέτρων των εξαρτώμενων από τον ομιλητή HMM. Η BBN αναφέρει ότι μοντέλα ανεξάρτητα του ομιλητή δημιουργημένα χρησιμοποιώντας αυτή τη μέθοδο λειτουργούν καλά σε σύγκριση με μοντέλα δημιουργημένα χρησιμοποιώντας παραδοσιακή δειγματοληψία. Επειδή το τελικό μοντέλο δημιουργείται από μέσους όρους παρά από υπολογισμούς σε συγχωνευμένα δεδομένα, είναι ευκολότερη η ενημέρωση της BBN προσέγγισης με δεδομένα από καινούργιους ομιλητές.

Η Voice Control Systems χρησιμοποιεί μια εντελώς διαφορετική προσέγγιση στην ανεξάρτητη από τον ομιλητή προσέγγιση για μικρά και που προκαλούν σύγχυση λεξιλόγια, όπως ο συλλαβισμός. Απομακρύνουν χαρακτηριστικά λέξεων που προκαλούν σύγχυση γύρω από τη λέξη στόχο. Αυτό λέγεται *διακριτική προπαρασκευή (discriminative training).* Η Voice Control Systems τη χρησιμοποιεί για να επιτυγχάνει πιο ακριβή αναγνώριση για δύσκολα σύνολα λέξεων.

Όταν γίνεται σωστά, η δειγματοληψία παράγει καλά μοντέλα ανεξάρτητα του ομιλητή. Είναι ,παρόλα αυτά, επίπονη εργασία και επιβραδύνει την ανάπτυξη της εφαρμογής, κάνοντας τη δημιουργία μοντέλων μεγάλου λεξιλογίου ιδιαίτερα δύσκολη. Αυτό έχει ωθήσει τους ερευνητές να ψάξουν εναλλακτικές μεθόδους για τη δημιουργία μοντέλων. Μία μοντελοποίηση υπολέξεων αναφέρθηκε στο κεφ. 3 (παρ. 3.1.5) και συζητείται περαιτέρω στην επόμενη παράγραφο.

Οι Rabiner & Juang (1993, κεφ. 5) παρέχουν μια τεχνική περιγραφή τεχνικών και θεμάτων δειγματοληψίας. Οι Wilpon & Rabiner (1985) περιγράφουν μια κοινώς χρησιμοποιούμενη τεχνική ταξινόμησης ονομαζόμενη K-μέσοι όροι. Οι Kubala & Schwartz (1991) παρουσιάζουν την BBN προσέγγιση στη δειγματοληψία και οι Meisel (1991) εξηγούν την τεχνολογία της SSI.

2.3.2 Μοντελοποίηση υπολέξεων

Η μοντελοποίηση των υπολέξεων είναι σχεδιασμένη για ταχεία κατασκευή μεγάλου λεξιλογίου. Διευκολύνει την ανάπτυξη λεξιλογίου για συστήματα μεγάλου λεξιλογίου και αρχικά χρησιμοποιήθηκε αποκλειστικά από συστήματα προσαρμοσμένα σε ομιλητή. Η επέκταση της μοντελοποίησης των υπολέξεων σε συστήματα ανεξάρτητα από τον ομιλητή προσφέρει ταχεία ανάπτυξη λεξιλογίου σ' αυτά τα συστήματα και επιτρέπει την επέκταση της ανεξαρτησίας από τον ομιλητή σε εφαρμογές που απαιτούν μεγάλα λεξιλόγια.

Η αρχική δημιουργία ενός συστήματος μοντελοποίησης υπολέξεων κατευθύνεται προς :

- Την δημιουργία μιας βάσης δεδομένων υπολέξεων
- Την ανάπτυξη μιας τεχνικής για τη σύνδεση των υπολέξεων με σκοπό τη δημιουργία λέξεων μοντέλων
- Τον καθορισμό της μορφής των δεδομένων εισόδου του χρήστη

Αυτές οι διαδικασίες είναι μακρύχρονες και απαιτούν επίμονη εργασία. Η βάση δεδομένων περιέχει όλα τα στοιχεία υπολέξεις του συστήματος. Τα αναμενόμενα δεδομένα εισόδου δίδονται γενικώς μέσω του πληκτρολογίου, αλλά οι πωλητές έχουν ενσωματώσει διαφορετικά συστήματα συλλαβισμού συμβατά στα συστήματά τους. Κάποια επιτρέπουν τον πρότυπο συλλαβισμό, ενώ άλλα αναμένουν μια μορφή φωνητικού συλλαβισμού. Ο

δακτυλογραφημένος συλλαβισμός είναι ευκολότερος για σχεδιαστές εφαρμογών (και χρήστες) για να μάθουν να το χρησιμοποιούν, αλλά για τα Αγγλικά η αντιστοιχία ανάμεσα στον συλλαβισμό και την προφορά είναι αναξιόπιστη. Ο φωνητικός σθλλαβισμός είναι πιο αξιόπιστος αλλά πιο δύσκολος να μαθευτεί. Σαν αποτέλεσμα, οι πωλητές καθυστερούν να προωθήσουν στην αγορά τα εργαλεία δημιουργίας μοντέλων λέξεων.

Η μοντελοποίηση υπολέξεων χρησιμοποιεί HMM's. Μερικοί πωλητές βασίζονται σε ήδη υπάρχοντα `σώματα` ομιλούμενης γλώσσας που περιέχουν μεγάλους αριθμούς δειγμάτων από ένα εύρος ομιλητών. Άλλων τεχνολογιών σχεδιαστές συγκεντρώνουν τα δικά τους δείγματα. Η προτίμηση για εσωτερική συλλογή δεδομένων εξαρτάται από τον πωλητή και το περιβάλλον ομιλίας. Η AT&T, για παράδειγμα, συνέλεξε τα δικά της δείγματα για το δικό της σύστημα μοντελοποίησης υπολέξεων *FlexWord,* που είναι σχεδιασμένο για εφαρμογές βασισμένες στην τηλεφωνία.

> *Οι υπολέξεις στο FlexWord του Conversant δημιουργήθηκαν από δείγματα. Χρησιμοποιήσαμε ένα σύνολο από προτάσεις φωνητικά ισορροπημένες που δημιουργήθηκαν στο MIT. Μετά συλλέξαμε δείγματα ομιλίας αντιπροσωπευτικά από ολόκληρες τις Ηνωμένες Πολιτείες (**Kathy Breslin, Marketing Manager, AT&T Conversant System, AT&T Global Business, personal communication, 1994).***

Ο AT&T Conversant ήταν το πρώτο εμπορικό προϊόν ανεξάρτητο του ομιλητή που πρόσφερε μοντελοποίηση υπολέξεων ως ένα χαρακτηριστικό ανάπτυξης εφαρμογών βασισμένες στην τηλεφωνία.

Αφού συλλεχθούν τα δεδομένα για ένα σύστημα μοντελοποίησης υπολέξεων, τμηματοποιούνται σε μονάδες υπολέξεων. Τα βήματα υπολογισμού και δημιουργίας μοντέλου είναι παρόμοια με εκείνα που περιγράφονται στη μεθοδολογία δειγμάτων. Οι υπολέξεις που έχουν συλλεχθεί τοποθετούνται στη βάση δεδομένων και δοκιμάζονται. Μόλις συμπληρωθεί η βάση των υπολέξεων, δεν απαιτείται περαιτέρω συλλογή δεδομένων εκτός εάν το σύστημα μεταφερθεί σε νέους πληθυσμούς.

Το τελευταίο βήμα της μοντελοποίησης υπολέξεων είναι η δημιουργία του μοντέλου λέξεων χρησιμοποιώντας την καθιερωμένη τεχνική εισόδου. Τα δεδομένα εισόδου τμηματοποιούνται σε μεγάλα κομμάτια υπολέξεων, μεταφρασμένα σε ακουστικές συσχετίσεις, ταιριασμένα με υπάρχοντα μοντέλα υπολέξεων, και ξανασυναρμολογημένα ως ένα πλήρες μοντέλο λέξεων (ή βασική μορφή για συστήματα προσαρμοσμένα στον ομιλητή). Το μοντέλο αποθηκεύεται στο λεξικό του συστήματος. Ο Mark Mandel του Dragon Systems παρομοιάζει αυτήν τη διαδικασία με τη χρήση παιχνιδιών ψευτομαστορέματος (tinker toys). Ο Jay Wilpon της AT&T Bell Laboratories το συγκρίνει με πρόσβαση σε κωδικούς προφοράς σε λεξικά.

Ο Chin-Hui Lee, et al. (1993) και οι Huang & Lee (1993) περιγράφουν τρίφωνη και υπολέξεων μοντελοποίηση. Ο Markowitz (1993a) συγκρίνει τη μοντελοποίηση υπολέξεων με τη δειγματοληψία.

2.3.3 Νευρωνικά Δίκτυα για Μοντελοποίηση ανεξάρτητη από τον Ομιλητή

Η ικανότητα των νευρωνικών δικτύων να εκτελέσουν πολύπλοκη ταξινόμηση δύσκολων δεδομένων τα κάνει ιδανικά για διαφοροποίηση ανάμεσα σε ομιλητές και για τη

δημιουργία μοντέλων ανεξάρτητων του ομιλητή. Οι σχεδιαστές νευρωνικών δικτύων έχουν διάφορες προσεγγίσεις στην μοντελοποίηση ομιλητή. Όπως και με όλη τη σχεδίαση νευρωνικών δικτύων απαιτούν μεγάλους αριθμούς δειγμάτων ομιλίας. Όπως και με τη τεχνική δειγματοληψίας της BBN, τα δείγματα ομιλίας μπορούν να προέρχονται από ένα μικρό αριθμό ομιλητών. Κάποιοι σχεδιαστές έχουν προσαρμόσει υπάρχοντα δίκτυα εξαρτώμενα από τον ομιλητή σε αναγνώριση ανεξάρτητη του ομιλητή (TDNN), αλλά οι περισσότεροι ερευνητές δημιουργούν νέα μοντέλα δικτύων.

Μια προσέγγιση είναι η σχεδίαση ενός δικτύου πρόβλεψης χρησιμοποιώντας ένα πολλαπλών επιπέδων perceptron (MLP) ή άλλη αρχιτεκτονική εμπροσθοτροφοδοτούμενη. Αυτά τα δίκτυα μπορούν να προβλέψουν τις ακουστικές παραμέτρους ενός εισερχόμενου διανύσματος ομιλίας βασιζόμενα στις παραμέτρους προηγούμενων διανυσμάτων. Ονομαζόμενη ως *νευρωνική πρόβλεψη* , αυτή η προσέγγιση είναι συγκρίσιμη με τη γραμμική προβλεπτική κωδικοποίηση.

Κάποιοι ερευνητές καταρτίζουν δίκτυα ώστε να μπορούν να ταξινομούν τους ομιλητές σε ομάδες αντιπροσωπευτικών ομιλητών. Κάθε νέος ομιλητής ταξινομείται μέσα στις αντιπροσωπευτικές ομάδες ομιλητών. Μόλις γίνει αυτή η ταξινόμηση , γίνεται η αναγνώριση με τη σύγκριση των δεδομένων ομιλίας με τα μοντέλα αναφοράς της επιλεγόμενης ομάδας.

Οι Iso & Watanabe (1990) σχεδίασαν ένα νευρωνικό μοντέλο πρόβλεψης. Οι Nakamura & Akabene (1991) περιγράφουν ένα δίκτυο ομάδων και ο Nakamura (1992) κατασκεύασε ένα μεγάλο TDNN για αναγνώριση ανεξάρτητη του ομιλητή.

ΠΡΟΣΑΡΜΟΓΗ ΟΜΙΛΗΤΗ

Εμπορικές εφαρμογές της προσαρμογής ομιλητή εφαρμόσθηκαν πρώτα σε μοντέλα εξαρτώμενα από τον ομιλητή (speaker-dependent) και χρησιμοποιήθηκαν κατά την διάρκεια του κύκλου εγγραφής (enrollment cycle). Το 1988, το σύστημα επίδειξης Dragon Dictate της Dragon System έγινε το πρώτο εμπορικό σύστημα που χρησιμοποίησε on-the-fly προσαρμογή για αναγνώριση μεγάλου λεξιλογίου. Αυτό συγχώνευσε προσαρμογή για το ακουστικό μοντέλο του ομιλητή, το μοντέλο γλώσσας και τον θόρυβο βάθους (background noise). Η προσαρμογή ομιλητή έγινε η ουσιαστική τεχνολογία (core) τόσο για την εξαρτώμενη από ομιλητή όσο και για την ανεξάρτητη από ομιλητή μοντελοποίηση.

Όταν αναφέρονται στα συστήματα τους, οι περισσότεροι πωλητές χρησιμοποιούν τον όρο *προσαρμογή (adaptation)* όταν αναφέρονται στις λειτουργίες προσαρμογής ομιλητή που αναφέρθηκαν σε αυτό το κεφάλαιο. Μερικοί πωλητές, όπως η Dragon Systems, επεκτείνει τον όρο ώστε να συμπεριλάβει την προσαρμογή του μοντέλου ενός συστήματος εσωτερικής γλώσσας (συνήθως ένα μοντέλο N-gram) και / ή την ικανότητα ενός συστήματος να ρυθμίσει τον θόρυβο βάθους (background).

2.4 ΠΡΟΣΑΡΜΟΓΗ ΟΜΙΛΗΤΗ (SPEAKER ADAPTATION)

Θέλαμε το φωνητικό μας μοντέλο να έχει προσαρμοστικότητα απαραίτητη για την αυτοδύναμη αναγνώριση κατά μήκος μιας περιοχής με ομιλητές των οποίων τα φωνητικά αναπνευστικά χαρακτηριστικά διέφεραν ευρέως. Ακόμη, θέλαμε ρυθμούς αναγνώρισης που να εξαρτώνται από τον ομιλητή (speaker-dependent). (John Hampshire II & Alex Waibel, Carnegie Mellon University, "The Meta-Pi network: Connectionist rapid adaptation for high-performance multi-speaker phoneme recognition," 1990, p.165).

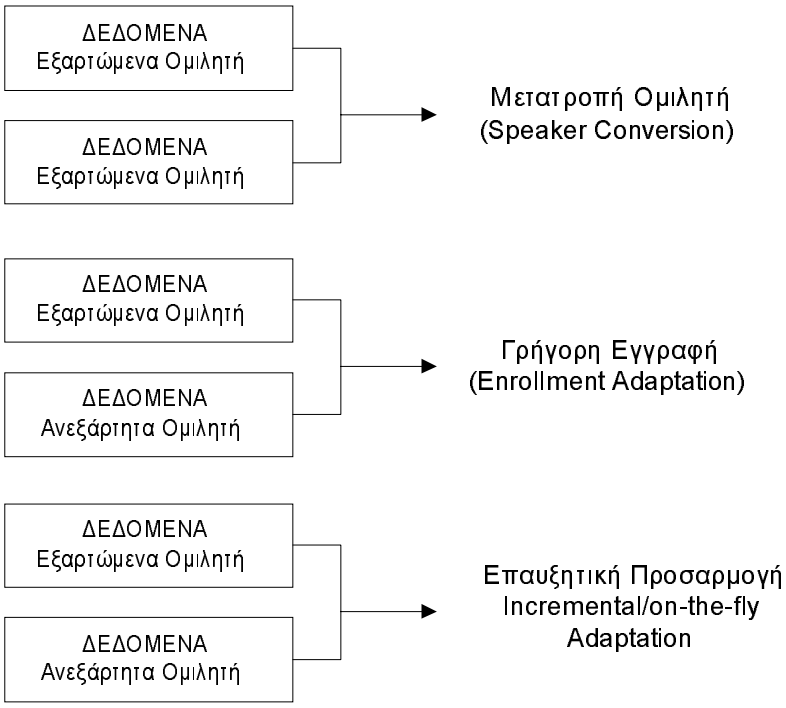
Άσχετα από τις προηγούμενες μεθοδολογίες μοντελοποίησης, Προσαρμογή Ομιλητή σημαίνει τροποποίηση των προτύπων αναφοράς που αποτελούν πηγή αναφοράς (reference patterns) παρά δημιουργία νέων προτύπων.

> *Πρέπει να γίνει φανερό, ότι τα μοντέλα στο DragonDictate της είναι τα ίδια όπως στα μοντέλα που είναι πλήρως ανεξάρτητα από τον ομιλητή (full speaker-independent). Το DragonDictate είναι μοντέλο που προσαρμόζεται στον ομιλητή (speaker adaptive). Δηλαδή, υπάρχει αρκετό από το μοντέλο μίας λέξης το οποίο είναι επαρκώς ανεξάρτητο από τον ομιλητή για αποδεκτή αρχική αναγνώριση, αλλά άμεσα τροποποιούνται με προσαρμογή ώστε να συμπεριλάβουν και τα ιδιαίτερα χαρακτηριστικά του ομιλητή (Mak Mandel, Dragon Systems Inc., personal communication, 1993).*

Τέτοιες μετατροπές βοηθούν τα συστήματα αναγνώρισης να επεξεργαστούν εισερχόμενα πρότυπα τα οποία διαφέρουν από εκείνα τα οποία χρησιμοποιούνται για την παραγωγή των προτύπων αναφοράς. Ένα άλλο πλεονέκτημα της προσαρμογής του ομιλητή (Speaker adaptation) είναι ότι η επεξεργασία της προσαρμογής μπορεί να παράγει καλή αναγνώριση για ένα ομιλητή χρησιμοποιώντας ένα μικρό ποσό από πληροφορία που είναι εξαρτώμενη από τον ομιλητή (speaker-dependend).

> *Με μόνο 300 προτάσεις προσαρμογής, ο ρυθμός λάθους του συστήματος προσαρμογής μας ήταν ο ίδιος όπως σε ένα σύστημα το οποίο είναι εξαρτώμενο από τον ομιλητή και έχει εκπαιδευθεί με 600 προτάσεις.. Το παραπάνω δείχνει ότι η αναγνώριση φωνής με προσαρμογή ομιλητή (speaker adaptive speech recognition) χρησιμοποιεί δεδομένα εκπαίδευσης (training data) που είναι περισσότερο αποτελεσματικά από ότι η αναγνώριση φωνής που εξαρτάται από τον ομιλητή (speaker-dependent speech recognition) (Xuedong Huang & Kai-Fu Lee, Carnegie Mellon University, " On speaker-independent, speaker-dependent, and speaker-adaptive speech recognition," 1993,p.151).*

Η προσαρμογή ομιλητή μπορεί να πραγματοποιηθεί με πολλούς τρόπους. Η εικόνα 2.1 απεικονίζει τις ποιο συχνά χρησιμοποιούμενες μεθόδους.



Εικόνα 2.1 Μερικές μορφές προσαρμογής ομιλητή

ΤΑΧΕΙΑ ΚΑΤΑΧΩΡΗΣΗ (RAPID ENROLLMENT)

Το πρώτο εμπορικό σύστημα για χρησιμοποίηση γρήγορης εγγραφής ήταν το PE-100 της Speech Systems, Inc. SSI έκανε ένα πείραμα το 1986 να προσαρμόσει τα μοντέλα ανεξάρτητα-ομιλητή στο λεξολόγιο του PE-100. Αργότερα εγκατέλειψαν την τεχνική για χάρη ξεχωριστού γυναικείων και ανδρικών μοντέλων ανεξάρτητων-ομιλητή . Τα μεγάλα συστήματα ορθογραφικών λεξιλόγιων της IBM τα χρησιμοποιούν αποκλειστικά. Τα ορθογραφικά συστήματα των Kurzweil AI and Philips συνδυάζουν γρήγορη καταχώρηση (rapid enrollment) μαζί με on-the-fly προσαρμογή.

Η μετατροπή Ομιλητή (Speaker conversion) περιλαμβάνει την μετατόπιση ενός προτύπου αναφοράς εξαρτώμενου-ομιλητή για ένα άτομο σε πρότυπο αναφοράς εξαρτώμενου-ομιλητή για ένα δεύτερο ομιλητή. Αυτό γεινλά λειτουγεί στο επίπεδο λέξης. Η προσαρμογή καταχώρησης (enrllment adaptation) χρησιμοποιεί δεδομένα από μία αρχική διαδικασία καταχώρηση (η οποία ονομάζεται ταχεία καταχώρηση – rapid enrollment) για να

τροποποιήσει μοντέλα ανεξάρτητα-ομιλητή, δημιουργώντας τα έτσι ώστε να μοιάζουν με την φωνή ενός ξεχωριστού ομιλητή. Η Γρήγορη Εγγραφή (rapid enrollment) γενικά περικλείει τον χρήστη να διαβάζει ένα προκαθορισμένο κείμενο ή να παρέχει άλλες υπαγορευμένες είσοδος (dictated input). Αυτό συμβαίνει γενικά σε μία και μοναδική περίοδο που διαρκεί μία εως δύο ώρες.

Μία άλλη προσέγγιση είναι η συλλογή δεδομένων προσαρμογής on-the-fly. Η on-the-fly προσαρμογή πραγματοποιείται μετατρέποντας μοντέλα αναφοράς ενώ ο ομιλητής χρησιμοποιεί το σύστημα. Σχεδιάστηκε για να αυξήσει την αποδοχή του χρήστη, αποβάλλοντας την καταχώρηση (enrollment), να και θα μπορούσε να συνδυαστεί με την ταχεία καταχώρηση (rapid enrollment) για την παραγωγή γρηγορότερων ασφαλέστερων κερδών

2.4.1 Προσαρμογή Προτύπων (Template adaptation)

Προσαρμογή ομιλητή που χρησιμοποιεί ίχνη είναι σπάνια. Εφαρμόσθηκε σε πρότυπα ολόκληρης λέξης (whole-word pattern) και απαιτεί ίχνη που κατασκευάστηκαν χρησιμοποιώντας έντονη εκπαίδευση. Η μετατροπή ομιλητή (speaker conversion) πραγματοποιήθηκε χρησιμοποιώντας ένα μικρό ποσό δεδομένων από ένα νέο ομιλητή για να τροποποιήσει κωδικολέξεις σε ένα διανυσματικώς κβαντισμένο κωδικό-βιβλίο. Καταχώρηση (enrollment) ή on-the-fly προσαρμογή των ανεξάρτητων-ομιλητή προτύπων αναφοράς ομαδοποιεί τα ακουστικά δεδομένα από τον νέο ομιλητή μαζί με αυτά από τα πρότυπα αναφοράς.

Scott Instruments (τμήμα της Voice Control Systems, Inc) χρησιμοποίησε on-the-fly προσαρμογή μαζί με τα δικά του πρότυπα ανεξάρτητου-ομιλητή. Οποτεδήποτε ένας συνδυασμός ανιχνεύθηκε ως λάθος από τον ομιλητή, το σύστημα πρόσθεσε την σωστή ακουστική πληροφορία στο σωστό μοντέλο αναφοράς και την αφαίρεσε από το λανθασμένο μοντέλο αναφοράς.

Rabiner & Juang (1993, chapter 5) παρέχει μία λεπτομερή τεχνική ανάλυση της γκάμας των τεχνικών προσαρμογής. Smith, et al. (1990) περιγράφει την προσέγγιση του προτύπου προσαρμογής της Scott Instruments.

2.4.2 HMM Προσαρμογή

Προσαρμογή ομιλητή του HMM μπορεί να εφαρμοσθεί στις λέξεις, αλλά περισσότερο ταυτίστηκε με μεγάλα λεξικά, συστήματα υπολέξεων. (Δές παράγραφο 2.3.2). Η χρήση προσαρμογής ομιλητή με μοντέλα υπολέξεων μεταβάλει τα τρίφθογγα (ή άλλες μονάδες υπολέξεων) τα οποία αποτελούν τις βασικές μορφές του μοντέλου αναφοράς μιάς εφαρμογής ή συστήματος. Η διαδικασία της προσαρμογής λέξεων ή υπολέξεων λέγεται απεικόνιση (mapping). Οι μετατροπές γενικά σώζονται σε ένα αρχείο και αναπαριστούν το μοντέλο ενός και μοναδικού ομιλητή.

Δύο κοινές προσεγγίσεις στην HMM προσαρμογή είναι η διανυσματικώς κβαντισμένη κωδικο-βιβλίο προσαρμογή (vector quantization codebook adaptation) και η HMM προσαρμογή με ακουστικές παράμετρους (HMM acoustic-parameter adaptation). Η

διανυσματικώς κβαντισμένη κωδικο-βιβλίο προσαρμογή (vector quantization codebook adaptation) συνεπάγεται την μετατροπή των κωδικών λέξεων στο διανυσματικώς κβαντισμένο κωδικό-βιβλίο. Οι τροποποιήσεις κάνουν τα ακουστικά πρότυπα (patterns) του κωδικο-βιβλίου να μοιάζουν με την φωνή ενός συγκεκριμένου ομιλητή. Αυτή είναι μία ενναλακτική προσέγγιση για προσαρμογή ομιλητή διότι η χρήση των διανυσματικώς κβαντισμένων κωδικό-βιβλίων μειώνει την αναζήτηση. Η κωδικό-βιβλίο αναπαράσταση προσφέρει επίσης την την πιθανότητα της πραγματοποίησης συνολικών μετακινήσεων διανυσμάτων βασισμένη σε μικρές ποσότητες δεδομένων που προέρχονται από ένα νέο ομιλητή.

HMM προσαρμογή με ακουστικές παράμετρους (HMM acoustic-parameter adaptation) περικλείει την τροποποίηση των ακουστικών παραμένων του HMM. Αυτοί οι παράμετροι ρυθμίζονται έτσι ώστε να φτικτούν να μοιάζουν περισσότερο με τα δεδομένα που παρέχονται από τον νεό ομιλητή.

Λεπτομερείς τεχνικές περιγραφές αρκετών προσεγγίσεων της προσαρμογής μπορούν να βρεθούν στα βιβλία των Huang & Lee (1993), Rabiner & Juang (1993, κεφ. 5) και στο Schwartz, et all. (1987).

2.4.3 Προσαρμογή Ομιλητή με χρήση Νευρωνικών Δικτύων

Αντίθετα από άλλες μεθοδολογίες που περιγράφησαν στην προσαρμογή ομιλητή, τα νευρωνικά δίκτυα σχεδιάστηκαν για προσαρμογή ομιλητή είτε τροποιώντας τα δικά τους εσωτρικά μοντέλα αναφοράς είτε δημιουργώντας μόνιμα μοντέλα ομιλητή-εξαρτώμενα για νεόυς ομιλητές. Πράγματι, αυτοί χρησιμοποιούν τα δικά τους εσωτερικά μοντέλα αναφοράς για να ταξινομήσουν τα πρότυπα ομιλίας των νέων χρηστών σαν να ήταν μέλη μίας ήδη υπάρχουσας τάξης αναφοράς ή ομάδας ομιλητών (speaker cluster). Η διαδικασία είναι παρόμοια με αυτή που χρησιμοποιείται από τα νευρωνικά δίκτυα για να κάνουν αναγνώριση εξαρτώμενη από τον ομιλητή (δες παράγραφο 2.3.3).

TDNN δίκτυα της CMU έχουν χρησιμοποιηθεί για την σχεδίαση μίας ποικιλίας συστημάτων προσαρμογής ομιλητή. Το Meta-Pi δίκτυο υπερδομής της CMU, για παράδειγμα, είναι μία παραλλαγή του TDNN. Η ανάπτυξη του είχε παρακινηθεί από την ανάγκη της επέκτασης του εξαρτώμενου από ομιλητή TDNN σε εργασίες πολύ-ομιλητή. Συνδέει αρκετά εξαρτώμενα από ομιλητή TDNN μονάδες τα οποία εργάζονται παράλληλα για να ταξινομήσουν την φωνή του νέου ομιλητή.Μετά την εφαρμογή του δικτύου Meta-Pi σε ένα ομιλητή προσδιορισμένου ως MHT, οι Hampshire και Waibel ανέφεραν:

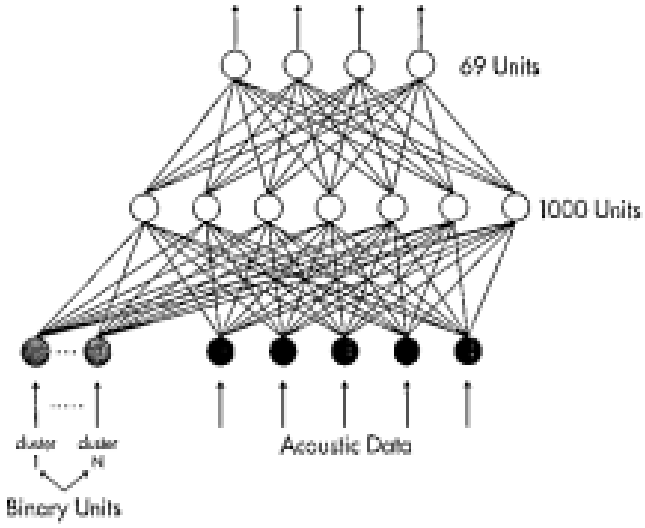
Πράγματι, η υπερδομή μαθαίνει να μοντελοποιεί τον ομιλητή MHT με ένα δυναμικό συνδυασμό από άλλους άνδρες ομιλητές και ακόμη επιτυγχάνει 99,9% ρυθμό αναγνώρισης στην φωνή του MHT (John Hampshire II & Alex Waibel, Carnegie Mellon University, ' The Meta-Pi Network,' 1990, p.167).

Μία άλλη ομάδα ερευνητών συνδύασε ένα TDNN εξατώμενου απο ομιλητή εκπαιδευμένου να κάνει αναγνώριση λέξης με ένα άλλο δίκτυο feedforward δίκτυο εκπαιδευμένου να εκτελεί προσαρμογή ομιλητή. Ο συνδυασμός βελτίωσε την απόδοση των TDNN δικτύων για νέους ομιλητές, αλλά ακόμη ήταν λιγότερο ακριβές από τα παραδοσιακά μοντέλα εξαρτώμενααπό ομιλητή για αυτούς τους ομιλητές.

Το νευρωνικό δίκτυο ομάδας ομιλητών (speaker cluster) (SCNN) είναι ένα υβριδικό MLP-HMM σύστημα το οποίο κατασκευάστηκε χρησιμοποιώντας μία πολυ-σταδιακή (multi- stage) διαδικασία ανάπτυξης . Κατά την διάρκεια του πρώτου στάδιου της ανάπτυξης,

το MLP εκπαιδεύτηκε να ομαδοποιεί τους ομιλητές σε δύο ομάδες βασιζόμενο πάνω σε ακουστικές παράμετρους που έχουν σχέση με το φύλο (sex-linked). Το δίκτυο είχε δύο κόμβους εξόδου: ένα για να εκφράσει την πιθανότητα ότι ο ομιλητής ήταν γυναίκα και ο άλλος για να υποδείξει την πιθανατότητα ότι ο ομιλητής ήταν άνδρας. Ενα άλλο δίκτυο κατασκευάστηκε για να ταξινομήσει τους ομιλητές σε περισσότερες από μία ομάδες (cluster).

Το MLP δίκτυο (σχήμα 2.2) υλοποιεί αναγνώριση φωνής. Αναγνωρίζει ένα φθόγγο (phone) από μία πρόταση που εκστομίσθηκε από ένα άγνωστο ομιλητή. Χρησιμοποιεί την ομιλούσα είσοδο και τις ομάδες με τις παραμέτρους αναγνώρισης-ομιλητή που ορίσθηκαν σε προηγούμενα πειράματα ώς οι δικές το πηγές δεδομένων. Κάθε κόμβος εξόδου εκφράζει την πιθανότητα ότι τα δεδομένα αναπαριστούν ένα συγκεκριμένο φθόγγο.



Το SNN βρέθηκε ότι βελτιώνει την ακρίβεια αναγνώρισης για νέους ομιλητές αλλά όπως στο Meta-Pi σύστημα, η ακρίβεια του δεν προσεγγίζει αυτήν των παραδοσιακών μοντέλων.

Οι Hampshire και Waibel (1990) περιγράφουν το Meta-Pi δίκτυο και οi Fukuzawa και Σια (1992) παρέχουν μία τεχνική ανάλυση του TDNN feeedforward συστήματος προσαρμογής. Οι Konig & Morgan (1993) περιγράφουν το SCNN.

2.5 Η ΣΥΝΕΧΗΣ ΣΕΙΡΑ ΤΗΣ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΟΜΙΛΗΤΗ

(The speaker-modelling continuum)

Οι μεθοδολογίες μοντελοποίησης ομιλητή που περιγράφηκαν από τεχνολογική άποψη χαρακτηρίζονται από διαφορετικές τεχνολογικές προσεγγίσεις και σκοπούς. Οταν ειδωθούν από την μεριά των λειτουργικών απαιτήσεων μιας εφαρμογής, οι διαφορές μεταξύ τους είναι λιγότερο καθαρές.

Πρέπει να αναγνωρισθεί ότι ασχολούμαστε με ένα συνεχές (continuum). Η προσέγγιση εξαρτώμενου ομιλητή μπορεί να μοντελοποιήσει ομάδες ανθρώπων όπως και ξεχωριστές….αλλά οι προσπάθειες είναι διαφορετικές (Jeff Hill, Vice President of New product Development, Voice Processing Corp., personal communication, 1993).

Εξάρτηση από ομιλητή, μοντελοποίηση πολυ-ομιλητή, και ανεξαρτησία από ομιλητή μοιάζουν με σημεία κατα μήκος ενός συνεχούς, όπως αυτό που απεικονίζεται στο σχήμα 2.3. Αναγνώριση της εισόδου ως την φωνή ενός μοναδικού ατόμου και αναγνώριση όλων των ομιλητών χρησιμοποιώντας μία καθορισμένη γλώσσα αναπαριστούν τα άκρα του συνεχούς.

Η μοντελοποίηση εξαρτώμενου από ομιλητή δεν θα έπρεπε να συγχισθεί με την εξακρίβωση ομιλητή (system verification). Η επαλήθευση ομιλητή η οποία είναι ένα συστατικό των συστημάτων ασφαλείας, έχει τον ρόλο της εξακρίβωσης της ισχυριζόμενης ταυτότητας του ομιλητή παρά την κατανόηση των λεγόμενων του ομιλητή. Αντίθετα, η κύρια λειτουργία της μοντελοποίησης εξαρτώμενου από ομιλητή είναι να μεγιστοποιήσει την ασφάλεια αναγνώρισης για την φωνή του ατόμου το οποίο παρείσχε τα δείγματα. Η εξάρτηση από ομιλητή δεν θα έπρεπε να χρησιμοποιηθεί ως εργαλείο ασφάλειας για να εμποδίσει την πρόσβαση στο σύστημα από μη εξουσιοδοτημένους ομιλητές.

Μοντέλα λέξεων εξαρτώμενα από ομιλητή δημιουργήθηκαν για ένα ομιλητή και μπορούν να χρησιμοποιηθούν από άλλους ομιλητές του ίδου φύλου των οποίων τα φωνητικά χαρακτηριστικά είναι παρόμοια με αυτά των ομιλητών των οποίων η φωνή χρησιμοποιήθηκε για να δημιουργηθούν τα μοντέλα. Θα είναι λιγότερο ακριβή για άλλους ομιλητές, αλλά η συνολική ακρίβεια μπορεί να βελτιωθεί χρησιμοποιώντας υψηλώς διαφοροποιημένα σύνολα λέξεων. Παρόμοια, μοντέλα πολυ-ομιλητών μπορούν να χρησιμοποιηθούν από μεμονωμένα άτομα τα οποία δεν συνεισφέραν δείγματα κατά την διάρκεια της καταχώρισης. Μερικές φορές, όταν ο πληθυσμός των ομιλητών είναι μικρός και ομογενής, ένα μοντέλο πολυ-ομιλητή μπορεί να χρησιμοποιηθεί σαν να είχε σχεδιαστεί για αναγνώριση ανεξάρτητη από ομιλητή. Τέτοιες επεκτάσεις της μοντελοποίησης πολυ-ομιλητή πρέπει να γίνονται με προσοχή.

Η διάταξη της μοντελοποίησης ανεξάρτητης από ομιλητή στο σχήμα 2.3 επιδιώκει να δώσει έμφαση στο γεγονός ότι μοντέλα ανεξάρτητα από ομιλητή δεν αναμένεται να λειτουργούν άψογα με την φωνή ανθρώπων οι οποίοι δεν προέρχονται από τον πληθυσμό από τον οποίο τα δείγματα εξήχθησαν.

Βάζοντας ένα σύστημα Ελέγχου φωνής του Ενωμένου Βασιλείου στην Βόρεια Σκωτία η αναγνώριση για ορισμένες λέξεις δεν θα είναι αποτελεσματική εξαιτίας της διαλέκτου..

Στο μέλλον, νευρωνικά δίκτυα ή παρεμφερείς τεχνολογίες πιθανόν να βοηθήσουν να αρθούν οι φραγμοί για παγκόσμια αναγνώριση. Ομως κανένα υπάρχων εμπορικό σύστημα αναγνώρισης είναι ικανό να αναγνωρίσει όλους τους ομιλητές από κάθε πλυθυσμό. Αυτό συμβαίνει γιατί οι περισσότερες εφαρμογές αναγνώρισης ανεξάρτητα από ομιλητή, περικλείουν μιά διαδικασία εφεδρείας (backup procedure) για ομιλητές που δεν μπορούν να χρησιμοποιήσουν το σύστημα αναγνώρισης.

2.6 ΘΕΜΑΤΑ ΣΕ ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΕΞΑΡΤΩΜΕΝΟΥ ΑΠΟ ΟΜΙΛΗΤΗ

Τυπικά, συστήματα εξαρτώμενου από ομιλητή μπορούν να επιτύχουν καλύτερη απόδοση αναγνώρισης από ότι τα συστήματα ανεξάρτητου από ομιλητή…Οπωσδήποτε, αυτό επιτεύχθηκε με το κόστος μίας νέας συνόδου

καταχώρησησ η οποία έγινε για κάθε ένα νέο ομιλητή. (Connectionist Speech Recognition: A Hybrid Approach, 1994, p.5).

Η λειτουργία καταχώρησης (enrollment process) αποτελεί μία σημαντική πηγή ενδιαφέροντος στην μοντελοποίηση εξαρτώμενης από ομιλητή. Είναι η μοναδική πηγή δημιουργίας μοντέλου λέξης το οποίο απαιτεί για κάθε χρήστη να προμηθεύσει με ένα τουλάχιστο ομιλών δείγμα κάθε λέξης στην εφαρμογή.

Οι χρήστες μπορεί να ενοχληθούν από την λειτουργία καταχώρησης (enrollment process), καθιστώντας την ακατάλληλη για εφαρμογές μεγάλων λεξιλόγιων.

Τα περισσότερα συστήματα εξαρτώμενα από ομιλητή επιτρέπουν την εκπαίδευση να γίνεται με μικρές αυξήσεις. Αλλά ακόμα και αυτό γίνεται απαράδεκτο όταν ο αριθμός των λέξεων στην εφαρμογή υπερβαίνει τις εκατό. Γίνεται δε αδιανόητο όταν πρόκειται για διακόσιες λέξεις. Παρ'ολα αυτά, μία από τις πρώτες εκδόσεις του μεγάλου συστήματος ορθογραφικού λεξιλόγιου Kurzweil AI, απαίτησε από τους χρήστες να διδάξουν και τις επτακόσιες λέξεις στο σύστημα.

Ακόμα και για μικρές εφαρμογές λεξιλόγιου η λειτουργία καταχώρησης (enrollment process) πρέπει να προγραμματισθεί προσεκτικά και να σχεδιασθεί να μεγιστοποιήσει την συνεργασία και το ενδιαφέρον των χρηστών. Μπορεί για παράδειγμα να χρησιμοποιηθεί σαν ένα μέσο εξοικίωσης των χρηστών με την εφαρμογή και το σύστημα αναγνώρισης φωνής.

Ενα άλλο θέμα σχετικό με συστήματα εξαρτώμενου ομιλητή καθώς και προσαρμογής ομιλητή, είναι η φόρτωση και εκφόρτιση μοντέλων ομιλητή. Κάθε φορά ένας αναρμόδιος ομιλητής αρχίζει να χρησιμοποιεί ένα σύστημα αξαρτώμενου από ομιλητή, αλλά τα μοντέλα ιχνών/λέξεων των χρηστών πρέπει να φορτωθούν στην μνήμη. Αν η ταχύτητα είναι ένα θέμα ή η εφαρμογή χαρακτηρίζεται από συχνές αλλαγές ομιλητή, οι απαιτήσεις για φόρτωμα μοντέλου πιθανόν να καταρτίσουν την αναγνώριση εξαρτώμενη από ομιλητή απαράδεκτη. Για παράδειγμα, σε ορισμένα περιβάλλοντα η δημιουργία επειγόντων νοσοκομειακών εκθέσεων, περικλείει παρεμβαλλόμενες συνεισφορές από αρκετούς επαγγελματίες. Με τέτοιες ρυθμίσεις, ο απαιτούμενος χρόνος για να ξεφορτωθεί το μοντέλο από προηγούμενο ομιλητή και να φορτωθεί το καινούργιο μοντέλο μπορεί να εμφανισθεί υπερβολικός, καθιστώντας την ανεξαρτησία ομιλητή την μόνη αποδεκτή δυνατότητα για αναγνώριση φωνής.

Εφόσων ένα μοντέλο ανεξάρτητου από ομιλητή σχετίζεται με τα ακουστικά πρότυπα του ατόμου που το εκπαίδευσε, είναι πρόκληση να χρησιμοποιηθεί ως συσκευή ασφάλειας. Αν τα ακουστικά πρότυπα ενός μη εξουσιοδοτημένου ατόμου είναι παρόμοια με αυτά ενός εξουσιοδοτημένου ομιλητή, το σύστημα ανεξάρτητου από ομιλητή πιθανόν να λειτουργήσει καλά και με την νέα φωνή. Μία καλα σχεδιασμένη εφαρμογή με ακουστικά ευδιάκριτες λέξεις στα ενεργά του λεξιλόγια θα ήταν ικανό με ασφάλεια να αναγνωρίσει εισόδους από μη εξουσιοδοτημένους ομιλητές ακόμα και αν οι φωνές τους μόνο μέσα σε κάποια όρια υα μπορούσαν να μοιάζουν με αυτές των ατόμων που τα εκπαίδευσαν.

2.6.1 Ασφαλεια Σε Μοντελοποιηση Ανεξαρτητου Απο Ομιλητη

Η καταχώρηση (enrollment) μπορεί να είναι ενοχλητική και εργαστηριακώς εντατική, αλλά η τεχνολογία ανεξάρτητη από ομιλητή είναι ακόμα ασφαλέστερη από την τεχνολογία εξαρτημένης από ομιλητή για σύνολα λέξεων όπως τα ψηφία. Η σύγχιση του '5' και '9'

μπορεί να εμφανιστεί σχεδόν σε κάθε ργαλείο αναγνώρισης φωνής. Αν αυτό το λάθος φαίνεται παράξενο, είναι χρήσιμο να σημειωθεί ότι οι τηλεφωνήτριες έχουν εκπαιδευθεί να ξεχωρίσουν αυτή την διαφορά λέγοντας 'nigh-in' για το '9' ενώ το στρατιωτικό προσωπικό εκπαίδευθηκε να λέει 'niner'. Η ικανότητα ενός συστήματος αναγνώρισης να σπάσει την καταχώρηση σε μικρά τμήματα και να επιτρέψει την καταχώρηση ή την επανεκπαίδευση ενώ η εφαρμογή είναι on-line μπορεί να βοηθήσει ώστε να μειωθούν τα αρνητικά φαινόμενα της καταχώρησης. Για να διατηρηθεί η υψηλή ακρίβεια, αυτές οι τεχνικές ικανότητες θα επρεπε να συνοδεύονται κ εκτός από τον ενθουσιασμό της υλοποίησης του συστήματος μεταξύ των χρηστών και από την προσεκτική σχεδίαση ενός προγράμματος επανασχεδίασης.

Τα μοντέλα λέξεων των συστημάτων ανεξάρτητου από ομιλητή μπορούν να είναι πολύ ακριβή ακόμα και σε περιβάλλοντα ομιλίας με θόρυβο. Η λειτουργία καταχώρησης (enrollment process) καθαυτή βελτιώνει την συνολική ακρίβεια εγκλιματίζοντας τους χρήστες στον αναγνωριστή (recognizer) και στην εφαρμογή. Η ακρίβεια επίσης μπορεί να βελτιωθεί όταν η καταχώρηση γίνεται στο περιβάλλον στο οποίο η εφαρμογή θα χρησιμοποιηθεί. Αυτό διευκολύνει τον αναγνωριστή να διακρίνει μεταξύ του θορύβου περιβάλλοντος και της φωνής.

Ο αριθμός των συμβόλων (token) που απαιτείται για την επίτευξη καλής ακρίβειας εξαρτάται από την σχεδίαση του συστήματος αναγνώρισης, το μέγεθος λεξιλογίου, την παρουσία μπερδεμένων λέξεων, την ακρίβεια των απαιτήσεων της εφαρμογής, τις απαιτήσεις της εργασίας και την φύση του περιβάλλοντος ομιλίας. Γενικά η απόδοση αυξάνεται με επιπρόσθετη εκπαίδευση, ειδικά για σύνολα από υψηλώς μπερδεμένα λεξιλόγια.

Η ακρίβεια μπορεί να εξασθενήσει ως αποτέλεσμα αλλαγών στην φωνή των χρηστών. Οι αλλαγές αυτές πιθανόν αναπαριστούν κανονικές μετατοπίσεις στην φωνή τού ομιλητή ή του περιβάλλοντος. Αν για παράδειγμα οι ομιλητές χρησιμοποιήσουν το σύστημα σε διαφορετικές χρονικές στιγμές στο εργασιακό τους πρόγραμμα, πιθανόν να υπάρχει ανάγκη να παράσχουν σύμβολα όταν οι φωνές τους είναι φρέσκες ή όταν είναι κουρασμένοι. Σε περιπτώσεις που περικλείει ποικίλους βαθμούς κούρασης, σύμβολα μπορεί να παρθούν από κουρασμένες και ξεκούραστες φωνές.

Μερικές φορές, η εκτέλεση της εργασίας θα αλλάξει τον τρόπο που ένα άτομο ομιλεί. Ορειβασία, ανέβασμα και προστατευτικό κάλυμμα κεφαλής είναι πιθανόν να μεταβάλλουν την φωνή των χρηστών. Σε αυτές τις περιπτώσεις, η καταχώρηση πιθανόν να θεωρηθεί άχρηστη εάν γίνει κάτω από τις συνθήκες της εργασίας.

Οι νέοι χρήστες τείνουν να ομιλούν στο σύστημα αναγνώρισης με ένα υπερβολικά προσεκτικό τρόπο. Αυτή η πρακτική εξαφανίζει τον εγλκιματισμό στο σύστημα και πιθανόν να απαιτήσει συλλογή από επιπλέον σύμβολα.

Η λειτουργία καταχώρησης (enrollment process) από μόνη της καλλιεργεί ένα αφύσικο στυλ ομιλίας το οποίο καλείται *φαινόμενο απαγγελίας (recitation effect)*. Το φαινόμενο απαγγελίας εμφανίζεται όταν οι ομιλητές χρησιμοποιούν ένα επίπεδο, μηχανικό στύλ. Μοντέλα λέξεων τα οποία δημιουργήθηκαν από ομιλία σύμφωνα με το φαινόμενο απαγγελίας θα είναι επιρρεπή σε λάθη, διότι διαφέρουν από το στυλ που χρησιμοποιήθηκε όταν οι λέξεις αυτές μιλήθηκαν στο κείμενο της εφαρμογής.

Η ανακαταχώρηση μπορεί να προγραμματισθεί κατά την συντήρηση της εφαρμογής. Για την μείωση πιθανών απαγοητεύσεων χρηστών, πολλά συστήματα ανεξάρτητα από ομιλητή παρέχουν εργαλεία επανεκπαίδευσης για ενημέρωση, βελτίωση ή διόρθωση μεμονωμένων μοντέλων λέξεων ενώ η εφαρμογή τρέχει.

2.6.2 Εισαγοντας Λεξεις Σε Μοντελοποιηση Εξαρτωμενη Απο Ομιλητη

Οι περισσότεροι αναγνωριστές εξαρτώμενοι από ομιλητή, μην έχοντας ενσωματωμένο λεξιλόγιο, επιτρέπουν τους σχεδιατές εφαρμογών να επιλέξουν και να ορίσουν όλα τα απαραίτητα λεξιλόγια. Κάθε λέξη που δημιουργείται πρέπει να διδαχθεί από όλους τους χρήστες. Μερικά προιόντα ενημερώνουν τον χρήστη ότι υπάρχουν αδίδαχτα στοιχεία; άλλα πάλι προιόντα δεν το κάνουν. Γενικά, νέα στοιχεία στα λεξιλόγια διδάσκονται με τέτοιο τρόπο σαν να ήταν το αρχικό λεξιλόγιο εφαρμογής.

2.7 ΘΕΜΑΤΑ ΣΕ ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΠΟΛΥ-ΟΜΙΛΗΤΗ

Η μοντελοποίηση πολύ-ομιλητή γενικά υλοποιείται σε εφαρμογές με μικρό ή μεσαίο μέγεθος λεξιλόγιο όταν ούτε είναι δυνατό αλλά ούτε επιθυμητό να φορτωθούν και να ξεφορτωθούν μοντέλα ομιλητών. Η μοντελοποίηση πολύ-ομιλητή είναι ιδιαίτερα χρήσιμη όταν η εφαρμογή χαρακτηρίζεται από ταχεία μετατόπιση από ένα χρήστη σε άλλο. Είναι επίσης χρήσιμη για μονάδες προφορικής πρόσληψης (verbal sign-on modules) μεγαλύτερων εφαρμογών. Μία μονάδα πρόσληψης (sign-on module), για παράδειγμα, αποτελείται από την εισαγωγή του προσωπικού κώδικα αναγνώρισης. Σε πολλές βιομηχανικές εφαρμογές κώδικες πρόσληψης (sign-on codes) αποτελούνται από το όνομα του ατόμου ή τον αριθμό του εργαζόμενου. Η λειτουργία τους είναι να πουν στο σύστημα ποια μοντέλα ομιλητών να προσεγγίσουν. Δεν είναι κατάλληλα σαν συσκευές προστασίας γιατί τα μοντέλα πολύ-ομιλητή μπορούν εύκολα να χρησιμοποιηθούν από μη εξουσιοδοτημένα άτομα των οποίων οι φωνές είναι παρόμοιες με αυτές εξουσιοδοτημένων ατόμων.

Δεν είναι φρόνιμο να χρησιμοποιείται μοντελοποίηση πολύ-ομιλητή για μεγάλες εφαρμογές λεξιλογίου διότι οι εξεζητημένες διαφοροποιήσεις που πρέπει να κάνουν δεν μπορούν να γίνουν από ένα μοντέλο πολύ-ομιλητή. Αυτά αξιώνουν περισσότερο σοφιστική μοντελοποίηση ανεξάρτητη ομιλητή ή μοντελοποίηση προσαρμοσμένη σε ομιλητή (speaker adaptive).

2.7.1 Ακρίβεια (Accuracy) Σε Μοντελοποίηση Πολύ-Ομιλητή

Εξαιτίας του γεγονότος ότι τα μοντέλα ομιλητών δημιουργήθηκαν από τις φωνές δύο ή περισσότερων ατόμων, η συνολική ακρίβεια των συστημάτων πολύ-ομιλητή μπορεί να είναι αποδεκτή αλλά είναι φτωχότερη από αυτή της αναγνώρισης ανεξάρτητης από ομιλητή. Αυτό εξαιτίας της επέκτασης της μεταβλητότητας στις παραμέτρους αναγνώρισης για μοντέλα λέξης τα οποία καλύπτουν περισσότερα από ένα ομιλητή και εξαιτίας του γεγονότος ότι η τεχνολογία που χρησιμοποιήθηκε σχεδιάστηκε να αναπαριστά τα χαρακτηριστικά της φωνής ενός μοναδικού ομιλητή. Η ακρίβεια είναι καλύτερη όταν η ομάδα ομιλητών είναι μικρή και ομογενής; είναι φτωχότερη όταν ένα μοντέλο πολύ-ομιλητή χρησιμοποιείται σαν να ήταν μοντέλο ανεξάρτητο από ομιλητή. Τέτοια χρήση μοντελοποίησης πολύ-ομιλητή θα πρέπει να γίνεται με προσοχή.

2.7.2 Εισάγοντας Λέξεις Σε Μοντελοποίηση Πολύ-Ομιλητή

Απαξ και η εφαρμογή με μοντελοποίηση πολύ-ομιλητή αναπτυχθεί, η προσθήκη του νέου λεξιλογίου στο σύστημα μπορεί να είναι προβληματική. Απαιτεί την προμήθεια δειγμάτων για όλα τα νέα λεξιλόγια από αρκετούς ομιλητές των οποίων οι φωνές κωδικοποιήθηκαν στα μοντέλα. Αυτός που αναπτύσσει εφαρμογές (application developer) ή ο διαχειριστής συστήματος (system manager) πρέπει να εξασφαλίσει ότι ένα σημαντικός αριθμός από δείγματα έχουν προμηθευθεί για την ασφαλή αναγνώριση των νέο - προστιθέμενων λέξεων.

2.8 ΘΕΜΑΤΑ ΣΕ ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΑΝΕΞΑΡΤΗΤΟΥ ΑΠΟ ΟΜΙΛΗΤΗ

Η μοντελοποίηση ανεξάρτητου από ομιλητή είναι βασική για εφαρμογές που σχεδιάστηκαν για να είναι προσβάσιμες από χρήστες μιας φοράς. Πολλές εφαρμογές βασισμένες σε τηλέφωνα ή τηλεφωνικούς θάλαμους εμπίπτουν σε αυτήν την κατηγορία. Η ανεξαρτησία από ομιλητή επίσης συνεισφέρει στην δυνατότητα αποδοχής των εφαρμογών μεγάλων λεξιλογίων.

2.8.1 Ακρίβεια (Accuracy) Σε Μοντελοποίηση Ανεξάρτητη Από Ομιλητή

Εν ολίγοις, τα μοντέλα ανεξάρτητα από ομιλητή είναι λιγότερο ακριβή από παρεμφερή καλοσχεδιασμένα μοντέλα εξαρτημένα από ομιλητή. Αυτή είναι μια κατανοητή συνέπεια του βαθμού μεταβλητότητας η οποία πρέπει να αποκτηθεί από τα μοντέλα ανεξάρτητα από ομιλητή.

2.8.1.1 Δημιουργία Μοντέλων Ανεξάρτητα Από Ομιλητή Χρησιμοποιώντας Δειγματοληψία

Τα μοντέλα ανεξάρτητα από ομιλητή τα οποία δημιουργήθηκαν χρησιμοποιώντας δειγματοληψία ποικίλουν πάρα πολύ με την ποιότητα. Όταν γίνεται καλά, η εφαρμογή δειγματοληψίας μπορεί να παράγει μοντέλα υψηλής ποιότητας. Διαφορές στην ποιότητα είναι αποτέλεσμα:

- Του αριθμού των δειγμάτων που χρησιμοποιήθηκαν για την δημιουργία των μοντέλων
- Της αντιπροσωπευτικότητας των δειγμάτων σε σχέση με τον πληθυσμό των χρηστών και του περιβάλλοντος ομιλίας.
- Της ποιότητας των αλγορίθμων που χρησιμοποιήθηκαν για την δημιουργία των μοντέλων.

Ο αριθμός των δειγμάτων που πρέπει να συλλεχθούν για την επίτευξη ενός υψηλού βαθμού ακρίβειας εξαρτάται από:

- Το μέγεθος και την πυκνότητα του αναμενόμενου πληθυσμού ομιλητών
- Των χαρακτηριστικών του θορύβου του περιβάλλοντος ομιλίας
- Της ακρίβειας των απαιτήσεων της εφαρμογής
- Τα χαρακτηριστικά του λεξιλογίου

Ετερογενείς πληθυσμοί απαιτούν μεγαλύτερη δειγματοληψία από ότι ομογενείς πληθυσμοί. Όμως, όπως έχει αναφερθεί στην παράγραφο 2.5 τα μοντέλα ανεξάρτητα από ομιλητή δεν σχεδιάστηκαν για να είναι παγκόσμια. Κατασκευάστηκαν για συγκεκριμένους πληθυσμούς χρηστών και καθορισμένα περιβάλλοντα ομιλίας. Οι μηχανικοί ανάπτυξης τέτοιων εφαρμογών αναγνωρίζουν αυτούς τους περιορισμού και καταλαβαίνουν ότι όταν μοντέλα ανεξάρτητα από ομιλητή τα οποία κατασκευάστηκαν για ένα πληθυσμό χρησιμοποιηθούν για ένα άλλο πληθυσμό, τότε η ακρίβεια αποκλίνει. Αυτό είναι ιδιαίτερα πιθανόν να συμβεί όταν διάλεκτοι συναντούνται.

Θορυβώδη περιβάλλοντα ομιλίας και αυτά με μεταβαλλόμενα χαρακτηριστικά θορύβου απαιτούν επιπρόσθετη δειγματοληψία. Τέτοια δειγματοληψία σχεδιάστηκε για την σύλληψη φωνής ενσωματωμένης στον αναμενόμενο θόρυβο περιβάλλοντος ή για την αναγνώριση δειγμάτων θορύβου τα οποία πρέπει να απομακρυνθούν από την είσοδο.

Δυσνόητες λέξεις και σημαντικά στοιχεία λεξιλογίων απαραίτητα για την λειτουργία της εφαρμογής απαιτούν μεγαλύτερη δειγματοληψία. Ακριβή αναγνώριση για πολυσύλλαβες λέξεις, όπως «Μασσαχουσέτη», μπορεί να επιτευχθεί με λίγα έως 15 δείγματα (tokens) στην περίπτωση που δεν υπάρχουν παρόμοιες λέξεις. Αντίθετα, χίλια ή και περισσότερα δείγματα ανά λέξη απαιτούνται για την δημιουργία μοντέλων για σύνολα από μονοσύλλαβες λέξεις.

Όπως μα την καταχώρηση εξαρτημένης από ομιλητή (speaker-dependent enrollment), η δειγματοληψία για την ανάπτυξη μοντέλων ανεξάρτητα από ομιλητή μπορεί να υφίσταται το *φαινόμενο απαγγελίας (recitation effect).* Μία μέθοδος για την προμήθεια περισσότερο φυσικών δειγμάτων είναι η δημιουργία ενός πρωτότυπου συστήματος για την συλλογή δειγμάτων από ομιλητές ενώ αυτοί χρησιμοποιούν το σύστημα. Μία άλλη μέθοδος είναι η χρησιμοποίηση προσαρμογής on-the-fly-.

2.8.1.2 Δημιουργώντας Μοντέλα Ανεξάρτητα Από Ομιλητή Χρησιμοποιώντας Μοντελοποίηση Υπο-λέξης

Η δημιουργία μοντέλων λεξιλογίου ανεξάρτητα από ομιλητή τα οποία χρησιμοποιούν μοντελοποίηση υπο-λέξης είναι εξαιρετικά γρήγορα, εύκολα και γενικά ανέξοδα σε σύγκριση με την δειγματοληψία. Επίσης είναι λιγότερο ακριβή από ότι τα μοντέλα επιπέδου λέξης (word-level).

Ο μόνος λόγος για την ανωτερότητα των μοντέλων τα οπία βασίζονται σε λέξεις είναι ότι προσομειώνουν τις λέξεις με πολύ ψηλότερο βαθμό λεπτομέρειας. Τα φαινόμενα της συνάρθρωσης (coarticulation) και (context dependence) είναι ενσωματωμένα στα μοντέλα (Bahl, Lalit, et al, IBM Speech Recognition Group, "Acoustic Marcov models used in the Targona speech recognition system,", 1988, p. 498).

Μοντέλα τα οποία κατασκευάστηκαν από μοντελοποίηση υπο-λέξης δεν είναι μόνο λιγότερο συντονισμένα στην φωνή του πληθυσμού στόχου αλλά γενικά δεν προσομειώνουν καλά ούτε και το περιβάλλον ομιλίας. Αυτοί οι περιορισμοί καθιστούν δύσκολη την

παραγωγή υψηλής ποιότητας που απαιτείται από τα μοντέλα ανεξάρτητα από ομιλητή για μερικές εφαρμογές. Ακόμα και συστήματα τα οποία χρησιμοποιούν προσεκτικά σχεδιασμένα μοντέλα υπο-λέξης ανεξάρτητα από ομιλητή, όπως το AT&T Conversant, πιθανόν να βασίζονται σε μοντέλα τα οποία βασίζονται σε δειγματοληψία για σημαντικές ή συγκεχυμένες ομάδες λέξεων όπως οι αριθμοί..

Ένα σημαντικό τμήμα της έρευνας κατευθύνεται προς την ανάπτυξη τεχνικών και εργαλείων μοντελοποίησης υπο-λέξης. Ένα από τα αποτελέσματα αυτής της εργασίας θα είναι η βελτίωση της ακρίβειας.

2.8.2 Εισάγοντας Λέξεις Σε Μοντελοποίηση Ανεξάρτητη από Ομιλητή

Τα συστήματα ανεξάρτητα από ομιλητή περιέχουν λεξιλόγια τα οποία κατασκευάστηκαν από την ανάλυση μεγάλο αριθμό δειγμάτων ή από την συγχώνευση υπο-λέξεων. Το ενσωματωμένο λεξιλόγιο όμως του συστήματος, μπορεί να αποτύχει να περιέχει την αναγκαία καθορισμένη εφαρμογή ορολογίας. Για την διευθέτηση αυτής της παράλειψης, όλοι οι πωλητές προσφέρουν custom εργαλεία ανάπτυξης λεξιλογίου. Οι υπόλοιπες παράγραφοι προσφέρουν μία ανασκόπηση θεμάτων σχετικών με την μοντελοποίηση ομιλητή.

2.8.2.1 Δειγματοληψία Από Μηχανικούς Ανάπτυξης Εφαρμογών

Μερικοί πωλητές προσφέρουν εργαλεία μοντελοποίησης σε μηχανικούς ανάπτυξης εφαρμογών με λεπτομερείς οδηγίες για να κατευθύνουν την διαδικασία δειγματοληψίας. Αυτά τα εργαλεία μπορεί να είναι εξαιρετικά, αλλά η δειγματοληψία επίσης απαιτεί προσεκτική αρχική σχεδίαση, καλή κατανόηση του πληθυσμού χρηστών, αποτελεσματικά δεδομένα συλλογής και προσεκτικό έλεγχο. Ολοι αυτοί οι παράγοντες επηρεάζουν την ποιότητα αναγνώρισης. Όταν η δειγματοληψία γίνεται σωστά, το αποτέλεσμα είναι αναγνώριση υψηλής ποιότητας. Διαδικασίες δειγματοληψίας που σχεδιάστηκαν ή δεν εκτελέστηκαν καλά μπορεί να κοστίσουν τόσο σε χρόνο όσο και σε χρήμα. Ακόμα και όταν οι πωλητές προσφέρουν δειγματοληψία και εργαλεία δημιουργίας μοντέλων στους πελάτες τους, η καλύτερη μοντελοποίηση γενικά γίνεται από πωλητές συστημάτων ανεξάρτητα από ομιλητή διότι και εμπειρία έχουν με την εκτίμηση των δειγμάτων αλλά και με την δημιουργία μοντέλων.

Σε μερικές περιπτώσεις, όταν τα συστήματα χρησιμοποιούνται αρχικά από ένα μοναδικό ομιλητή ή από ένα μικρό αριθμό από ομιλητές, τα μοντέλα μπορούν να χρησιμοποιηθούν και από τους ίδιους τους τελικούς χρήστες (end user). Η Voice Processing Corporation επιτρέπει σε μοντέλα ανεξάρτητα από ομιλητή να εισαχθούν στα δικά τους μικρά συστήματα λεξιλογίων.

2.8.2.2 Χρησιμοποιώντας Μοντελοποίηση Υπο-λέξης Από Μηχανικούς Ανάπτυξης Εφαρμογών

Οταν ο πωλήτής ενός συστήματος ορθογραφίας (dictation) προμηθεύει ένα λεξιλόγιο το οπίο περιέχει περισσότερο από 100,000 λέξεις, οι περισσότερες απαιτήσεις του λεξιλόγιου μπορούν να συναντηθούν με την πρόσβαση στο λεξικό. Τα προιόντα ορθογραφίας ποικίλουν ανάλογα με το αν μοντέλα λέξεων που προστέθηκαν από τους χρήστες είναι ανεξάρτητα από το σύστημα ή εξαρτωμένα από το σύστημα. Για παράδειγμα η Philips Dictation Systems και η IBM, παράγουν συστήματα ανεξάρτητα από ομιλητή παρόλο που διαφέρουν πάρα πολύ στον τρόπο που πραγματοποιούνται.

2.8.3 Τροποποιώντας τα μοντέλα

Οπως έχει ήδη προαναφερθεί (παράγραφο 2.5) τα μοντέλα ανεξάρτητα από ομιλητή δεν μπορούν να διαχειρισθούν όλους τους ομιλητές με κάθε πιθανή διάλεκτο (μητρική ή ξένη) ομιλώντας σε όλα τα περιβάλλοντα. Εφόσων δημιουργήθηκαν απο δείγματα λέξεων ή υπολέξεων, τα μοντέλα ανεξάρτητα από ομιλητή αντανακλούν τα δεδομένα που χρησιμοποιήθηκαν για να τα δημιουργήσουν. Οποτε η διάλεκτος το περιβάλλον ομιλίας ή το κανάλι ομιλίας (μικρόφωνο, τηλέφωνο, δες κεφ. 7) διαφέρουν από αυτά που χρησιμοποίηθηκαν για να δημιουργήσουν το μοντέλο, το μοντέλο είναι πιθανόν να απαιτήσει τροποποίηση.

Αναγνωρίζοντας την ανάγκη να προσαρμοστούν τα μοντέλα ομιλητή σε μία εφαρμογή, οι περισσότεροι πωλητές προσφέρουν μαζί με τα μοντέλα ανεξάρτητα από ομιλητή και τα εργαλεία στους μηχνικούς ανάπτυξης εφαρμογών για μικρορυθμίσεις σε εκείνα τα μοντέλα φωνής του αναμενόμενου πληθυσμού χρηστών. Εταιρείες με μοντέλα βασισμένα στην δειγματοληψία επιπέδου λέξης (word-level) προσφέρουν εργαλεία δειγματοληψίας και οδηγίες. Μοντέλα υπο-λέξεων μπορούν να τροποποιηθούν εισάγοντας νέα δεδομένα. Για παράδειγμα τα συστήματα HARK της BBN περιέχουν εργαλεία τέτοια εργαλεία που τροποποιούν τα υπάρχοντα μοντέλα ομιλητή εισάγοντας νέα δεδομένα. Τα εργαλεια αυτά δημιουργούν ένα νέο σύνολο από μοντέλα για την πρωτεύουσα εφαρμογή (target application) και δεν δημιουργούν μόνιμες αλλαγές στα πρωτότυπα μοντέλα που οι πωλητές προμηθεύουν.

2.9 ΘΕΜΑΤΑ ΣΕ ΜΟΝΤΕΛΑ ΠΡΟΣΑΡΜΟΓΗΣ

Πιθανόν κάποιος να αμφισβητήσει ότι ένα ιδανικό σύστημα είναι εκείνο που αρχίζει με σύστημα ανεξάρτητο από ομιλητή και προσαρμοζόμενο σε ένα ομιλητή, αυξάνεται με το χρόνο (Xuedong Huang & Kai-Fu Lee, Carnegie Mellon University, "On speaker-independent, speaker-dependent, and speaker-adaptive speech recognition," 1993, p. 150).

Ο βαθμός της προσαρμογής για ένα ιδανικό σύστημα ποικίλει ανάλογα με την σχεδιαζόμενη χρήση του. Η προσαρμογή ωφελεί περισσότερο εκείνες τις εφαρμογές στις οποίες οι χρήστες θα προσπελάσουν την εφαρμογή κατ' επανάληψη. Είναι ιδιαίτερα πολύτιμο για εφαρμογές μεγάλων λεξιλόγιων για τα οποία εκτεταμένη καταχώρηση δεν θα ήταν ανεχτή.

Η προσαρμογή ομιλητή είναι ακατάλληλη για εφαρμογές που περικλείουν σύντομη αλληλεπίδραση με χρήστης του παλιού καιρού. Αλλά ακόμα και όταν οι χρήστες προσπελαύνουν την εφαρμογή σε τακτική βάση, ο βαθμός της προσαρμογής ομιλητή πρέπει να εκτιμηθεί ώστε να καθορίσει το όφελος της χρήσης ενός συστήματος προσαρμογής

ομιλητή παρά ένα σύστημα ανεξάρτητο ομιλητή ή ένα σύστημα το οποίο δεν κάνει καθόλου αναγνώριση φωνής. Μερικές μεγάλες εφαρμογές λεξιλογίων, για παράδειγμα, πιθανόν να περιέχουν ταχεία ολίσθηση από ένα ομιλητή σε ένα άλλο σε συνθήκες μεγάλης πίεσης και όταν δύο ή τρία λεπτά καθυστέρησης μπορεί να φανούν μη ανεκτά. Η ταχεία καταχώρηση πιθανόν να μην φανεί «ταχεία» σε χρήστες οι οποίοι πρέπει να προμηθεύσουν μίας ή δύο ώρες ομιλίας. Όπως και με την καταχώρηση εξαρτημένης από ομιλητή, η εμπειρία μπορεί να είναι απογοητευτική για μερικούς ομιλητές κατά την διάρκεια της διαδικασίας. Αυτός είναι ένας λόγος για τον οποίο η Philips προσφέρει μοντέλα ανεξάρτητα από ομιλητή σαν τμήμα του εξοπλισμού αναγνώρισης φωνής..

Τέτοια ερεθίσματα πρέπει να ισορροπούνται απέναντι στην μεγαλύτερη ακρίβεια των μοντέλων προσαρμογής ομιλητή καθώς συγκρίνονται με τα μοντέλα ανεξάρτητα από ομιλητή. Θα μπορούσαν επίσης να συγκριθούν με την μεγαλύτερη αποδοτικότητα της χρήσης συστημάτων φωνής έναντι των χειρόγραφων εγγράφων ή με την οικονομία στο κόστος και την ταχύτητα κατά την χρήση αναγνώρισης φωνής παρά χρησιμοποιώντας αντιγραφή του κειμένου.

2.9.1 Ακρίβεια στην Προσαρμογή Ομιλητή

Το κύριο πλεονέκτημα της χρήσης προσαρμογής ομιλητή είναι η βελτίωση της ακρίβειας που είναι αποτέλεσμα των μετατροπών στα ακουστικά χαρακτηριστικά των υπαρχόντων μοντέλων αναφοράς. Η αρχική ποιότητα αναγνώρισης πιθανόν να είναι ασήμαντη για συστήματα τα οποία βασίζονται σε προσαρμογή on-the-fly, μία διαδικασία η οποία μπορεί να θεωρηθεί σαν μία σοφιστική παραλλαγή της καταχώρησης (enrollment), Σε μεγάλες εφαρμογές λεξιλόγιων, η προσαρμογή ομιλητή ικανοποιεί την λεπτομερή μοντελοποίηση η οποία απαιτείται για την λεπτή διάκριση ανάμεσα σε ένα μεγάλο αριθμό από υποψήφιες λέξεις.

2.9.2 Εισάγοντας Λέξεις στην Προσαρμογή Ομιλητή

Η ικανότητα της εισαγωγής λέξεων σε ένα σύστημα ή μια εφαρμογή εξαρτάται ολόκληρη από την τεχνολογία που χρησιμοποιήθηκε για τη δημιουργία των μοντέλων ομιλητών. Μεγάλα συστήματα λεξιλογίων τα οποία περιέχουν βασικές μορφές (Baseforms) ανεξάρτητες από ομιλητή οι οποίες δημιουργήθηκαν από μοντελοποίηση υπο-λέξεων, περιέχουν μεγάλα εφεδρικά λεξιλόγια. Αν μία λέξη δεν βρέθηκε στο λεξιλόγιο, τότε τα θέματα προμηθεύθηκαν από τον πωλητή του συστήματος ή από τεχνικές μοντελοποίησης υπο-λέξεων.

2.10 ΘΕΜΑΤΑ ΟΜΙΛΗΤΩΝ

Αγχος-Δυναμικός τόνος, «lamps and goats» και αποδοχή είναι τα θέματα ομιλητών τα οποία θα συζητηθούν στις παρακάτω παραγράφους.

2.10.1 Αγχος-δυναμικός τόνος

Το άγχος-δυναμικός τόνος μπορεί να είναι αποτέλεσμα της συγκίνησης του χρήστη, ένα αποτέλεσμα της εργασίας, μία αντίδραση σε μία επείγουσα κατάσταση ή μία αντίδραση σε ένα θορυβώδες περιβάλλον ομιλίας. Ολες οι μορφές άγχους-Δυναμικού τόνου παράγουν αλλαγές στην ομιλία. Αυτές οι μεταβολές περικλείουν πιο σύντομη ομιλία; περισσότερη φωνητική προσπάθεια που χρησιμοποιήθηκε για ομιλία; ασήμαντη υποστήριξη αναπνοής, πιο σφιχτή θέση των σιαγόνων, τα οποία προκαλούν διαφορετική θέση; και κατά συνέπεια βραχνάδα στην φωνή. Τέτοιες αλλαγές σαν αυτές είναι εμφανής και μπορεί να έχουν ισχυρή αρνητική επίδραση στην ακρίβεια της αναγνώρισης. Σε μερικές περιπτώσεις έχει παρατηρηθεί παρέκκλιση 60 % όταν μοντέλα αναφοράς που δημιουργήθηκαν σε συνθήκες έλλειψης άγχους χρησιμοποιήθηκαν με στρεσαρισμένη φωνή.

Ασήμαντη ανταπόκριση από ένα σύστημα αναγνώρισης είναι πιθανόν να δημιουργήσει άγχος-δυναμικό τόνο και να οδηγήσει σε μεγαλύτερη επιδείνωση της απόδοση από τον άνθρωπο και τον αναγνωριστή. Στις περισσότερες εφαρμογές, εάν η αρχική ομιλία εισόδου δεν αναγνωρίζεται σωστά, το σύστημα αναγνώρισης θα ρωτήσει τον χρήστη να επαναλάβει την είσοδο. Συχνά, οι ομιλητές θα επαναλάβουν την είσοδο, όπως θα το έκαναν εάν επικοινωνούσανε με ανθρώπους οι οποίο αποτυγχάνουν να καταλάβουν: δηλαδή θα χρησιμοποιούσαν περισσότερη καθαρή προφορά ομιλίας και πιθανόν και πιο αργό ρυθμό ομιλίας. Εφόσων αυτά τα πρότυπα δεν ταιριάζουν με τα αποθηκευμένα μοντέλα, η απόδοση του αναγνωριστή είναι πιθανόν να εκφυλισθεί περισσότερο.

Όπως έχει ήδη αναφερθεί στην παράγραφο 2.6.1, σε μερικές περιπτώσεις είναι πιθανόν να εκτελεστεί η εκπαίδευση σε περιβάλλον έντασης. Εφόσων μία τέτοια εκπαίδευση δεν είναι συνήθως προαιρετική, οι ερευνητές αναπτύσσουν μεθόδους οι οποίες να μεταβάλλουν τα κανονικά πρότυπα ομιλίας ώστε να μοιάζουν με δυναμικού τόνου ομιλία (stressed). Η έρευνα έχει περιπλακεί ακόμη περισσότερο από το γεγονός ότι τα φαινόμενα του δυναμικού τόνου ποικίλουν ανάλογα στις λέξεις και μέσα στις προτάσεις.

2.10.2 Lamps and Goats (Πρόβατα και)

Μερικά προβλήματα μοντελοποίησης εμφανίζονται να συνδέονται αποκλειστικά με τους ομιλητές. Η φωνή των περισσότερων ανθρώπων εύκολα μοντελοποιείται χρησιμοποιώντας μία ή περισσότερες από τις μεθοδολογίες που περιγράφηκαν στο Τεχνολογικό μέρος της εργασίας αυτής. Στην βιομηχανία της αναγνώρισης ομιλίας αυτοί οι ομιλητές ονομάζονται (lamps) πρόβατα. Αυτοί είναι η απόλαυση των μηχανικών εφαρμογών.

Αλλοι ομιλητές έχουν πρότυπα φωνής τα οποία είναι δύσκολο από τα συστήματα αναγνώρισης φωνής να συλληφθούν. Αυτοί οι ομιλητές ονομάζονται (goats). Μοντελοποίηση εξαρτημένη από ομιλητή (speaker-dependent), προσαρμοσμένη σε ομιλητή (speaker-adaptive) και πολύ-ομιλητή (multi-speaker) απαιτούν επιπρόσθετη εκπαίδευση από τους παραπάνω ομιλητές και σε μερικές περιπτώσεις δεν θα επιτύχουν να δημιουργήσουν ένα καλό σύνολο από μοντέλα. Όταν αυτοί χρησιμοποιήσουν συστήματα ανεξάρτητα από

ομιλητή πιθανόν να αναγκασθούν να προσεγγίσουν την εναλλακτική συσκευή εισόδου ή τον εφεδρικό ανθρώπινο τρόπο που παρέχεται από το σύστημα.

Αν και δεν είναι ξεκάθαρο γιατί μερικοί ομιλητές έχουν δυσκολία, γενικά υποθέτεται ότι αυτοί διαθέτουν ασυνήθη φωνητικά χαρακτηριστικά ή ότι δεν είναι συνεργάσιμοι. Ενώ η μία ή και οι δύο από τις παραπάνω περιπτώσεις μπορεί να είναι η πηγή του προβλήματος είναι επίσης πιθανόν ότι

η μειωμένη απόδοση αναγνώρισης για ένα συγκεκριμένο ομιλητή μπορεί να είναι αποτέλεσμα από μία αλληλεπίδραση μεταξύ λεξιλόγιου, συσκευής αναγνώρισης και ομιλητή και ότι δεν είναι αποτέλεσμα μόνο του ομιλητή (David Pisoni, "Automatic measurement of speech recognition.performance: A comparison of six speaker-dependent recognition devices,: 1986, p.18).

Αυτή η θέση υποστηρίχθηκε από επιπρόσθετη έρευνα που έγινε στα πλαίσια του Ευρωπαικού προγράμματος ESPRIT από το υποπρόγραμμα SAM (Speech Assessment Methodology) και η οπία τελείωσε το 1992 (δες κεφάλαιο 8). Στόχος του SAM ήταν να καθορισθούν πρότυποι ελέχγοι (standard tests) για προιόντα αναγνώρισης φωνής χρησιμοποιώντας μόνο πρότυπα HW και SW. Παρ'ολο που υπήρχε ένας μεγάλος αριθμός από συντελεστές οι οποίοι επηρεάζουν την ακρίβεια των προιόντων τα οποία ελέχθησαν σε πιλοτική μελέτη, μία από τις πιο σημαντικές πηγές μεταβολής της ακρίβειας ήταν ο ομιλητής. Η αλληλεπίδραση μεταξύ ομιλητή και συσκευής αναγνώρισης συναντήθηκε περισσότερο από 25 % στην συνολική μεταβολή.

Εάν σε περισσότερους από ένα ή δύο ομιλητές συναντήθηκε μειωμένη αναγνώριση είναι απίθανο ότι ο μηχανικός σχεδίασης συνάντησε ένα κοπάδι από goats. Οι πιθανότερες πηγές είναι μάλλον ή ελαττωματική σχεδίαση ή ακατάλληλο εργαλείο αναγνώρισης ή μειωμένη συνεργασία χρηστών (δες παράγραφο 2.10.3)

2.10.3 Εγκριση-Αποδοχή (Acceptance)

Αντίθετα από άλλους τρόπους εισόδου, η αναγνώριση ομιλίας είναι εξαιρετικά ευαίσθητη στην συνεργασία μεταξύ των χρηστών του συστήματος. Η έλλειψη αποδοχής-έγκρισης μπορεί να είναι αποτέλεσμα ελλειπής σχεδίασης εφαρμογής, μειωμένης ακρίβειας ή παρεξήγηση μεταξύ της ροής μίας εφαρμογής και των τρόπων με τους οποίους οι ομιλητές καταλαβαίνουν την εργασία.

Η σχεδίαση μίας εφαρμογής και η ποιότητα της αναγνώρισης μπορεί να είναι μικρότερης σπουδαιότητας στην επιτυχία μίας εφαρμογής από ότι οι ακόλουθοι συντελεστές οι οποίοι είναι αποδεκτοί από τους χρήστες:

- Αντίληψη πλεονεκτημάτων του συστήματος
- Συμμετοχή στην ανάπτυξη του συστήματος
- Φόβος Ηλεκτρονικών Υπολογιστών / Τεχνολογίας
- Φόβος αλλαγής

Πράγματι, η ίδια εφαρμογή που κατασκευάστηκε με το ίδιο προιόν αναγνώρισης μπορεί να επιτύχει την μία στιγμή και να αποτύχει κάποια άλλη εξαιτίας του βαθμού συνεργασίας των χρηστών του συστήματος.

Όταν μία εφαρμογή σχεδιάστηκε για μία συγκεκριμένη ομάδα επαναλαμβανόμενων χρηστών, οι συντελεστές οι οποίοι είναι αποδεκτοί από τους χρήστες και ορίστηκε

παραπάνω αποδεικνύει ότι χρειάζεται ενεργή συμμετοχή των αντιπροσώπων του πληθυσμού ομιλητών στην σχεδίαση και έλεγχο της εφαρμογής. Επίσης ο ίδιος κατάλογος συντελεστών αποδεικνύει ότι χρειάζονται αποτελεσματικά, και καλοσχεδιασμένα προγράμματα τα οποία είναι προσανατολισμένα προς τον χρήστη και την εκπαίδευση (user orientation and training program).

Ένα εντυπωσιακό παράδειγμα της σπουδαιότητας της αποδοχής από τον χρήστη παρέχεται από την χρηματοοικονομική βιομηχανία. Δύο χρηματιστηριακά γραφεία αποφάσισαν να χρησιμοποιήσουν το σύστημα VoiceTrader της Verbex Voice System το οποίο παρουσιάστηκε από την Banker's Monthly ως ένα από τα αξιολογότερα προιόντα για το 1989 στην υπηρεσία για την χρηματοικονομικό και τραπεζικό κλάδο. Όταν ο Smith Barney Shearson εκλέχθηκε να χρησιμοποιήσει το εν λόγω σύστημα, αυτοί ενέπλεξαν τους χρηματιστές ενεργά στην διαδικασία και εισήγαγαν ένα χαρακτηριστικό στην εφαρμογή τους το οποίο βοηθά τους χρήστες να κάνουν μία γρήγορη, (on-the-fly) εκπαίδευση όταν χρειάζεται. Ο Smith Barney ανέφερε συνολική ακρίβεια αναγνώριση φωνής 96-99 %, άσχετα από το επίπεδο θορύβου στο δωμάτιο και τις αλλαγές στην φωνή των χρηματιστών στην διάρκεια μίας χρηματιστηριακής συνόδου. Αυτό σημαίνει ότι το σύστημα αναγνώρισε την σωστή λέξη ή έκφραση 96-99 % φορές. Η Natwest Market Ltd στο Λονδίνο, και η οποία ήταν επίσης χρήστης του ίδιου συστήματος είχε την αντίθετη εμπειρία. Το σύστημα τους είχε πλήρη αποτυχία στο dealing room άσχετα αν κατά την διάρκεια του ελέγχου και της εκπαίδευσης είχε ακρίβεια 100 %. Λανθασμένες αναγνωρίσεις ή καθόλου αναγνώριση ανάγκασε τους χρήστες να νιώσουν έντονη αποστροφή προς την τεχνολογία και να σταματήσουν να χρησιμοποιούν το σύστημα. Η Netwest αναγνώρισε ότι ένα μεγάλο μέρος στην μειωμένη απόδοση του συστήματος ήταν εξαιτίας της «ψυχολογίας» των χρηστών οι οποίοι δεν προετοιμάστηκαν στο να δουν το VoiceTrade σαν ένα βοήθημα. Αντίθετα, αυτοί το είδαν νε φόβο και έδειξαν ελάχιστη προθυμία να παράσχουν καλή εκπαίδευση ή να προβούν σε επανεκπαίδευση των προβληματικών λέξεων.

Όταν ο πληθυσμός των χρηστών αποτελείται από ένα μεγάλο αριθμό από χρήστες του παλιού καιρού (one-time users), η προηγούμενη λίστα των συντελεστών αποδοχής υποστηρίζει για την υλοποίηση ενός μικρού πρωτότυπου για την εκτίμηση της απόκρισης του χρήστη, θέματα μοντέλου ομιλητή και συνολική σχεδίαση Interface. Ασχετα από την φύση του πληθυσμού χρηστών, ο μηχανικός ανάπτυξης ποτέ δεν πρέπει να υποτιμήσει την ικανότητα ενός χρήστη στο να αποτύχει να συμμορφωθεί με το δικό του παράδειγμα (Judith Tschirgi, Director Services & Speech Technology, AT&T Network Systems, Personal Communication.).

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΑΡΑΡΤΗΜΑ Ι

ΕΦΑΡΜΟΓΕΣ ΑΠΟ ΤΗΝ ΚΙΝΗΤΗ ΤΗΛΕΦΩΝΙΑ