



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
Μεταπτυχιακό Ηλεκτρονικού Αυτοματισμού

Δημητρίου Ιακωβίδη (Α.Μ.97510)

Οι Παύσεις και η Χρονική Δομή του Λόγου

Εργασία στο μάθημα: Επικοινωνία με Ομιλία
Διδάσκων: Γεώργιος Κουρουπέτρογλου

Αθήνα 1999

Για την πραγματοποίηση της εργασίας αυτής, χρησιμοποιήθηκε ως πηγή το κεφάλαιο *Pauses and the Temporal Structures of Speech*, που συνέγραψε η Brigitte Zellner – Laboratoire d’analyse informatique de la parole (LAIP), Université de Lausanne, CH-1015 LAUSANNE, Switzerland – του συγγράμματος “Fundamentals of speech synthesis and speech recognition”, Eric Keller, John Willey & Sons (1994).

Στοιχεία από Ελληνική και ξένη βιβλιογραφία ελήφθησαν για τον εμπλουτισμό και την πιο εύστοχη μετάφραση στα Ελληνικά του Αγγλικού κειμένου και των επιστημονικών όρων που αυτό περιείχε. Επίσης, αρκετά είναι τα σημεία εκείνα στα οποία δίδονται περιφραστικές καθώς και γραφικές επεξηγήσεις, με στόχο την κατανόηση του κειμένου από τον αναγνώστη.

Στο τέλος της εργασίας, υπάρχει ένας ιδιαίτερα χρήσιμος πίνακας με επεξηγήσεις και ορισμούς επιστημονικών όρων της Αγγλικής γλώσσας και τις αντιστοιχίες τους στην Ελληνική.

Ιακωβίδης Δημήτριος
Μεταπτυχιακό Ηλεκτρονικού Αυτοματισμού,
Αθήνα 1999.

ΠΕΡΙΕΧΟΜΕΝΑ

<i>ΠΕΡΙΕΧΟΜΕΝΑ</i>	3
<i>ΠΕΡΙΛΗΨΗ</i>	5
<i>ΕΙΣΑΓΩΓΗ</i>	7
<i>ΠΑΥΣΕΙΣ</i>	8
Η Κατηγοριοποίηση των Παύσεων.....	8
Η Φυσική και Γλωσσολογική Κατηγοριοποίηση.....	8
Μια Ψυχολογολογική Κατηγοριοποίηση.....	9
Οι Πηγές Προέλευσης των Παύσεων.....	10
Η Χρονική Δομή του Λόγου και οι Περιορισμοί που υπαισέρονται σε αυτή.....	10
Η Λειτουργία των Διαλέξικών Παύσεων που γίνονται αντιληπτές (μεγαλύτερες από 200ms).	12
Οι Παύσεις και η Έννοια της Λεκτικής Ευφράδειας.....	14
Ενδιαφέρουσες Μελέτες - Παραδείγματα.....	16
<i>Η Δομή της Διάρκειας του Λόγου</i>	17
<i>Ένα Εργαλείο Γραμματικής Ανάλυσης: Οι Εκτελεστικές Δομές</i>	18
Είναι οι Εκτελεστικές Δομές Βασισμένες στις Συντακτικές και/ή στις Προσωδιακές Δομές ;..	20
Οι Εκτελεστικές Δομές και οι Συντακτικές Δομές.....	20
Εκτελεστικές Δομές και Προσωδιακές Μονάδες.....	21
Πως οι Προσωδιακές Δομές Γεννώνται στον Ανθρώπινο Εγκέφαλο ;.....	22
<i>Η Αυτόματη Γένεση των διαφόρων Χρονικών Διαρκειών: Αλγόριθμοι</i>	23
Νόμο-Κυβερνώμενα Συστήματα.....	23
Στατιστικά συστήματα.....	27
Ο Αλγόριθμος Keller-Zellner.....	28
Μοντέλα Πρόβλεψης της θέσης των Κύριων Ορίων μιας Φράσης και Εισαγωγής Παύσεων χρησιμοποιώντας μια Στοχαστική Ανεπιρέαστη Από Συμφραζόμενα Γραμματική (ΣΑΑΣΓ)...	30
Τα Μοντέλα Πρόβλεψης Σχηματικά.....	30
Τα Δεδομένα Ομιλίας.....	31
Εκμάθηση των Θέσεων των Κύριων Ορίων των Φράσεων και των Θέσεων των Παύσεων χρησιμοποιώντας μια ΣΑΑΣΓ.....	32
Υπολογισμός των Παραμέτρων για την Πρόβλεψη χρησιμοποιώντας μια ΣΑΑΣΓ.....	33
Μοντέλο πρόβλεψης με τη χρήση Νευρωνικού Δικτύου.....	35
Το Μοντέλο Campbell.....	36
Το Μοντέλο των Barbosa-Bailly.....	37
Πρόβλεψη των Διαρκειών των ΟΔΑΚ.....	37
Τα Δεδομένα Ομιλίας.....	39
Συγχωνεύοντας το φαινόμενο των Παύσεων στον Αλγόριθμο Ανακατανομής.....	40
Αυτόματη Εκμάθηση.....	43
<i>Κατακλείδα</i>	44

<i>ΑΝΑΦΟΡΕΣ - ΒΙΒΛΙΟΓΡΑΦΙΑ</i>	46
<i>ΠΙΝΑΚΑΣ ΑΓΓΛΙΚΩΝ ΟΡΩΝ</i>	52
<i>ΠΙΝΑΚΑΣ ΣΥΝΤΜΗΣΕΩΝ</i>	59

ΠΕΡΙΛΗΨΗ

Σκοπός της εργασίας αυτής είναι η μελέτη της χρονικής δομής του λόγου και ειδικότερα των παύσεων, με στόχο τη μοντελοποίηση τους και την κατασκευή αλγορίθμων, ικανών να παράγουν παύσεις και χρονικές διάρκειες, κατά τρόπο που να προσεγγίζουν αυτές που εμφανίζονται κατά την ανθρώπινη ομιλία. Οι αλγόριθμοι αυτοί, εφαρμόζονται σε συστήματα μετατροπής κειμένου σε ομιλία. Το αποτέλεσμα της εφαρμογής τους είναι το πιο φυσικό άκουσμα της τεχνητής ομιλίας, που κατά συνέπεια οδηγεί στο να γίνεται πιο κατανοητή.

Οι παύσεις, κατηγοριοποιούνται, είτε φυσικά και γλωσσολογικά, είτε βάσει ψυχογλωσσολογικών κριτηρίων. Στην εμφάνιση των παύσεων μπορεί να οδηγήσουν, παράγοντες που προέρχονται από την ίδια τη φυσιολογία του ανθρώπου, τα προσωπικά ιδιαίτερα χαρακτηριστικά κάθε ανθρώπου, το ρυθμικό χαρακτήρα της ομιλίας, τη γνωστική επεξεργασία που προηγείται της άρθρωσης, την πολυπλοκότητα του αντικειμένου που περιγράφεται μέσω της ομιλίας, ή ακόμη και από την ψυχολογική κατάσταση που μπορεί να βρίσκεται κάποιος (π.χ. άγχος, κούραση κτλ.). Η εμφάνιση των παύσεων, ο ρυθμός ομιλίας και λάθη που συμβαίνουν σε αυτή, εξαρτώνται και από την ευφράδεια κάθε ομιλητή. Η κατανομή της ευφράδειας των ομιλητών, για ένα δείγμα πληθυσμού, έχει δειχθεί ότι είναι κανονική (Γκαουσιανή).

Για την κατασκευή ενός συστήματος μετατροπής κειμένου σε ομιλία, θα πρέπει να ληφθούν υπόψη τουλάχιστον οι βασικότεροι από τους παράγοντες που συμβάλλουν στη σωστή δόμηση της διάρκειας του λόγου. Ένα “εργαλείο” με το οποίο μπορούμε να εντοπίσουμε σε ένα κείμενο τη θέση κάποιων παύσεων είναι οι αποκαλούμενες *εκτελεστικές δομές (performance structures)*. Η δομή της πρότασης αναλύεται σε “φυσικές” ομάδες λέξεων, μεταξύ των οποίων τείνουν να εμφανίζονται οι παύσεις. Οι ομάδες αυτές, στατιστικά, σημασιολογικά και συντακτικά χαρακτηρίζονται από ισχυρή *ενότητα (cohesion)* και ψυχογλωσσολογικά κατατάσσονται κάτω από την έννοια της εκτελεστικής δομής. Υπάρχουν ενδείξεις ότι υπάρχει άμεση σχέση μεταξύ εκτελεστικών και προσωδιακών δομών ενώ η συσχέτισή τους με τις συντακτικές δεν είναι ισχυρή.

Διάφοροι αλγόριθμοι έχουν επινοηθεί για την παραγωγή χρονικών διαρκειών και την πρόβλεψη της θέσης των παύσεων. Γενικά, τα μοντέλα που έχουν αναπτυχθεί βάσει των αλγορίθμων αυτών, κατατάσσονται σε *νομο-κυβερνώμενα (rule-governed)* και σε *στατιστικά (statistical)*. Η γενική ιδέα, λειτουργίας των νομο-κυβερνώμενων μοντέλων έχει να κάνει με την εφαρμογή ενός αριθμού κανόνων στα δεδομένα μας. Στα στατιστικά μοντέλα κυριαρχεί η εκμετάλλευση των στατιστικών δεδομένων με τα οποία τροφοδοτούμε το σύστημα, και τα οποία σχετίζονται με τις διάρκειες διαφόρων τμημάτων του λόγου, ανάλογα με τη σχετική τους θέση σε μία φράση, τον τονισμό, κ.α.

Στα πλαίσια της μελέτης αλγορίθμων, παρουσιάζονται αναλυτικά το μοντέλο z-score και το μοντέλο που χρησιμοποιεί μια στοχαστική γραμματική ανεπηρέαστη από τα συμφραζόμενα (ΣΑΑΣΓ). Και τα δυο αυτά μοντέλα χρησιμοποιούν δομές νευρωνικών δικτύων (στατιστικό μοντέλο), τα οποία αρχικά εκπαιδεύονται με κάποια δεδομένα

ομιλίας. Σε συνδυασμό, στο μοντέλο z-score χρησιμοποιείται ο αλγόριθμος ανακατανομής, ενώ στο μοντέλο ΣΑΑΣΓ χρησιμοποιείται ο μέσα-έξω αλγόριθμος. Τα αποτελέσματα των αλγορίθμων αυτών είναι ιδιαίτερα ικανοποιητικά.

ΕΙΣΑΓΩΓΗ

Η σύνθεση ομιλίας που μοιάζει στο άκουσμα της φυσική, απαιτεί στενό έλεγχο επάνω στη χρονική δομή (temporal structure) της ροής του λόγου. Αυτό περιλαμβάνει ένα πλήρως προγνωστικό σχήμα για τη δομή της διάρκειας και κυρίως την παράταση των τελευταίων συλλαβών των λεξημάτων (lexemes) όπως επίσης και για τη δομή των παύσεων σε μια φράση. Στην εργασία αυτή, παρουσιάζεται μια περιγραφή της χρονικής δομής και μια σύνοψη των πολυάριθμων παραγόντων που την τροποποιούν. Στο δεύτερο μέρος, γίνεται μια εισαγωγή σε προγνωστικά σχήματα για τη χρονική δομή του λόγου ("εκτελεστικές δομές"), κι επιδεικνύεται η δυνατότητά τους να χαρακτηρίζουν ολόκληρη τη προσωδιακή δομή του λόγου.

Η σύνθεση ομιλίας από κείμενο απαιτεί τη μετατροπή των συλλαβών σε ακούσματα ομιλίας. Για να μοιάζει αληθινός στο ανθρώπινο αυτί, ο τεχνητός λόγος πρέπει να έχει φυσική ακουστική, διαμόρφωση της άρθρωσης (vocal inflection) (π.χ. σύμφωνα με τις περιστάσεις ή την ψυχική διάθεση), ρυθμό και τονισμό. Με άλλα λόγια, η σύνθεση ομιλίας απαιτεί προσωδιακά χαρακτηριστικά. Χρονικά φαινόμενα, όπως οι παύσεις, παράταση στην άρθρωση των συλλαβών και συνολική δομή χρονισμού, αποτελούν ένα βασικό μέρος από αυτές τις προσωδιακές πλευρές του λόγου. Μερικά τέτοια φαινόμενα (όπως παράταση στην άρθρωση των συλλαβών στο τέλος των φράσεων) εφαρμόζονται σε συσκευές σύνθεσης ομιλίας για αρκετό καιρό, ενώ άλλα (όπως παύσεις και ρυθμός στην ομιλία) δεν είναι ακόμη ευρέως χρησιμοποιούμενα. Μια πλήρως ανθρωποφανής εφαρμογή αυτών των χρονικών χαρακτηριστικών του λόγου, αναμένεται να οδηγήσει σε πολύ σημαντικές περαιτέρω βελτιώσεις στη σύνθεση ομιλίας αποδίδοντάς την με ακόμη μεγαλύτερη "ευφράδεια", περισσότερο ανθρωποφανή, και πιθανώς ακόμη λίγο πιο κατανοητή.

Προικίζοντας τη σύνθεση ομιλίας με προσωδιακές παραμέτρους, σημαίνει ότι ο τόνος της ομιλίας, ο τονισμός, το μήκος των συλλαβών και ο ρυθμός παραγωγής της ομιλίας θα πρέπει να παράγονται με βάση ένα δεδομένο κείμενο. Είναι λοιπόν σημαντικό να λάβουμε υπόψη μας πως τα χρονικά φαινόμενα προκύπτουν στον ανθρώπινο λόγο, και πως αυτά σχετίζονται με το κείμενο από το οποίο παράγονται. Σε επίπεδο ακουστικού σήματος, οι υψηλής στάθμης χρονικές παράμετροι μεταφράζονται όχι μόνο σε αντίστοιχες χαμηλής στάθμης χρονικές μεταβολές, αλλά επίσης σε τροποποιήσεις της θεμελιώδους συχνότητας και της έντασης. Κάτι άλλο που θα πρέπει να λάβουμε υπόψη μας λοιπόν, είναι η σχέση μεταξύ των χρονικών φαινομένων, που θεωρήσαμε ως δεδομένη σε προσωδιακό επίπεδο, και της ακριβούς ακουστικής εφαρμογής αυτών των φαινομένων.

Όπως θα δούμε και αργότερα στο κεφάλαιο, οι χρονικές διάρκειες των τμημάτων ή των συλλαβών και τα φαινόμενα των παύσεων δείχνουν να είναι δυο όψεις του ίδιου νομίσματος. Θα ξεκινήσουμε περιγράφοντας τα φαινόμενα των παύσεων, και στο δεύτερο μέρος του κεφαλαίου, θα δείξουμε πώς τα φαινόμενα των παύσεων μπορούμε να τα δούμε σα μια εξελιγμένη θεωρία που σχετίζεται με τα χρονικά φαινόμενα σε μια γενική, προσωδιακή διάρθρωση του λόγου.

ΠΑΥΣΕΙΣ

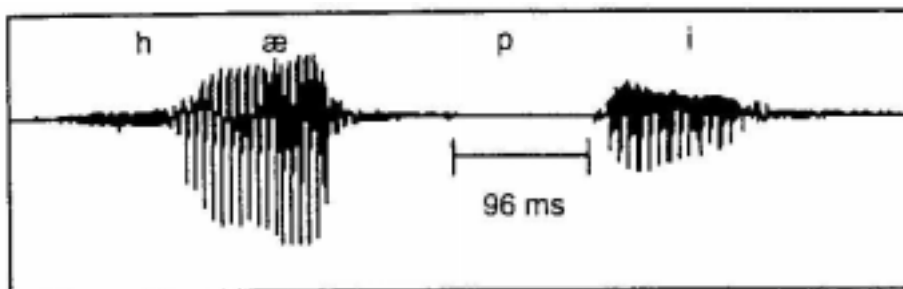
Η Κατηγοριοποίηση των Παύσεων

Από περιγραφικής άποψης, δύο κατηγοριοποιήσεις των παύσεων γενικά χρησιμοποιούνται. Η πρώτη είναι μια φυσική και γλωσσολογική κατηγοριοποίηση, και η δεύτερη είναι μια ψυχολογική και ψυχογλωσσολογική κατηγοριοποίηση.

Η Φυσική και Γλωσσολογική Κατηγοριοποίηση

Σύμφωνα με τον παραδοσιακό γλωσσολογικό ορισμό, η κανονική ροή του λόγου θεωρείται ότι διακόπτεται από μια *φυσική παύση* όποτε μια σύντομη σιγή μπορεί να παρατηρηθεί στο ακουστικό σήμα (π.χ. ένα τμήμα χωρίς σημαντικό πλάτος). Ποια ακριβώς χρονική διάρκεια της σιγής θεωρείται αρκετή ώστε να αποτελέσει παύση, εξαρτάται από το γλωσσικό της πλαίσιο (βλ. Dechert & Raupach, 1980):

- *Ενδοτμηματικές παύσεις (Intra-segmental pauses)* είναι εκείνες που σχετίζονται με το κλείσιμο της φωνητικής οδού στην κανονική παραγωγή της ομιλίας. Παράδειγμα: Στην αγγλική λέξη "happy" (σχήμα 1), η παύση στο χρόνο εκκίνησης ομιλίας (VOT: Voice Onset Time) για το σύμφωνο /p/ αντιστοιχεί σε μια σιγή διάρκειας 96ms.

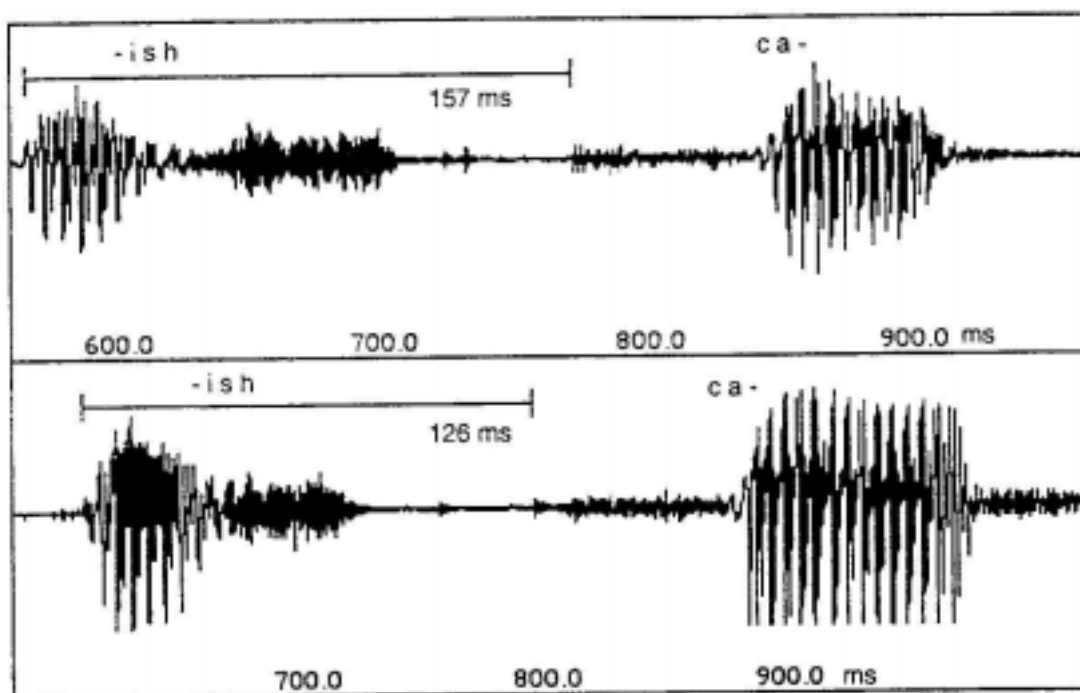


Σχήμα 1. Το ακουστικό σήμα της αγγλικής λέξεως "happy" παρουσιάζει μία ενδοτμηματική παύση.

- *Διαλεξικές παύσεις (Inter-lexical pauses)* είναι εκείνες που μπορεί να εμφανίζονται μεταξύ δυο λέξεων. Αποτελούν τον πρώτο χωρισμό του λόγου σε τμήματα, ή την διατύπωση που πιθανώς θα εξυπηρετήσει την κατανόηση της ειπωθείσας φράσης. Ένα παράδειγμα είναι η διαφοροποίηση μεταξύ "A Turkish (carpet salesman)" - που σημαίνει: ένας Τούρκος πωλητής χαλιών - και "(a Turkish carpet) salesman" - που σημαίνει: ένας πωλητής τουρκικών χαλιών - (σχήμα 2).

Σε ακουστικό επίπεδο, η διαφοροποίηση αυτών των δύο προτάσεων επιτυγχάνεται με μεταβολές στη μελωδία και μεταβολές στη χρονική διάρκεια. Στην πρώτη πρόταση, η χρονική διάρκεια του συνδυασμού των τελευταίων συλλαβών της λέξης "Turkish" και η διαλεξική παύση ισοδυναμούν με 157ms, ενώ στη δεύτερη πρόταση το ίδιο ακουστικό τμήμα διαρκεί 126ms.

Είναι σημαντικό να σημειωθεί πως ότι ακούγεται στο αυτί σαν "φαινόμενο παύσης" στην πραγματικότητα μεταφράζεται σε επίπεδο σήματος σε πολύπλοκο χειρισμό της χρονικής διάρκειας και της θεμελιώδους συχνότητας που εκτείνεται στην τελευταία συλλαβή και στην παύση που προσαρτείται. Η παραδοσιακή γλωσσολογική και ψυχολογολογική έννοια ενός μονοδιάστατου "φαινομένου παύσης" έχει βελτιωθεί από τις πιο σύγχρονες έρευνες. Σήμερα αναγνωρίζεται ότι αυτά τα λεγόμενα "Χρονικά" προσωδιακά φαινόμενα είναι στην πραγματικότητα εγγενώς πολύπλοκα, πολυμεταβλητά και υπερτιμηματικά, από τη στιγμή που γενικά εκτείνονται σε έναν αριθμό τμημάτων ή συλλαβών, και εμπλέκουν την αλληλεπίδραση μεταξύ ενός αριθμού φυσικών παραμέτρων.



Σχήμα 2. Ακουστικά σήματα που αντιστοιχούν στην προφορά της φράσης "a Turkish (carpet salesman)" (επάνω) και "a (Turkish carpet) salesman" (κάτω).

Μια Ψυχολογολογική Κατηγοριοποίηση

Ακόμη κι αν η συζήτηση είναι περιορισμένη αυστηρά στα χρονικά φαινόμενα, οι παύσεις που γίνονται αντιληπτές δεν είναι στην πραγματικότητα το ισοδύναμο των φυσικών παύσεων. Αυτό, οφείλεται στο νόμο της αντίληψης, ο οποίος είναι πολύ γνωστός στους φυσιολόγους: Είτε στο πεδίο της όρασης, είτε στις ακοής, είτε στις αφής, το κατώφλι για την ουσιαστική αντίληψη βρίσκεται πάνω από το πραγματικό ερέθισμα. Μάλιστα, τα πλάτη των καμπυλών που μετρήθηκαν με λεπτομέρεια σε τεστ αντίληψης (π.χ. αποκρίσεις σε λεπτώς διαφοροποιούμενα ερεθίσματα) διαφέρουν συστηματικά από τις καμπύλες που μετρήθηκαν απευθείας στο φυσικό ερέθισμα.

Αναφερόμενοι στα χρονικά φαινόμενα, ένας αριθμός από ειδικούς στην ψυχοακουστική έχουν καταγράψει τη σημαντικότητα αυτών των όψεων της αντίληψης για την κατανοητικότητα που έχουμε στην επεξεργασία του λόγου (βλ. Zwicker & Feldkeller 1981, Botte et al. 1989). Μια σωστή αντίληψη του συγκροτημένου λόγου απαιτεί ικανότητα επεξεργασίας 15 έως 30 περίπου διαφορετικών ήχων ("φωνημάτων") ανά δευτερόλεπτο. Διακυμάνσεις που υπερβαίνουν το 20% της χρονικής διάρκειας αυτών των ήχων έχει βρεθεί να σχετίζονται στον τομέα της αντίληψης με την αλάθητη κατανόηση της ροής του λόγου.

Μερικές παύσεις είναι ευκολότερα αντιληπτές από άλλες, και γενικά, τέτοιες παύσεις φαίνεται να υποστηρίζουν συγκεκριμένες λειτουργίες μέσα στο μήνυμα, όπως παραδείγματος χάριν γραμματικές λειτουργίες, σημασιολογική εστίαση, έκφραση δισταγμού κ.τ.λ. Επίσης, οι παύσεις γίνονται ευκολότερα αντιληπτές αν η χρονική τους διάρκεια είναι γύρω στα 200 - 250ms. Αυτό φαίνεται να είναι το ακουστικό κατώφλι για την αντίληψη των παύσεων (Goldman-Eisler 1968, Grosjean & Deschamps 1975). Σε γλώσσες που έχουν παρατηρηθεί συστηματικές παύσεις, μπορούμε να διακρίνουμε δύο τύπους παύσεων, τις λεγόμενες "σιωπηρές" (silent) και τις "γεμάτες" (filled):

- Οι *Σιωπηρές παύσεις* (*silent pauses*), αντιστοιχούν στο σιωπηρό τμήμα που μπορεί να γίνει αντιληπτό στο σήμα της ομιλίας. Τέτοιες παύσεις μπορεί να παράγονται σε συνδυασμό με μία εισπνοή, κατάποση, με οποιοδήποτε λάρυγγο-φωνητικό αντανακλαστικό ή μια σιωπηρή εκπνοή.
- Οι *Γεμάτες παύσεις* (*filled pauses*), αντιστοιχούν σε ένα φωνούμενο τμήμα (voiced section) που μπορεί να γίνει αντιληπτό στο σήμα της ομιλίας. Οι περισσότερες γεμάτες παύσεις σε γλώσσες όπως τα Αγγλικά και τα Γαλλικά αποτελούν τη μακρόσυρτη εκφορά των φωνηέντων, τις επαναλήψεις φράσεων, λέξεων, συλλαβών, ήχων και τις λαθεμένες εκκινήσεις στην προσπάθεια εκφοράς του λόγου (Grosjean & Deschamps 1975, Bloodstein 1981).

Γενικά, στους φυσιολογικούς ομιλητές (ομιλητές χωρίς καμία παθολογία στο λόγο), οι σιωπηρές και γεμάτες παύσεις εμφανίζονται ανάμεσα στις λέξεις. Παύσεις των 200ms ή μεγαλύτερες σπάνια παρατηρούνται μέσα σε μια λέξη.

Οι Πηγές Προέλευσης των Παύσεων

Οι θέσεις και οι χρονικές διάρκειες των σιωπηρών και γεμάτων παύσεων υπόκεινται σε δύο τύπους περιορισμών. Κατά πρώτον, εξαρτώνται από τη φυσιολογική πλευρά της δραστηριότητας του μηχανισμού ομιλίας, και κατά δεύτερον αντικατοπτρίζουν τη γνωστική επεξεργασία.

Η Χρονική Δομή του Λόγου και οι Περιορισμοί που υπεισέρχονται σε αυτή.

Όπως οποιαδήποτε άλλη ανθρώπινη δραστηριότητα, η παραγωγή του λόγου δεν μπορεί να ασκείται ακατάπαυστα. Διακοπές είναι απαραίτητες, από τη στιγμή που κάποιος άλλος μηχανισμός δραστηριότητας εξελίσσεται παράλληλα και αλληλεπιδρά με άλλες διαφορετικού τύπου μηχανισμούς λειτουργίας. Εφόσον η λειτουργία του μηχανισμού της κίνησης είναι γενικώς αναγνωρισμένο ότι εκτελείται σύμφωνα με τρία διαδοχικά στάδια (προσχεδιασμός, εκτέλεση, παύση), οποιαδήποτε διακοπή μιας δραστηριότητας για να γίνει αποδεκτή μια άλλη, συνήθως πραγματοποιείται κατά το στάδιο της παύσης.

Ακόμη, από άποψη φυσιολογίας αναπόφευκτες παύσεις συνήθως προκύπτουν κατά την αναπνοή στη φάση της εισπνοής, από τη στιγμή που η φωνητική δραστηριότητα είναι περίπλοκα συνδεδεμένη με την αναπνοή. Καταλήγοντας, η παραγωγή ομιλίας μπορεί να θεωρηθεί ως ρυθμική δραστηριότητα, όπου οι ομάδες των λέξεων παράγονται με συγκεκριμένο ρυθμό. Όλα αυτά τα γεγονότα πιθανώς συμβάλλουν στην παραγωγή παύσεων κατά τακτά χρονικά διαστήματα.

Επιπρόσθετα, ένας αριθμός από άλλες παραμέτρους μπορεί να δειχθεί ότι επηρεάζει την εμφάνιση των παύσεων:

Ανεξάρτητοι φυσιολογικοί παράγοντες: Από τη στιγμή που η δραστηριότητα του μηχανισμού της ομιλίας είναι κατά μεγάλο μέρος μια ανεξάρτητη δραστηριότητα, η εμφάνιση των παύσεων εξαρτάται κατά πολύ από τον συγκεκριμένο ομιλητή. Γενικά, ασθενής αναπνοή, χαμηλός μυϊκός τόνος, και αργός ρυθμός άρθρωσης, συνδέονται με μεγαλύτερο αριθμό παύσεων, ενώ αντίθετα συμβαίνει όταν ο ρυθμός άρθρωσης είναι γρήγορος και υπάρχει καλή αναπνευστική χωρητικότητα.

Χρονικοί περιορισμοί και η έννοια του ρυθμικού διαστήματος: Οι παύσεις τείνουν να προκύπτουν μεταξύ ρυθμικών ομάδων. Ο Dauer δείχνει ότι τα δεδομένα από μια συγκριτική μελέτη γλωσσών με χρονισμό στη συλλαβή (syllable-timed) και γλώσσες δυναμικού τονισμού (stress-timed) (π.χ. Γαλλικά και Αγγλικά αντίστοιχα) υποστηρίζουν την υπόθεση γενικών χαρακτηριστικών στη ρυθμική δομή (Dauer 1983). Σύμφωνα με αυτή την υπόθεση, ο προσχεδιασμός της ομιλίας βασίζεται σε μια φυσιολογική ρυθμιστική μονάδα που επιτρέπει περίπου δύο "πράξεις" ανά δευτερόλεπτο.

Στοιχεία από χρονομετρήσεις της γενικής συμπεριφοράς του μηχανισμού κίνησης οδηγούν στην υποστήριξη της ιδέας ότι υπάρχει κάποιο βιολογικό ρολόι το οποίο παρακολουθεί κάθε ρυθμική δραστηριότητα. Οι χρονικές διάρκειες κάποιων τμημάτων του λόγου δείχνουν δύο χαρακτηριστικά τα οποία υποδεικνύουν την ύπαρξη ενός εσωτερικού ωρολογιακού μηχανισμού: κατά πρώτον, ορισμένα τμήματα του λόγου είναι αρκετά συχνά, και κατά δεύτερον οι διακυμάνσεις για κάποιες διάρκειες (π.χ. η χρονική διάρκεια μεταξύ δύο μεταπτώσεων ΣΦ – CV transitions - στα Γαλλικά) δείχνουν συχνές διορθώσεις – πάρα πολύ σαν να γίνονται σταθερές ρυθμίσεις, έτσι ώστε να προσεγγίσουν συγκεκριμένα μοντέλα χρονικών πλαισίων. Σύμφωνα με τους Barbosa και Bailly (1993), αυτό το εσωτερικό ρολόι είναι "μια λειτουργία που κρατά το ρυθμό για τον συγχρονισμό των σημάτων διέγερσης που μεταδίδονται στους μυς". Προγραμματίζει

τη διάρκεια των ρυθμικών μονάδων, και η τακτική λειτουργία του ρολογιού συντηρείται μέσω των παύσεων. Με άλλα λόγια, οι χρονικές διάρκειες των παύσεων περιορίζονται σε έναν αριθμό από μονάδες ρολογιού. Προφανώς, αυτή η έννοια της τακτικότητας είναι πολύ ελκυστική, ιδιαίτερα στο πλαίσιο μιας νομοκρατούμενης φυσικής σύνθεσης ομιλίας. Ωστόσο, ο Keller (1990) μας θυμίζει ότι σε βιολογικό επίπεδο, δεν υπάρχει καμία εδραιωμένη ζεύξη μεταξύ νευρωνικού παλμού μιας κεντρικής προέλευσης και χρονισμού ομιλίας. Θα ήταν μάλιστα ασφαλέστερο να ερμηνεύσουμε τα υπάρχοντα στοιχεία λέγοντας ότι πιθανώς να υπάρχει κάποια εσωτερική ρύθμιση του χρόνου, αλλά ο ακριβής φυσιολογικός μηχανισμός είναι ακόμα άγνωστος.

Ο Dauer παρέχει ακόμη ένα παρόμοιο επιχειρήμα. Σύμφωνα με αυτόν τον συγγραφέα, η τακτικότητα στις γλώσσες δυναμικού τονισμού εκδηλώνεται στο διάστημα τονισμού, όπου το μέσο διάστημα κυμαίνεται από 1,9 μέχρι 2,3 τονισμούς ανά δευτερόλεπτο. Αυτοί οι τονισμοί προκύπτουν με πολύ μεγαλύτερη τακτικότητα απ'ότι οι συλλαβές, όπου το μέσο διάστημα μεταξύ των συλλαβών κυμαίνεται από 4,5 έως 8 συλλαβές ανά δευτερόλεπτο. Όπως το βλέπει ο Dauer, η ρυθμική ομαδοποίηση εφαρμόζεται σε όλες τις γλώσσες και αλληλεπιδρά με τη δομή των συλλαβών τους, με τους κανόνες της πτώσης των φωνηέντων (vowel reduction) και με την έμφαση που μπορεί να δοθεί στις λέξεις. Βάσει αυτών των εννοιών για τη ρύθμιση της διάρκειας, μπορεί να προταθεί ότι η ρυθμική ομαδοποίηση του λόγου διευκολύνει την εμφάνιση των παύσεων σε συγκεκριμένα διαστήματα.

Η Λειτουργία των Διαλεξικών Παύσεων που γίνονται αντιληπτές (μεγαλύτερες από 200ms).

Πέρα από τις πηγές προέλευσης των παύσεων που σχετίζονται σε μεγάλο βαθμό με τη φυσιολογία όπως είδαμε παραπάνω, είναι επίσης δυνατό να αναγνωρίσουμε και κάποιες που σχετίζονται με τη γνωστική δραστηριότητα. Μολονότι αυτές οι τελευταίες παύσεις είναι λιγότερο συχνές, ακολουθούν κάποιο πρότυπο και άρα μπορούν να χρησιμοποιηθούν για την παραγωγή συνθετικής ομιλίας που να ακούγεται φυσική.

Οι παύσεις ως αντικατοπτρισμός γνωστικής δραστηριότητας: Σύμφωνα με τους Goldman-Eisler (1968-1972), μια παύση αντικατοπτρίζει εξωτερικά γνωστικές διεργασίες που εμπλέκονται στην παραγωγή του λόγου. Με αυτή την έννοια, οι παύσεις προσφέρουν πρόσθετο χρόνο κατά τον οποίο μπορεί να προσχεδιαστεί και να προγραμματιστεί το τελικό προϊόν της ομιλίας. Αυτή η υπόθεση εξηγεί μερικές κοινές παρατηρήσεις, όπως όταν ένας ομιλητής σκέφτεται πολύ χρόνο μέχρι να αποδώσει μια πολύ γρήγορη, καθαρή και καλά δομημένη απάντηση. Από την άλλη πλευρά, μπορεί επίσης να παρατηρηθεί ότι μερικές φορές, ο ομιλητής ξεκινάει να απαντάει αμέσως, και μετά σταματά ή ξαναγυρίζει για να αποσαφηνίσει το μήνυμα που θέλει να μεταδώσει. Σε αυτή την περίπτωση, η υπόθεση λειπει ότι "ο λόγος πέρασε μπροστά από τη γνωστική δραστηριότητα" και ότι η παύση αντικατοπτρίζει το χρόνο που χρειάζεται η διαδικασία γνωστικού προσχεδιασμού για να τον προφτάσει.

Η υπόθεση των Goldman-Eisler προβλέπει ακόμη τις διαφορές που συνήθως παρατηρούνται μεταξύ του αυθόρμητου λόγου και του λόγου που προέρχεται από

ανάγνωση κειμένου. Ο αυθόρμητος λόγος είναι πολύ πιο επιρρεπής στις παύσεις γνωστικής προέλευσης απ'ότι ο λόγος που προέρχεται από ανάγνωση κειμένου. Υπό αυτή την έννοια, όλα αυτά δεν αποτελούν απλά ενοχλητικές μεταβλητές. Από τη στιγμή που πιθανώς αντικατοπτρίζουν γνωστική δραστηριότητα, συμβάλλουν στο να καταλάβει ο ακροατής πώς ο ομιλητής δομεί τη φράση του. Και από τη στιγμή που οι παύσεις φανερά συμμετέχουν ώστε να αποδίδεται πιο κατανοητικά η ανθρώπινη επικοινωνία, μπορεί να ειπωθεί ότι κάποιοι μελλοντικοί εξελιγμένοι μηχανισμοί αναγνώρισης ομιλίας μπορούν να επωφεληθούν από αυτού του τύπου τις πληροφορίες για να βελτιώσουν τις επιδόσεις τους στην αναγνώριση.

Παύσεις που συμπεριφέρονται σαν "φάροι" για τις φράσεις: Ένας αριθμός ψυχολογολογικών ερευνών έχει δώσει μερικές καθαρές ενδείξεις στο πώς οι παύσεις μοντελοποιούνται. Σε μια μελέτη για τα Γαλλικά, για παράδειγμα, οι Grosjean και Deschamps (1975) έδειξαν ότι όσο πιο πολύπλοκη είναι η προς εκτέλεση επικοινωνιακή εργασία, τόσο μεγαλύτερος είναι και ο αριθμός των παύσεων (Πίνακας 1). Όταν γίνεται περιγραφή ενός κινουμένου σχεδίου, οι παύσεις είναι μεγαλύτερες και συχνότερες απ'ότι να απαντά κάποιος σε μια συνέντευξη (Grosjean & Deschamps 1975).

	<i>Μέση χρονική διάρκεια των σιωπηρών παύσεων</i>
Περιγραφή ενός κινουμένου σχεδίου:	1320 ms
Συνέντευξη:	520 ms

Πίνακας 1. Χρονική διάρκεια παύσεων ανάλογα με τη γλωσσική εργασία.

Και στις δυο αυτές περιπτώσεις επικοινωνίας, η χρονική διάρκεια αυτών των παύσεων είναι αρκετά εντυπωσιακή (1320 και 520ms), ειδικότερα όταν αυτό το δει κανείς σε σχέση με τη χρονική διάρκεια μιας τυπικής συλλαβής (περίπου 200-300ms), ή ενός συνήθους φωνήεντος (περίπου 100-150ms) (O'Shaughnessy 1981). Με άλλα λόγια, οι παύσεις, μπορεί να τοποθετούνται σα "φάροι" στο λόγο, με στόχο να δομήσουν ολόκληρη τη φράση για τον ομιλητή και τον ακροατή. Υποδιαιώνοντας το λόγο σε μικρότερα τμήματα, οι παύσεις πιθανότατα συμβάλλουν κατά πολύ στο να γίνει πιο κατανοητός.

Επιπρόσθετα, υπάρχει μια σχέση μεταξύ της χρονικής διάρκειας και της συχνότητας των παύσεων από τη μια, και της ιεραρχίας των συντακτικών συστατικών από την άλλη. Για παράδειγμα, κατά τη φωναχτή ανάγνωση, υπάρχει μια τάση οι παύσεις να είναι μεγαλύτερες, όσο πιο περίπλοκο είναι το συντακτικό πλαίσιο. Ακόμα, θα δειχθεί παρακάτω ότι ο χωρισμός σε χρονικά τμήματα δεν είναι πραγματικά ισοδύναμος με τη συντακτική δομή των φράσεων.

Περιστασιακοί περιορισμοί: Τελικά, είναι επίσης σημαντικό να λαμβάνεται υπόψη και το πλαίσιο των περιστάσεων, γιατί όταν ο ομιλητής πιέζεται χρονικά μπορεί να ευνοούνται ή να δυσχεραίνονται οι εκφραστικές του ικανότητες (για παράδειγμα, είναι δύσκολο να αποδώσεις μια πολύ σημαντική πληροφορία γρήγορα σε ένα δωμάτιο με

θόρυβο). Όσο πιο δύσκολη είναι η κατάσταση όπου διεξάγεται η επικοινωνία, τόσο περισσότερες παύσεις, δισταγμοί, και τραυλίσματα είναι δυνατό να συμβούν.

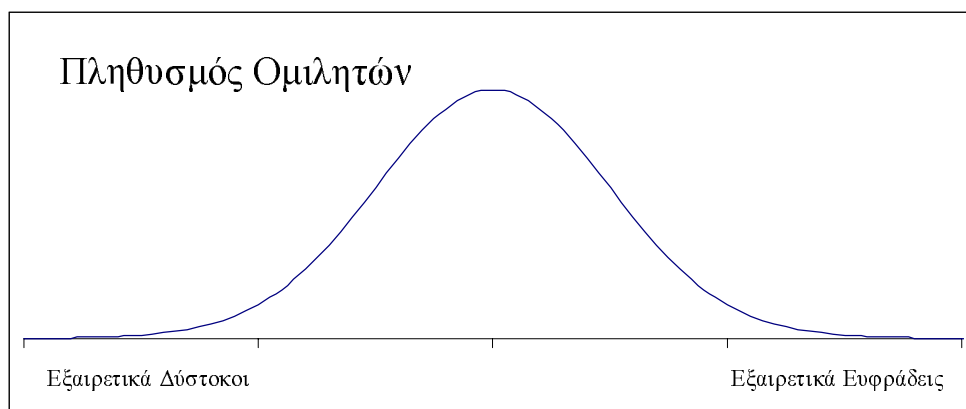
Η παραγωγή ομιλίας είναι μια ρυθμική δραστηριότητα όπως επίσης και μια υποβόσκουσα γνωστική διεργασία, με αποτέλεσμα οι ομιλητές να παράγουν τις παύσεις αυθόρμητα. Το γεγονός αυτό, όπως και οι διασυνδέσεις μεταξύ των χρονικών και συντακτικών δομών, μας οδηγούν να πιστεύουμε ότι οι αντιληπτές διαλεξικές παύσεις πρέπει να είναι προβλέψιμες και αλγοριθμικά πραγματοποιήσιμες στη σύνθεση ομιλίας.

Οι Παύσεις και η Έννοια της Λεκτικής Ευφράδειας

Τα φαινόμενα που περιγράφονται εδώ μπορούν να συμπεριληφθούν στη συνολική έννοια της *λεκτικής ευφράδειας*: ένας ομιλητής λέμε ότι έχει *ευφράδεια λόγου* όταν ομιλεί εύκολα, με ομαλές ενάρξεις και μεταβάσεις, και με σχετικά γρήγορο ρυθμό (Pfuwadel 1986). Αντίθετα, ένας ομιλητής λέμε ότι δεν έχει ευφράδεια λόγου, όταν διστάζει, παράγει παύσεις εκεί που δε θα έπρεπε, και κάνει πολλά λάθη ομιλίας. Είναι ένα καλά καταγεγραμμένο γεγονός της ανθρώπινης επικοινωνιακής συμπεριφοράς ότι οι ομιλητές κυμαίνονται από εξαιρετικά ευφράδεις έως καθόλου. Ακόμη και χωρίς να έχουν κάποια νευρολογική διαταραχή, μερικοί ομιλητές είναι δυσνόητοι λόγω υπερβολικών δισταγμών, παύσεων και λαθών ομιλίας.

Από την Υπερβολική Δυστοκία στην Έκφραση ως την Εξαιρετική Ευφράδεια.

Είναι χρήσιμο να σκεπτόμαστε την ευφράδεια σαν να ήταν κατανεμημένη σε Γκαουσιανή καμπύλη. Στα αριστερά της καμπύλης θα τοποθετούνταν οι υπερβολικά δύσστοκοι ομιλητές, θα ακολουθούσαν διστακτικοί ομιλητές διαφορετικού βαθμού, στη συνέχεια η πλειοψηφία των ομιλητών που είναι λογικά ευφράδεις και στα δεξιά της καμπύλης τέλος, μια μικρή ομάδα υπερβολικά ευφράδων ομιλητών.



Σχήμα 3. Γκαουσιανή καμπύλη κατανομής ομιλητών.

Στον αυθόρμητο λόγο, δεν είναι ασυνήθιστο να ακούμε κάποιον να λει μια λέξη αντί για άλλη ή να μπερδεύει τις συλλαβές μιας λέξης. Αυτά είναι τα λεγόμενα λάθη εκτέλεσης (*performance errors*) ή φυσιολογικές δυστοκίες στην έκφραση (*normal dysfluencies*). Σε αντίθεση με τις παύσεις που είναι αρκετά συχνές, τα λάθη εκτέλεσης είναι μοναδικά. Παράγονται σε συγκεκριμένες στιγμές, σε εξαιρετικές περιπτώσεις, και γενικά δε θα παραχθούν ξανά κάτω από άλλες συνθήκες. Πάντως, η ασυνήθιστη εμφάνισή τους, όπως και η συχνή διόρθωση τέτοιου τύπου λαθών, δεν αφήνουν καμία αμφιβολία για το τι ήθελε να πει ο ομιλητής. Σε αντίθεση με τα λάθη που παράγονται από παιδιά ή από ανθρώπους που μαθαίνουν κάποια ξένη γλώσσα, η "ομιλητική τους ικανότητα" δεν αμφισβητείται.

Οι διαταραχές της ομιλίας στα αριστερά της καμπύλης περιλαμβάνουν επίσης και διάφορους βαθμούς τραυλίσματος. Σε αυτές τις περιπτώσεις, η χρονική δομή του λόγου αποκλίνει από το αναμενόμενο πρότυπο: η συχνότητα, η χρονική διάρκεια και οι θέσεις των παύσεων δεν είναι φυσιολογικές. Όταν ένας ομιλητής που δεν τραυλίζει διστάζει, οι (σιωπηρές/γεμάτες) παύσεις γενικά βρίσκονται στα συντακτικά ή προσωδιακά πλαίσια (Zellner-Bechel 1992). Ένας σχεδόν παθολογικά ή παθολογικά τραυλός, από την άλλη, παράγει δυστοκίες που συχνά απέχουν πολύ από τα συντακτικά και προσωδιακά πλαίσια (Starkweather 1987). Μια λαθεμένη κατάτμηση τείνει να διαταράσσει τον ακροατή, μερικές φορές πάρα πολύ, ενώ μια φυσιολογική κατάτμηση δύσκολα γίνεται συνειδητά αντιληπτή. Αυτό μας επιτρέπει να συμπεράνουμε ότι η μη φυσιολογική τοποθέτηση των παύσεων μπορεί να είναι τόσο καταστροφική στην αποκωδικοποίηση μιας φράσης προκειμένου να γίνει αντιληπτή, όσο η κανονική χρήση των παύσεων μπορεί να είναι χρήσιμη για την κατανόησή της.

Ακόμη και ομιλητές με πολύ μεγάλη ευφράδεια μεταβάλλουν τη χρονική δομή του λόγου τους. Ο Miller et al. έχει δείξει ότι στον αυθόρμητο λόγο, ένας ομιλητής μπορεί να μεταβάλλει το ρυθμό σημαντικά μέσα στην ίδια φράση, όπου οι πιο εκτεταμένες μεταβολές του ρυθμού επιτυγχάνονται με τροποποιήσεις του αριθμού και της διάρκειας των παύσεων (Miller et al. 1984). Ακόμη, οι μεταβολές του ρυθμού σχετίζονται με τη χρήση παραλλαγών στους ήχους της ομιλίας, από τη στιγμή που σε γρήγορο ρυθμό ομιλίας, οι ήχοι (ή τα αλλόφωνα –“allophones”) αποκλίνουν περισσότερο από τις "ιδανικές" φωνητικές παραλλαγές ενός δεδομένου φωνήματος απ'ότι σε πιο αργό ρυθμό (Lacheret-Dujour 1991).

Αυτός ο χαρακτηρισμός της ευφράδειας στη ροή της ομιλίας θέτει μια ξεχωριστή πρόκληση στη σύνθεση και αναγνώριση ομιλίας. Η τεχνολογία τώρα δεν εκμεταλλεύεται την εμφάνιση των παύσεων και των λαθών ομιλίας. Στο μέλλον, μπορεί να αποτελέσει ενδιαφέρουσα πρόταση η παραγωγή προτυποποιημένων (κανονικών) παύσεων στη συνθετική ομιλία. Η έννοια της "ευφράδειας" σχετίζεται άμεσα με την πρόβλεψη ενός αρκετά μεγάλου αριθμού φωνητικών παραμέτρων, όπως είναι ο ρυθμός ομιλίας, οι παύσεις, τα λάθη και η πτώση των φωνηέντων. Εάν όλες αυτές οι παράμετροι τεθούν υπό έλεγχο με κάποια λογική συνέπεια στην σύνθεση ομιλίας, η τεχνητή φωνή πιθανότατα θα δίνει μια εντύπωση μεγαλύτερης ευχέρειας κι ευφράδειας του λόγου. Παρόμοια, συσκευές αυτόματης αναγνώρισης ομιλίας που έχουν αναπτυχθεί για να

αντιμετωπίζουν τον αυθόρμητο λόγο, μπορεί αναπόφευκτα να πρέπει να μάθουν να αντιμετωπίζουν και το φαινόμενο των παύσεων, τα λάθη ομιλίας και κάποια μέρα ίσως και τη διόρθωση των λαθών ομιλίας.

Ενδιαφέρουσες Μελέτες - Παραδείγματα.

Πολλά από αυτά που αναφέρθηκαν παραπάνω, σχετικά με το πώς προκύπτουν οι παύσεις, μπορούμε να δούμε πως εφαρμόζονται στην πράξη, και να γίνουν περισσότερο κατανοητά, μελετώντας κάποια πειράματα που διεξήχθησαν σε διάφορες ομάδες ομιλητών.

Σε μια μελέτη της ομιλίας δώδεκα ζευγών φοιτητών, ο πρώτος από το κάθε ζεύγος περιέγραφε ένα διάγραμμα το οποίο μπορούσε να δει, ενώ ο δεύτερος ακούγοντας μόνο την περιγραφή, το σχεδίαζε. Έτσι, ήταν δυνατό να παρατηρηθεί ο τρόπος με τον οποίο προκύπτουν οι παύσεις, συγκριτικά, σε έναν αριθμό ομιλητών. Στον απόσπασμα 1, μπορούμε να διακρίνουμε ένα μέρος της ομιλίας αυτής, η οποία έλαβε χώρα στα Αγγλικά. Στις παρενθέσεις σημειώνεται ο χρόνος που διαρκούν οι παύσεις μεταξύ των λέξεων σε δευτερόλεπτα.

A: Halfway down the page (0.3) draw (0.6) a red (0.4) horizontal line (0.2) of about (0.5) two inches (16) on eh (1.1) the right hand side just above the line (1.9) in black (0.1) write ON (3.2)
B: ON (3.4)
A: above the line (14) draw (0.2) a black (0.65) triangle (1.0) ehm (1.9) a right-angle (0.2) triangle (1.9) starting to the left (0.2) of the red line (1.0) about (0.9) half a centimeter above it (4.0)

Απόσπασμα 1. Απόσπασμα ομιλίας στο οποίο αναγράφονται οι χρόνοι των παύσεων σε sec.

Στο απόσπασμα αυτό, ορίζονται οι ακόλουθοι τύποι παύσεων, όσο αναφορά το σχετικό τους μήκος:

- *Εκτεταμένες παύσεις (extended pauses).* Οι παύσεις αυτές, είναι μεγάλες κι εκτείνονται από 3.2 έως 16sec (που προκύπτουν, στα σημεία όπου ο ομιλητής παραθέτει επαρκείς πληροφορίες, ώστε αυτός που τον ακούει να μπορεί να σχεδιάσει ή να γράψει αυτό που του περιγράφει).
- *Μακρές παύσεις (long pauses).* Στο απόσπασμα αυτό, είναι οι παύσεις που χρονικά ανήκουν στο διάστημα μεταξύ 1.0 έως 1.9sec.
- *Βραχείες παύσεις (short pauses).* Στο απόσπασμα αυτό, είναι οι παύσεις που χρονικά ανήκουν στο διάστημα μεταξύ 0.1 έως 0.6sec.

Τα αποτελέσματα της μελέτης αυτής, όπως και μιας παρόμοιας μελέτης που είχε διεξάγει ο Chafe (1979), συμφωνούν, και οδηγούν και οι δύο στην επιβεβαίωση ότι το μήκος μιας παύσης είναι συνάρτηση του πόσο ο ομιλητής προσχεδιάζει την επόμενη φράση του.

Ο Chafe αναφέρει επίσης μια δουλειά, στην οποία ζητείται από τα υποκείμενά του, να εξιστορήσουν μια σειρά από γεγονότα που είδαν να διαδραματίζονται σε ένα βωβό φιλμ μικρού μήκους. Ο Chafe παρατηρεί “σημαντικούς δισταγμούς” (major hesitations) (σε ισοδυναμία με τον όρο εκτεταμένες παύσεις που χρησιμοποιήσαμε προηγουμένως).

And as he's holding onto the handlebars he takes off with them (1.1) Um – (0.7) then (0.4) uh – (2.1) a girl – on a bicycle (1.15) comes riding towards him . . . in the opposite direction.

Απόσπασμα 2. Chafe - απόσπασμα ομιλίας στο οποίο αναγράφονται οι χρόνοι των παύσεων σε sec.

Τα “Um” και “uh” του αποσπάσματος 2 (όπως και τα “eh” του αποσπάσματος 1), είναι επιμηκύνσεις της προφοράς των αντίστοιχων φωνηέντων, δηλαδή *γεμάτες παύσεις*. Ο συνολικός χρόνος καθυστέρησης μεταξύ του τέλους της πρώτης πρότασης και της αρχής της “a girl on a bicycle”, ήταν 6.25sec. Το γεγονός αυτό, αποτελεί ξεκάθαρη απόδειξη κάποιας σημαντικής και χρονοβόρας γνωστικής επεξεργασίας.

Για να προετοιμάσουμε το έδαφος, μπορεί να είναι χρήσιμο να εξετάσουμε μερικές λεπτομέρειες της χρονικής δομής του λόγου, όπως επίσης και διάφορους προγνωστικούς αλγόριθμους. Σε αυτό θα στραφούμε στη συνέχεια.

Η Δομή της Διάρκειας του Λόγου

Είδαμε ότι ο ρυθμός ομιλίας αποτελεί μια σημαντική πλευρά της λεκτικής ευφράδειας. Ο ρυθμός ομιλίας καθορίζεται όχι μόνο από τις παύσεις, αλλά και από το ρυθμό της άρθρωσης (Grosjean & Daschamps 1975). Μεταβολές στο ρυθμό της άρθρωσης επιβάλλονται από διάφορους παράγοντες, και μάλιστα πρωταρχικά από παραλλαγές της δομής της διάρκειας του λόγου.

Υπάρχει εκτεταμένη βιβλιογραφία σχετικά με αυτό το ερώτημα, αλλά θα αναφερθούμε μονάχα στην πολύ γνωστή μελέτη του Klatt (1976), μια και η μελέτη του ήταν προσανατολισμένη προς την υποστήριξη της σύνθεσης της ομιλίας. Στη μελέτη αυτή, οι παράγοντες που βρέθηκε να επηρεάζουν τη χρονική δομή των προτάσεων που εκφέρονται στην Αγγλική γλώσσα, καταχωρήθηκαν στη σχετική βιβλιογραφία. Κάθε παράγοντας εκτιμήθηκε για την ικανότητα να προσφέρουν στοιχεία αντίληψης επαρκή ώστε να παίρνουν γλωσσικές αποφάσεις. Επτά παράγοντες έχουν διατηρηθεί (και μετέπειτα συμπεριλήφθηκαν στο στατιστικό μοντέλο του ελέγχου της διάρκειας που χρησιμοποιείται στο MITalk):

1. Εξωγλωσσικοί παράγοντες (π.χ. η διάθεση του ομιλητή)
2. Παράγοντες που σχετίζονται με το λόγο (θέση μέσα σε μια παράγραφο)
3. Σημασιολογικοί παράγοντες (έμφαση και σημασιολογική πρωτοτυπία)
4. Συντακτικοί παράγοντες (επιμήκυνση της δομής μιας φράσης)
5. Παράγοντες που σχετίζονται με τις λέξεις (επιμήκυνση της λήγουσας)
6. Φωνολογικοί/φωνητικοί παράγοντες (εγγενής φωνολογική διάρκεια, τονισμός)
7. Φυσιολογικοί παράγοντες (εγγενής χρονική διάρκεια και ασυμπιεστότητα)

Μπορεί να δει κανείς ότι η χρονική διάρκεια επηρεάζει κάθε επίπεδο παραγωγής της ομιλίας. Πάντως, είναι χρήσιμο να διακρίνουμε δύο επίπεδα ελέγχου της διάρκειας, το εξωγενές και το ενδογενές (Ferreira 1993). Μονάδες που έχουν μήκος λέξης (λεξήματα, λέξεις που έχουν κλιθεί, τυποποιημένες εκφράσεις κ.α.), θεωρούνται πως έχουν ενδογενείς διάρκειες οι οποίες πιθανώς αποθηκεύονται σε κάποιο πνευματικό λεξικό. Κάθε φορά που χρησιμοποιούνται, η βασική *κατανομή* της διάρκειας των ποικίλων τμημάτων της δύσκολα θα είναι η ίδια. Καθώς αυτές οι μονάδες εξελίσσονται σε μεγαλύτερες οντότητες (φράσεις, εκφράσεις), "τραβιούνται" και "συμπιέζονται" ανάλογα με τις ανάγκες μεγαλύτερων απαιτήσεων στην ομιλία. Αυτές οι μεγαλύτερες απαιτήσεις αντιστοιχούν σε ένα εξωγενές επίπεδο του ελέγχου της διάρκειας. Θα υπάρχει περισσότερη ή λιγότερη έκταση, ανάλογα με το που μπαίνει η λέξη στη φράση, με το αν δίνεται έμφαση στη λέξη ή όχι, και σε ποια γραμματική ομάδα ανήκει η λέξη.

Έτσι, η πρώτη δουλειά σε ένα σύστημα μετατροπής γραπτού κειμένου σε ομιλία, είναι να αναλυθεί η δομή της πρότασης σε "φυσικές" ομάδες λέξεων. Οι παύσεις τείνουν να εμφανίζονται μεταξύ τέτοιων ομάδων. Η σωστή πρόβλεψη των θέσεων και της έκτασης των σιωπηρών και γεμάτων παύσεων καθορίζεται σε μεγάλο βαθμό από την έκταση των προσωδιακών μονάδων, τις λεγόμενες *εκτελεστικές δομές (performance structures)*.

Ένα Εργαλείο Γραμματικής Ανάλυσης: Οι Εκτελεστικές Δομές

Μια εκτελεστική δομή είναι μια ψυχογλωσσική δομή που αποτυπώνει τους διάφορους βαθμούς ενότητας (cohesion) μεταξύ των λέξεων μιας εκστομιζόμενης φράσης. Για παράδειγμα, στην προηγούμενη πρόταση, υπάρχει πολύ μεγαλύτερη ενότητα μεταξύ των λέξεων "τους", "διάφορους" και "βαθμούς" απ'ότι μεταξύ των "δομής" και της λέξης "που" που ακολουθεί. Με τον όρο "ενότητα" εννοείται η συχνότητα συνεμφάνισης (co-occurrence), σημασιολογικό-συντακτικές σχέσεις (όπως οι σχέσεις άρθρου-ουσιαστικού ή επιθέτου-ουσιαστικού), και η συντακτική σχέση (όπως συμφωνία ενικού/πληθυντικού). Υπάρχουν σημαντικά στοιχεία που αποδεικνύουν ότι οι ομιλητές φαίνεται ότι οργανώνουν το λόγο τους κάνοντας αναφορές σε μια τέτοια εσωτερική έννοια ενότητας μεταξύ των διάφορων τμημάτων μιας φράσης. Ουσιαστικά συγκλίνουσες εκτελεστικές δομές έχουν παρουσιασθεί από πολλά ψυχογλωσσικά πειράματα, όπως μνημονικές εργασίες, εργασίες διαισθητικής γραμματικής ανάλυσης που εκτελούνται από ανθρώπους που δεν είναι γλωσσολόγοι, όπως επίσης και από τη μέτρηση της χρονικής διάρκειας των παύσεων και των συλλαβών στο λόγο. Μια και η

εμπειρική μέτρηση της χρονικής διάρκειας των παύσεων και των συλλαβών έχει πιο άμεση σχέση με το πλαίσιο που εξετάζουμε, θα επικεντρωθούμε σε αυτού του τύπου τα αποδεικτικά στοιχεία.

Ξεκινώντας από τον Grosjean et al. (1979), κάποιοι συγγραφείς έχουν δείξει ότι η εμφάνιση και τα μήκη των παύσεων συσχετίζονται εντόνως με το βαθμό της ενότητας μεταξύ των λέξεων, που σημαίνει ότι οι παύσεις τείνουν να είναι μακρές και συχνές μεταξύ λέξεων που παρουσιάζουν σχετικά μικρή ενότητα και είναι πολύ μικρότερες και πολύ λιγότερο συχνές μεταξύ λέξεων που είναι ισχυρά εξαρτώμενες μεταξύ τους. Αυτή η ομαδοποίηση των λέξεων δείχνει να είναι ανεξάρτητη από αναπνευστικούς περιορισμούς: Οι ίδιες δομές βρίσκονται κατά τη φωναχτή ανάγνωση προτάσεων παραγόμενων με και χωρίς αναπνευστική διακοπή. Επιπλέον, όποια κι αν είναι η πειραματική μέθοδος, ο λαμβανόμενος χωρισμός σε τμήματα της πρότασης είναι παρόμοιος.

Για να μπορούμε να χρησιμοποιούμε αυτή την έννοια στη σύνθεση ομιλίας από κείμενο, πρέπει να προσδιοριστούν αξιόπιστοι προγνώστες. Κατ' αυτή τη διαδικασία το κύριο πρόβλημα είναι η αυτόματη αναγνώριση, που ξεκινάει και που τελειώνει μια εκτελεστική δομή. Σύμφωνα με τον Grosjean (Grosjean και Dommergues 1983, Monnin και Grosjean 1993), οι εκτελεστικές δομές έχουν τρία κύρια χαρακτηριστικά:

- (1) "Ευρυθμία": οι βασικές μονάδες τείνουν να είναι ίδιου μήκους.
- (2) "Ιεραρχία": οι βασικές μονάδες βρίσκονται μέσα σε μεγαλύτερες μονάδες οι οποίες είναι κι αυτές ενσωματωμένες μέσα σε ακόμα μεγαλύτερες μονάδες.
- (3) "Συμμετρία": οι εκτελεστικές δομές τείνουν να είναι ζυγισμένες, και η μεγαλύτερη παύση βρίσκεται κάπου στη μέση της πρότασης.

Η λεκτική ροή που φαίνεται στο σήμα της ομιλίας

ροή ομιλίας / **παύση** / ροή ομιλίας / **παύση** / ροή ομιλίας κ.τ.λ.

αντιστοιχεί στατιστικά, σημασιολογικά και συντακτικά σε

λέξεις με ισχυρή ενότητα / **παύση** / λέξεις με ισχυρή ενότητα / **παύση** / κ.τ.λ.

και αντιστοιχεί ψυχολογολογικά σε

εκτελεστική δομή / **παύση** / εκτελεστική δομή / **παύση** / εκτελεστική δομή / **παύση** / κ.τ.λ.

Πίνακας 2. Ομαδοποίηση λέξεων.

Ένα δεύτερο πρόβλημα αφορά την εσωτερική οργάνωση των δομών αυτού του τύπου. Για να απαντήσουμε σε αυτό το πρόβλημα, είναι απαραίτητο να ξεκαθαρίσουμε πως οι

ψυχογλωσσικές εκτελεστικές δομές που θεωρούμε ως βασική αρχή (και οι χρονικές δομές του λόγου που υπάγονται σε αυτές) σχετίζονται με τις συντακτικές δομές γλωσσικής βάσης και/ή με τις προσωδιακές δομές φωνητικής βάσης.

Είναι οι Εκτελεστικές Δομές Βασισμένες στις Συντακτικές και/ή στις Προσωδιακές Δομές ;

Οι Εκτελεστικές Δομές και οι Συντακτικές Δομές.

Δεδομένης της εκτενούς δουλειάς που έχει διεξαχθεί για τις συντακτικές δομές κατά τη δεκαετία του 1960 και του 1970, ήταν φυσικό οι εκτελεστικές δομές αρχικά να εξομοιωθούν με τις συντακτικές. Από τη στιγμή που οι εκτελεστικές δομές δείχνουν να είναι οργανωμένες όσο αναφορά την ενότητα μεταξύ διαφόρων λέξεων, η ενότητα αυτή, αντιλαμβανόταν ως άμεσα σχετιζόμενη με αυτή που εκφράζεται από τις συντακτικές δομές, δομές που είναι βασισμένες σε συγκεκριμένα κριτήρια για τη σύλληψη γραμματικών σχέσεων μεταξύ λεξικών μονάδων (lexical units) (Martin 1980).

Για το λόγο αυτό, ένας αριθμός μελετών στην αρχή, κατευθύνονταν προς τη δημιουργία εκτελεστικών δομών από συντακτική ανάλυση της πρότασης (π.χ. Grosjean 1980a, Grosjean και Dommergues 1983). Υπενθυμίζεται ότι τις συντακτικές δομές μπορούμε γενικά να τις δούμε σαν αναποδογυρισμένα δένδρα, όπου οι κόμβοι (σημεία διακλάδωσης) που βρίσκονται χαμηλότερα συνδέουν ισχυρά σχετιζόμενα, τυπικά κεντρικά στοιχεία και οι κόμβοι που βρίσκονται ψηλότερα συνδέουν μεγαλύτερα, λιγότερο άμεσα σχετιζόμενες ομάδες συντακτικών ομάδων. Μέσα σε αυτού του τύπου τη δομή, μεταβάσεις μεταξύ των τμημάτων του δένδρου, που εξαρτώνται από διαφορετικούς υψηλότερους κόμβους, σχετίζονται με μακρύτερες σιωπηρές παύσεις, σύμφωνα με την αρχή "όσο πιο περίπλοκη η μετάβαση, τόσο μεγαλύτερη είναι η παύση". Ωστόσο, τα αποτελέσματα που ελήφθησαν δεν ήταν ικανοποιητικά, γιατί συντακτικές και ψυχογλωσσικές δομές δεν βρέθηκε να είναι ομομορφικές (Grosjean και Dommergues 1983). Συχνά, οι εκτελεστικές δομές που παράγονταν εμπειρικά δεν ταίριαζαν με τις συντακτικά προτεινόμενες. Το αποτέλεσμα αυτό, ήταν ανεξάρτητο από την ακριβή συντακτική θεωρία που είχε υποθεθεί ότι ισχύει.

Πιο συγκεκριμένα, βρέθηκε ότι οι συντακτικές δομές είναι ανεπηρέαστες από το μήκος των συστατικών τους, ενώ οι κύριες εκτελεστικές μονάδες τείνουν να είναι περίπου ισοζυγισμένες, με την έννοια ότι περιέχουν περίπου τον ίδιο αριθμό λέξεων, και ότι το μήκος των μονάδων τους είναι παρόμοιο. Εάν μια πρόταση έχει για παράδειγμα, μια *βραχεία ουσιαστική φράση* που ακολουθείται από μια *μακρά ρηματική φράση*, το κυριότερο συντακτικό όριο βρίσκεται μεταξύ των δύο φράσεων. Ωστόσο, σύμφωνα με τα ψυχογλωσσικά αποδεικτικά στοιχεία, η κυριότερη "εκτελεστική" διακοπή τοποθετείται μεταξύ του ρήματος και της υπόλοιπης ρηματικής φράσης. Από την άλλη πλευρά, μια *μακρά ουσιαστική φράση* και μια *βραχεία ρηματική φράση*, υποδιαιρούνται διαφορετικά, έτσι ώστε τα συστατικά καταλήγουν να είναι περισσότερο ή λιγότερο ζυγισμένα.

Για παράδειγμα, όταν ζητείται από Γαλλόφωνους να κατατιμήσουν μια πρόταση σε τονικές ομάδες ή μονάδες, τότε αυτοί τείνουν να ζυγίζουν τον αριθμό των συλλαβών σε κάθε ομάδα (Gregoire 1899, Grosjean και Dommergues 1983, Monnin και Grosjean 1993, Fonagy 1992, Martin 1992). Το μέσο μήκος των μονάδων αυτών, κατά τη φωναχτή ανάγνωση απλών προτάσεων, υπολογίστηκε γύρω στις 3.46 συλλαβές με τυπική απόκλιση 1.43 (Monnin και Grosjean 1993).

Η συντακτική ανάλυση πολυπλοκότητας μίας πρότασης είναι αναμφίβολα ένας βασικός παράγοντας, αλλά από μόνος του, δεν προκύπτει να αποτελεί επαρκή προγνώστη εκτελεστικών δομών. Η Lehiste (1972) εξέτασε τις επιδράσεις των μορφολογικών και συντακτικών ορίων στη χρονική δομή των φράσεων που εκφέρονται. Κατέληξε στο ότι οι διεργασίες χρονικής αναπροσαρμογής τείνουν να αγνοούν τα όρια των μορφημάτων (morphemes) και των λέξεων. Η χρονική δομή, είναι σύμφωνα με τη Lehiste, ρυθμιζόμενη από τον αριθμό των συλλαβών παρά από τον αριθμό των τμημάτων ή την ύπαρξη ορίων. Ακόμη ο Ferreira (1993), έδειξε μέσω ενός αποτελεσματικού πειράματος ότι η χρονική δομή μιας πρότασης είναι άμεσα σχετιζόμενη με την προσωδιακή της δομή. Αν μια συντακτική παραλλαγή (αριθμός συντακτικών ορίων που ακολουθούν μια λέξη κλειδί) παρουσιαστεί μέσα σε κάποιο δεδομένο κείμενο, οι χρονικές μεταβλητές δεν επηρεάζονται σημαντικά. Ωστόσο, αν εισαχθεί μια προσωδιακή παραλλαγή (μια αλλαγή της έμφασης), χρονικές μεταβλητές όπως διάρκειες των τελευταίων λέξεων και παύσεις επηρεάζονται σημαντικά. Μπορούμε να συμπεράνουμε λοιπόν ότι, οι εκτελεστικές δομές εξαρτώνται από γλωσσικές μεταβλητές (όπως συντακτικά όρια) και από ψυχογλωσσικές μεταβλητές (όπως μήκος ή συμμετρία).

Εκτελεστικές Δομές και Προσωδιακές Μονάδες.

Εάν οι προσωδιακοί παράγοντες σχετίζονται, αλλά δεν είναι ομοιόμορφοι με τη συντακτική δομή (βλ. Hirst et al., in press), και αν οι εκτελεστικές δομές σχετίζονται, αλλά δεν είναι αποδεδειγμένα σύμφωνες με τη συντακτική οργάνωση, μένει να δει κανείς αν υπάρχει μια πιο κοντινή σχέση μεταξύ εκτελεστικών δομών και προσωδιακών μονάδων. Σύμφωνα με το σχήμα μας, κάτι τέτοιο θα έδινε:

Εκτελεστική δομή / παύση / εκτελεστική δομή / παύση / εκτελεστική δομή / κ.τ.λ. ισοδυναμεί με Προσωδιακή μονάδα / παύση / προσωδιακή μονάδα / παύση / προσωδιακή μονάδα / κ.τ.λ.

Πίνακας 3. Εκτελεστικές δομές και προσωδιακές μονάδες.

Υπάρχουν κάποιες ενδείξεις ότι αυτή η σχέση υπάρχει. Σύμφωνα με τον Hirst (in press), οι σιωπηρές και οι γεμάτες παύσεις αποτελούν τα ακουστικά σημεία τα οποία εμπεριέχουν τις προσωδιακές μονάδες (ή όπως στον τομέα αυτό: "προσωδιακή φράση",

"προσωδιακή δομή", κ.τ.λ.). Όπως τα ακουστά και τα φωνητικά χαρακτηριστικά της ομιλίας που εκτείνονται πέρα από τα όρια του φωνήματος (*suprasegmental*), τέτοιες προσωδιακές μονάδες μπορούν να οριστούν από τις *σχέσεις* που υπάρχουν μεταξύ των προσωδιακών παραμέτρων του λόγου (τονισμός, ρυθμός, διάρκεια κτλ.). Μερικά παρόμοια αποτελέσματα φαίνεται να προκύπτουν από τη δική μας δουλειά με αλγορίθμους (βλ. παρακάτω). Λαμβάνοντας λοιπόν υπόψη αυτή την εμφανή σχέση με την προσωδία, μπορεί να είναι χρήσιμο να εξετάσουμε πώς οι άνθρωποι φαίνεται να επεξεργάζονται τις προσωδιακές δομές.

Πώς οι Προσωδιακές Δομές Γεννώνται στον Ανθρώπινο Εγκέφαλο ;

Το βασικό ερώτημα αφορά τον τρόπο με τον οποίο μια εκτελεστική δομή – η διατύπωση μιας πρότασης – κατασκευάζεται καθώς μια φράση προετοιμάζεται ώστε να εκφραστεί. Σύμφωνα με τους Bachenko και Fitzpatrick (1990), υπάρχει ένας *ουδέτερος τρόπος διατύπωσης*, δηλαδή ένα πρότυπο διατύπωσης σε επίπεδο πρότασης, το οποίο είναι ανεξάρτητο από τη σημειολογία του λόγου. Σύμφωνα με αυτούς τους συγγραφείς, αυτός ο ουδέτερος τρόπος διατύπωσης εξαρτά όλες τις παραλλαγές, όπως αυτές του εστιασμού στην έκφραση, της έμφασης κ.τ.λ, που συνδέονται στο λόγο. Προτείνουν ότι αυτή η ουδέτερη από τη συνομιλία (*discourse-neutral*) διατύπωση χτίζεται επάνω στη γνώση των συντακτικών συστατικών, στη συνοχή των συστατικών μιας πρότασης, και στο μήκος της φράσης.

Με αυτό τον τρόπο, πρόσφατες φωνολογικές θεωρίες αντιλαμβάνονται την πρόταση ως ένα σύνολο προσωδιακών μονάδων που συνδυάζουν φωνολογικές και συντακτικές μονάδες. Κάθε προσωδιακή μονάδα προσδιορίζει ρυθμικά χαρακτηριστικά σε επίπεδο συλλαβών όπως και σε επίπεδο λέξης, φράσης και πρότασης. Για παράδειγμα, από προσωδιακής άποψης, η λέξη είναι ένα σύνολο από μέτρα (όσο αναφορά το ρυθμό), δηλαδή ένα ισχυρό-προς-αδύναμο σχήμα συλλαβών. Τέτοιες λέξεις ομαδοποιούνται σε μια *φωνολογική φράση* (Selkirk 1984). Αυτή η φωνολογική φράση δεν είναι απαραίτητα ισομορφική με την συντακτική φράση. Στο τέλος, οι φωνολογικές φράσεις ομαδοποιούνται σε *επιτονικές φράσεις*, οι οποίες με τη σειρά τους συνδυάζονται για να σχηματίσουν μια φραστική εκφορά του λόγου. Το ερώτημα είναι, σε ποίο σημείο κατά τη διάρκεια της διεργασίας παραγωγής του λόγου αυτή η ουδέτερη φωνολογική φράση κωδικοποιείται. Ο Ferreira (1993) λαμβάνει υπόψη του ότι η προσωδιακή συστατική δομή παράγεται *μη έχοντας γνώση* του τμηματικού περιεχομένου, ακριβώς όπως στο μοντέλο παραγωγής της γλώσσας του Levelt (για λεπτομέρειες βλ. Levelt 1989). Σύμφωνα με αυτή την υπόθεση, *αόριστα χρονικά_διαστήματα*, τοποθετούνται σε διάφορες περιοχές μιας πρότασης νωρίς κατά τη διεργασία προσχεδιασμού της ομιλίας. Αργότερα, οι παράγοντες κατάτμησης λαμβάνονται υπόψη, ενώ η διάρκεια των λέξεων και των παύσεων έχουν μια συμβιβαστική σχέση, ώστε να γίνει σεβαστή η συνολική διάρκεια μιας κατανεμημένης διακοπής στο λόγο. Σε φωνητικό επίπεδο, μια μεγάλη λέξη καταλαμβάνει περισσότερο χώρο στο χρονικό διάστημα απότι μια μικρή λέξη, ωστόσο μεγαλύτερο μήκος μιας λέξης συνεπάγεται και μια μικρότερη επακόλουθη παύση. Αυτό σημαίνει πως μέσα σε κάποιο διάστημα, το μέγεθος της λέξης και της παύσης έχουν

ανάδρομη σχέση. Το μέγεθος ενός χρονικού διαστήματος καθορίζεται από τον αριθμό των προσωδιακών χαρακτηριστικών. Όσο περισσότερα όρια προσωδιακών χαρακτηριστικών καταλήγουν σε μία δεδομένη λέξη, τόσο μεγαλύτερο θα είναι και το χρονικό διάστημα για τη λέξη αυτή.

Είναι δυνατόν έτσι, να βρεθεί μια γενική συμμετρία σχετικά με τα χρονικά φαινόμενα. Μπορεί λοιπόν ειπωθεί με κάποια βεβαιότητα πως η εκφορά του λόγου σχετίζεται με τις προσωδιακές δομές, και πως τα σχήματα του ρυθμού μας επιτρέπουν να καθιερώσουμε μια σχέση μεταξύ χρονικών φαινομένων περιεχομένου των γλωσσικών τμημάτων. Αυτό είναι το βασικό θεώρημα πάνω στο οποίο μπορούν να βασιστούν οι αλγόριθμοι για τη σύνθεση της ομιλίας.

Η Αυτόματη Γένεση των διαφόρων Χρονικών Διαρκειών: Αλγόριθμοι

Παρακάτω, εξετάζονται περιληπτικά οι σημαντικότεροι από τους αλγορίθμους που ανεπτύχθησαν για την πρόβλεψη και την αυτόματη γένεση των παύσεων, και των χρονικών διαρκειών γενικότερα, με βάση ένα δεδομένο κείμενο, ενώ στη συνέχεια δίδεται έμφαση στον αλγόριθμο Keller-Zellner, σε μοντέλα που χρησιμοποιούν κάποια *στοχαστική χωρίς πλαίσια γραμματική (Stochastic Context-Free Grammar – SCFG / ΣΑΑΣΓ)*, και στα μοντέλα Campbell και Barbosa-Bailly, που προβλέπουν τις χρονικές διάρκειες και τη γένεση των παύσεων μέσω του μοντέλου z-score.

Δύο τύποι μοντέλων χρονικής διάρκειας μπορούν να διακριθούν: τα *νόμο-κυβερνώμενα συστήματα (rule-governed systems)* και τα *στατιστικά συστήματα (statistical systems)*.

Νόμο-Κυβερνώμενα Συστήματα

Στα μοντέλα αυτά, ένας αριθμός κανόνων – γενικά σειριακών νόμων – εφαρμόζεται στο γλωσσικό υλικό το οποίο χαρακτηρίζεται από μια ενδογενή χρονική διάρκεια. Αυτό, καταλήγει σε μια διάρκεια του παραγόμενου προϊόντος. Σε αυτές τις χρονικές διάρκειες που εμφανίζονται στο παραγόμενο προϊόν, θεωρείται ότι λαμβάνονται υπόψη διάφορα φαινόμενα όπως γλωσσικοί περιορισμοί (π.χ. συντακτικοί παράγοντες) ή ψυχογλωσσικοί περιορισμοί (π.χ. μήκος).

Οι ακόλουθοι Αλγόριθμοι (GGL, CPC, PHI, Bachenko & Fitzpatrick, και MG) είναι όλοι τους χτισμένοι πάνω στις ίδιες βασικές έννοιες, καθώς χρησιμοποιούν έναν αριθμό από έννοιες της *γενετικής-μετασχηματιστικής θεωρίας (generative syntax)*¹.

¹ Περί της θεωρίας αυτής, αναλυτικότερα στοιχεία υπάρχουν ακολούθως του πίνακα αγγλικών όρων, στο τέλος της εργασίας.

Grosjean Grosjean & Lane – GGL. Ο στόχος αυτού του αλγόριθμου ήταν η πρόβλεψη των εκτελεστικών δομών. Για κάθε όριο λέξης, ο κυκλικός αυτός αλγόριθμος ταυριάζει έναν κατάλογο της γλωσσικής πολυπλοκότητας σε ένα μέτρο της απόστασης από το μέσο της πρότασης.

Η μέση συσχέτιση μεταξύ των προβλέψεων και των παρατηρήσεων είναι 0.83, γεγονός που οδηγεί στο συμπέρασμα ότι αυτός ο απλός αλγόριθμος είναι σχετικά καλός προγνώστης των χρονικών διαρκειών, ακόμη και για τα δεδομένα των παύσεων. Ωστόσο, η συντακτική ανάλυση τείνει να υπερεκτιμά τη σημαντικότητα των γλωσσικών σημείων που λειτουργούν μεταξύ των λέξεων (π.χ. σύνδεσμοι, αναφορικά κ.τ.λ) (Gee & Grosjean 1983).

Cooper & Paccia Cooper (1980) – CPC. Ο αλγόριθμος τους περιέχει 14 κανόνες για την πρόγνωση, στην Αγγλική γλώσσα, της πιθανότητας μια παύση να βρίσκεται σε κάθε διαλεξικό όριο, όπως επίσης τμηματικές επιμηκύνσεις, χρονικές διάρκειες τμημάτων και παύσεων. Αυτό βασικά μπορεί να δώσει εξηγήσεις για το βάθος των συντακτικών κόμβων.

Η μέση συσχέτιση που λαμβάνεται μεταξύ των μετρήσεων και των προβλέψεων είναι περίπου 0.75. Όταν εφαρμόζεται στα Γαλλικά, η συσχέτιση ελαττώνεται στο 0.57. Οι βασικές αντιδράσεις έγκεινται στο γεγονός ότι η εφαρμογή των κανόνων δεν είναι αρκετά ξεκάθαρη, και δεν υπάρχουν καθόλου διχοτομημένες συνθέσεις (bisection compounds) για προτάσεις ενός μεσαίου μεγέθους. (Gee & Grosjean 1983).

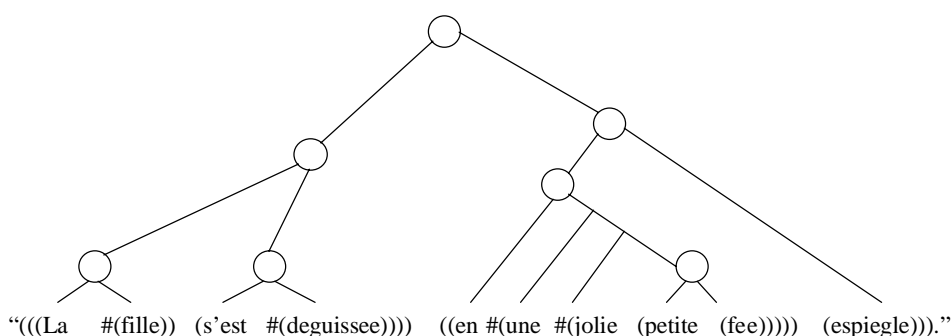
Gee & Grosjean (1983) Phonological Phrase Algorithm – PHI (Φωνολογικός Αλγόριθμος Φράσεων). Αυτός ο αλγόριθμος προτείνει ότι δυο παράγοντες κυβερνούν τη χρονική δομή της πρότασης: Η συντακτική δομή και ο διαχωρισμός που βασίζεται στο πληροφοριακό τους περιεχόμενο μεταξύ των λειτουργικών και των λεξικών λέξεων (lexical words). Οι λειτουργικές λέξεις θεωρούνται ότι είναι ασθενείς συλλαβές. Ο αλγόριθμος αυτός δεν απαιτεί ολόκληρη τη δενδρική δομή μιας πρότασης για να αρχίσει την κατάτμηση σε προσωδιακές μονάδες.

Η μέση συσχέτιση αναφέρεται να είναι πολύ υψηλή (0.96), πιθανώς γιατί ο αλγόριθμος ενσωματώνει φωνολογικές, συντακτικές και προσωδιακές πληροφορίες. Όταν εφαρμόζεται στα Γαλλικά, τα κυριότερα προσωδιακά όρια μπορούν άριστα να προβλεφθούν, μόνο που υπερεκτιμάται το μεταρθρωτικό όριο (post-verbal boundary).

Bachenko & Fitzpatrick (1990) – BF. Επηρεασμένοι από τον προηγούμενο αλγόριθμο, οι συγγραφείς αυτοί συνέλαβαν ένα σύστημα για την Αγγλική γλώσσα ικανό να προβλέπει τη γλωσσική έκφραση με βάση μια διαδικασία τοπικού προσδιορισμού (localising) των ορίων της φράσης. Πρώτα, προσδιορίζονται οι φωνολογικές λέξεις και φράσεις. Έπειτα, εφαρμόζονται οι κανόνες των εμφανών ορίων, προσδιορίζοντας μια σχετική ισχύ - δηλαδή τη δυνατότητα να γίνονται αντιληπτά - σε κάθε όριο μιας φράσης σύμφωνα με την συντακτική της κατηγοριοποίηση, το μήκος και τη συνοχή της.

Τα αποτελέσματα τους δείχνουν πως ένα 80% των πρωταρχικών ορίων - δηλαδή των ορίων που μπορούν να γίνουν πιο εύκολα αντιληπτά - προβλέφθηκε σωστά. Οι περιορισμοί του συστήματος αυτού αφορούν λάθη στη γραμματική ανάλυση των προτάσεων (parsing). Για παράδειγμα βρέθηκαν λαθεμένοι ορισμοί στα μέρη του λόγου και λαθεμένες αναλύσεις στους προσδιορισμούς πριν από το ρήμα (pre-head modifiers).

Monnin and Grosjean (1993) - MG. Ο αλγόριθμος αυτός, που αναπτύχθηκε ιδιαίτερα για τα Γαλλικά είναι επίσης πολλά υποσχόμενος, επειδή ο μέσος συσχετισμός μεταξύ των δομών της απόδοσης εννιά προτάσεων που διαβάζονται και η πρόβλεψη των δομών αυτών με τον αλγόριθμο είναι γύρω στο 0.94. Κύρια και δευτερεύοντα προσωδιακά όρια προβλέπονται σωστά. Αυτό έχει σαν αποτέλεσμα να χωρίζεται η πρόταση σε προσωδιακά τμήματα που δεν είναι απαραίτητα αντίστοιχα με τις επιφανειακές συντακτικές δομές. Πάραυτα ο αλγόριθμος αυτός, δεν έχει δοκιμαστεί διεξοδικά σε άλλες προτάσεις.



Σχήμα 4. Ένα δέντρο εκτελεστικής δομής (performance structure tree) σύμφωνα με τους κανόνες των Monnin και Grosjean. Ένας μόνο κόμβος χωρίζει τα «petite» και «fee», το οποίο προβλέπει μια μικρή τελική συλλαβή και μια συνεχόμενη παύση για το «petite». Σε αντίθεση με αυτό, πέντε κόμβοι διαχωρίζουν το «deguisee» και «en», το οποίο προβλέπει μια πολύ μεγαλύτερη τελική συλλαβή και διάρκεια παύσης σε αυτή την κύρια ένωση. Το σύμβολο #- καθορίζει μια συνένωση γραμματικών στα λεξικά σημεία. Τα δέντρα εκτελεστικής δομής (performance structure trees) είναι αρκετά διαφορετικά από τα δέντρα συντακτικής δομής, οποιαδήποτε και αν είναι η θεωρητική προσέγγιση.

Οι κανόνες Keller-Zellner (KZ) (1993): Αρχικά προερχόμενος από τον MG αλγόριθμο και ξεκινώντας για τα Γαλλικά, ο αλγόριθμος αυτός είναι αρκετά διαφορετικός τόσο σε μορφή όσο και σε περιεχόμενο. Βασικά, οι στόχοι των κανόνων αυτών είναι να ικανοποιήσουν τα κριτήρια απλότητας, σε ότι αφορά την *ψυχογλωσσική παραδοχή*, και την *υψηλή ικανότητα πρόβλεψης* για τα σύνολα των δεδομένων που είχαν στη διάθεσή τους.

Τα προσωδιακά συστατικά σχηματίζονται με βάση το απλό προσεγγιστικό συντακτικό. Δεν χρειάζονται συντακτικές δομές περισσότερο πολύπλοκες από αυτές που εφαρμόζονται σε μια μόνο πρόταση. Οι προσωδιακές ομάδες μπορούν να

προσδιοριστούν από την εφαρμογή του 1 και 2 σταδίου στους κανόνες των Monnin-Grosjean.

Τα τελικά διαστήματα συλλαβής+παύσης αυξάνουν σε διάρκεια καθώς το συστατικό προχωρά. Η αύξηση προχωρά από ένα εμπειρικό ελάχιστο προς ένα εμπειρικό μέγιστο. Η αρχική υπόθεση απαιτεί ισοδύναμα βήματα. Αυξημένες διάρκειες αντιστοιχούν σε μια ελάττωση του ρυθμού, φαινόμενο το οποίο συχνά παρατηρείται στην ομιλία.

Ρυθμική εναλλαγή παρατηρήθηκε σε δυο μέρη: μεταφωνητικά και στη μέση συστατικών που αποτελούνταν από 4-6 λέξεις. Οι ρυθμικές εναλλαγές εμφανίζονται όταν ένα στοιχείο επιμηκύνεται σε διάρκεια περισσότερο απ'ότι επιτρέπεται. Ως αποτέλεσμα, το στοιχείο που το ακολουθεί θα πρέπει να γίνει μικρότερο σε διάρκεια "ώστε να περατωθεί η διάρκεια του συστατικού στον απαραίτητο χρόνο". Συγκεκριμένα, αυτό σημαίνει πως πρέπει να θέσουμε ως βασική αρχή μια αντιστροφή των χρονικών διαρκειών στο ζεύγος των λέξεων που συμπεριλαμβάνονται σε μια εναλλαγή.

Monnin – Grosjean		Keller - Zellner		
Caelen -Haumont Σύνολο Δεδομένων	Κανονικά	Αργά	Κανονικά	Αργά
Πρόταση 1	.786	.895	.862	.845
Πρόταση 2	.289	.375	.811	.829
Πρόταση 3	.925	.808	.878	.751
Μέση Τιμή	.667	.693	.850	.808
Monnin -Grosjean Σύνολο Δεδομένων	Κανονικά	Αργά	Κανονικά	Αργά
Πρόταση 1	.890	.674	.873	.835
Πρόταση 2	.914	.796	.886	.954
Πρόταση 3	.981	.886	.773	.892
Πρόταση 4	.961	.826	.798	.850
Πρόταση 5	.947	.736	.827	.872
Πρόταση 6	.984	.711	.812	.835
Πρόταση 7	.931	.841	.754	.906
Πρόταση 8	.940	.585	.870	.809
Πρόταση 9	.968	.808	.701	.818
Μέση Τιμή	.946	.763	.810	.863

Πίνακας 4. Γραμμικοί συσχετισμοί μεταξύ προβλέψιμων και μετρήσιμων διαρκειών τελικών συλλαβών+παύσεων, σύμφωνα με δυο διαφορετικούς κανόνες.

Ο αλγόριθμος που συνεπάγεται από αυτά είναι αρκετά απλός και αναφέρεται ξανά στο τέλος του κεφαλαίου αυτού. Συσχετισμοί μεταξύ των στοιχείων των Caelen-Haumont και των Monnin-Grosjean, αναφέρονται στον πίνακα 4. Έχει βρεθεί πως οι συσχετίσεις εμφανίζουν κάποια τακτικότητα. Δεν πέφτουν ποτέ κάτω από μια γραμμική συσχέτιση .7, και γενικά τείνουν να βρίσκονται στην κλίμακα του 0.8.

Μια θεώρηση της εξέλιξης του F_0 και των τιμών της ενέργειας στο τέλος των προσωδιακών συστατικών που τίθεται εδώ, δείχνει κάποιες κανονικότητες. Οι τιμές F_0 αυξάνουν στο τέλος του κάθε συστατικού, εκτός από τα συστατικά εκείνα που

βρίσκονται στο τέλος μιας πρότασης. Οι τιμές της ενέργειας φθίνουν με κάποια τακτικότητα στο τέλος κάθε συστατικού. Αυτό, υπονοεί πως η χρονική δομή που διακρίνεται στο σημείο αυτό, αλληλεπιδρά άμεσα με τον έλεγχο που έχουμε πάνω στο F_0 και την ενέργεια.

Εφόσον δεν μπορούν να προβλέψουν *ολόκληρη* τη χρονική δομή μιας φράσης, τα προηγούμενα μοντέλα είναι, προφανώς, ατελή. Ακόμα κι αν οι διαλεξικές παύσεις και οι επιμηκύνσεις των τελευταίων συλλαβών παράγονται με επιτυχία, οι αλγόριθμοι αυτοί παραβλέπουν τις χρονικές διάρκειες των συλλαβών και των μερών του λόγου.

Στατιστικά συστήματα

Από την άποψη των συστημάτων από κείμενο-σε-ομιλία (Text-to-Speech systems, TtS), είναι σημαντικό να έχουμε τον έλεγχο του χρόνου ομιλίας σε κάθε επίπεδο της διαδικασίας για την παραγωγή μιας πρότασης. Το σύστημα MITalk - το Βρετανικό σύστημα από κείμενο-σε-ομιλία που αναπτύχθηκε στο Ινστιτούτο Τεχνολογίας της Μασαχουσέτης (Massachusetts Institute of Technology) (Allen et al., 1987) - ικανοποιεί την απαίτηση αυτή, και γι' αυτόν το λόγο αποτελεί ένα έξοχο σημείο εκκίνησης. Παρόλο που το σύστημα αυτό, ανήκει σε μια παλαιότερη γενιά συνθετών και ακούγεται λίγο τεχνητός για τα σημερινά δεδομένα, η χρονική του δομή βασίζεται σε μια μέθοδο με μεγάλη επιρροή που πρωτοαναπτύχθηκε από τον Klatt (1976). Αυτό, είναι ένα στατιστικό μοντέλο που χτίστηκε γύρω από τις χρονικές διάρκειες των μερών του λόγου, δηλαδή, τις χρονικές διάρκειες για κάθε φώνημα ξεχωριστά.

Με βάση τα στοιχεία που συνέλεξε από μια ποικιλία εργασιών που έγιναν από διάφορους συγγραφείς, το μοντέλο φωνήεντος του Klatt ξεκινά κάνοντας ένα υπολογισμό μιας ενδογενούς φωνολογικής διάρκειας για ένα δεδομένο τμήμα του λόγου. Έπειτα, η διάρκεια αυτή, μικραίνει ή μεγαλώνει συναρτήσει του τμήματος που ακολουθεί, της θέσης του μέσα στη φράση, της ύπαρξης ή της απουσίας τονισμού, και του μήκους της λέξης. Τα ασυνάρτητα λεξιλογικά δεδομένα του Lehiste (cf., Klatt, 1976) αναλύθηκαν ξανά, και βρέθηκε πως τέσσερις μόλις κανόνες στο μοντέλο του Klatt μπορούν να εξηγήσουν το 97% των συνολικών παραλλαγών στις φωνηεντικές διάρκειες που έχουν μετρηθεί.

Ένα κάπως παρόμοιο μοντέλο προτάθηκε από τον O'Shaughnessy (1981, 1984). Αυτό, είναι πιθανότατα το πιο σημαντικό στατιστικό μοντέλο για ένα κείμενο της καθομιλουμένης Γαλλικής γλώσσας. Με βάση αρκετά αναγνώσματα ενός μικρού κειμένου που περιείχε όλα τα Γαλλικά φωνήματα, προτάθηκε ένα μοντέλο διαρκειών ακουστικής κομμάτια κατάλληλα για σύνθεση βάσει κανόνων. Στο μοντέλο αυτό, 33 κανόνες προσδιορίζουν τις βασικές διάρκειες για διάφορες τάξεις των τμημάτων του λόγου, καθώς και τροποποιήσεις στη βασική αυτή διάρκεια, συναρτήσει του φωνητικού πλαισίου.

Για τις ηχητικές τάξεις που δεν ενέχουν προπαυσική έκταση του ήχου, το μοντέλο αυτό μπόρεσε να προβλέψει τις διάρκειες για 281 τμήματα ενός κειμένου, με μια τυπική απόκλιση των 9ms. Όμως ήταν λιγότερο ακριβές στην πρόβλεψη των προπαυσικών φωνηεντικών διαρκειών, εξαιτίας των περισσότερων παραλλαγών των τμημάτων στις θέσεις αυτές. Επιπλέον, το μοντέλο αυτό, δεν μπόρεσε να προβλέψει τις σιωπηρές διαλεξικές παύσεις.

Τα δυο αυτά στατιστικά μοντέλα έχουν δομηθεί γύρω από την ίδια βασική υπόθεση, μια υπόθεση η οποία είναι ανοιχτή σε μια σημαντική κριτική. Οι συγγραφείς υποθέτουν πως τα φαινόμενα τα σχετικά με τη χρονομέτρηση του λόγου, μπορούν να συμπεριληφθούν από ένα τμήμα του λόγου, λες και το κομμάτι αυτό “κατέχει μια ενδογενή τιμή-στόχο σε σχέση με την άρθρωση ή την ακουστική πραγμάτωση” (Fujimura, 1981). Παρόλα αυτά, πρόσφατες μετρήσεις έδειξαν πως οι χρονικές διάρκειες σε μέγεθος συλλαβής έχουν γενικά λιγότερες παραλλαγές σε σχέση με τις χρονικές διάρκειες τμημάτων μικρότερων της συλλαβής, και γι’αυτό αποτελούν πιο αξιόπιστα σημεία εκκίνησης για τον υπολογισμό των διαρκειών των τμημάτων που είναι μικρότερα της συλλαβής (Barbosa and Bailly, 1993). Αυτή η προσέγγιση της χρονικής διάρκειας λαμβάνει μεγαλύτερη υποστήριξη εξαιτίας της παρατήρησης πως οι παραλλαγές στον τονισμό καθώς κι εκείνες στο ρυθμό εκφοράς του λόγου τείνουν να τροποποιούν τα κομμάτια στο μέγεθος τουλάχιστον μιας συλλαβής, φυσικά μεγαλύτερων από μονά τμήματα του λόγου. Οι χρονικές διάρκειες των φαινομένων συνάρθρωσης (co-articulation) μπορούν επίσης να ωφεληθούν από αυτή την ανακατεύθυνση της προοπτικής. Ίσως να είναι πιο ωφέλιμο να δούμε τα τμήματα της γλώσσας και τα φαινόμενα συνάρθρωσης από την προοπτική του να συνενώσουμε ένα αριθμό τέτοιων τμημάτων σε μια συλλαβή, παρά από το ατομικό μέτρο διάρκειας που αντιπροσωπεύεται από το φώνημα.

Ανεξάρτητα από το αν ο υπολογισμός μας ξεκινά από ένα τμήμα ή από μια συλλαβή, η στατιστική προσέγγιση στο θέμα της διάρκειας, υπόσχεται πολλά. Αποτελέσματα του τύπου αυτού, υποδεικνύουν πως ένα πιθανό επόμενο βήμα σε αυτόν τον τομέα έρευνας είναι η ανάπτυξη ενός ισχυρού στατιστικού μοντέλου, ικανού να προβλέψει ολόκληρη τη δομή των χρονικών διαρκειών μιας πρότασης.

Ο Αλγόριθμος Keller-Zellner

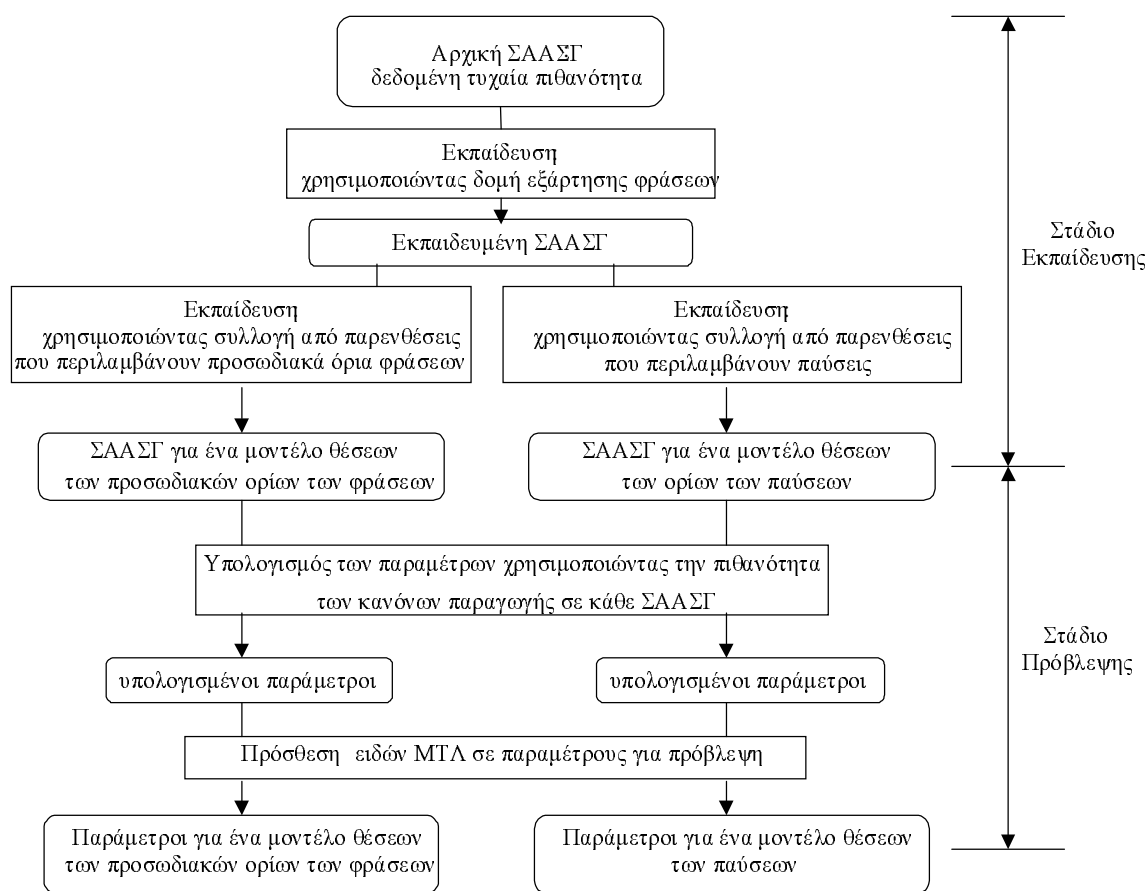
- (1) Αναγνώριση, από τα δεξιά προς τα αριστερά, του πυρήνα των προσωδιακών συστατικών: ουσιαστικών, ρημάτων και ανεξαρτήτων επιθέτων, επιρρημάτων και αντωνυμιών (όπως τα “*La chemise est sale*”, “*c'est bien*”, “*pense a ca*”).
- (2) Δημιουργία των προσωδιακών συστατικών ομαδοποιώντας τις λέξεις γύρω από τον πυρήνα. Όλες οι λέξεις στα αριστερά του πυρήνα προσαρτώνται στον πυρήνα που βρίσκεται στα δεξιά, εκτός από τα επίθετα και τις αντωνυμίες που έπονται δομικά και που προσαρτώνται στον πυρήνα που βρίσκεται στα αριστερά (“*la chemise blanche*”, “*donne-lui*”).

- (3) Υπολογισμός των προβλέψεων που μπορούν να γίνουν για τη διάρκεια τελευταίας συλλαβής+παύσης. Μέσα σε κάθε προσωδιακό συστατικό οι διάρκειες αυξάνουν από μια ελάχιστη σε μια μέγιστη διάρκεια. Αρχικά, η αύξηση αυτή φαίνεται να εμφανίζεται σε ίσα στάδια. (Η ελάχιστη και η μέγιστη διάρκεια φαίνεται να είναι 50 και 350 ms στη φυσιολογική ομιλία, 50 και 525 ms στην αργή ομιλία). Η πρώτη τελική συλλαβή σε ένα συστατικό έχει μια διάρκεια του μεγέθους του ελάχιστου+σταδιακού ms.
- (4) Ρυθμικές συμβιβαστικές ανταλλαγές:
1. Προ-φωνητική συμβιβαστική ανταλλαγή: Όταν ένα συστατικό ακολουθεί ένα ρήμα και υπάρχουν τουλάχιστον δυο λέξεις προηγούμενες του πυρήνα, τότε η διάρκεια της τελευταίας συλλαβής της πρώτης λέξης αυξάνεται σύμφωνα μ' εκείνη της δεύτερης λέξης. (Ανταλλαγή στις διάρκειες για τις λέξεις 1 και 2).
 - 2a. Ρυθμική εναλλαγή: Αν ένα συστατικό αποτελείται από τέσσερις ή περισσότερες λέξεις, και αν η λέξη 3 αποτελείται από δυο ή περισσότερες συλλαβές, η λέξη 2 αυξάνεται σε διάρκεια σύμφωνα με τη λέξη 3. (Ανταλλαγή στις διάρκειες για τις λέξεις 2 και 3).
 - 2b. Αν ο κανόνας 1 έχει ήδη εφαρμοστεί: Αν ένα συστατικό αποτελείται από τέσσερις ή περισσότερες λέξεις, και αν η λέξη 4 αποτελείται από δυο ή περισσότερες συλλαβές, η λέξη 3 αυξάνεται σε διάρκεια σύμφωνα με τη λέξη 4. (Ανταλλαγή στις διάρκειες για τις λέξεις 3 και 4).
 3. Συστατικά που αποτελούνται από μια λέξη: Τα συστατικά που περιέχουν μια μόνο λέξη δείχνουν μειωμένες διάρκειες στις τελικές συλλαβές. (Μειωμένες διάρκειες για συστατικά της μιας λέξεις κατά 50 ms).
- (5) Μέτρηση της τελευταίας συλλαβής+παύσης. Η μέτρηση ξεκινά με το φωνήεν της τελευταίας συλλαβής και τελειώνει με το πέρας της παύσης. Περιλαμβάνει οποιοδήποτε εμβόλιμο σύμφωνο μπορεί να εμφανιστεί, αποκλείει όμως το προαιρετικό φώνημα *schwa*, γεγονός που μας επιτρέπει να κάνουμε άμεσες συγκρίσεις με ομάδες στοιχείων μεταξύ βορείων ομιλητών και αυτών που ζουν στη νότια Ευρώπη. Οι επερχόμενες χρονικές μετρήσεις ήταν αρκετά όμοιες. Σε περιορισμένες ομάδες στοιχείων, το παρεμβαλλόμενο σύμφωνο καταπνιγόταν. Παρόλα αυτά, οι επερχόμενες διάρκειες βρέθηκαν να έχουν μεγαλύτερες παραλλαγές από αυτές που συμπεριλάμβαναν το σύμφωνο. Μετρήσεις για λέξεις που βρίσκονται στο τέλος μιας πρότασης, έγιναν γνωστές για λίγες μόνο προτάσεις και γι' αυτό μηδενίστηκαν σε όλες τις περιπτώσεις για στατιστικούς λόγους.

Μοντέλα Πρόβλεψης της θέσης των Κύριων Ορίων μιας Φράσης και Εισαγωγής Παύσεων χρησιμοποιώντας μια Στοχαστική Ανεπιρέαστη Από Συμφραζόμενα Γραμματική (ΣΑΑΣΓ).

Τα μοντέλα αυτά, αναπτύχθηκαν βασιζόμενα σε παρόμοιες ιδέες, καθώς τόσο οι θέσεις των ορίων μιας φράσης, όσο και η εισαγωγή των παύσεων έχουν παρόμοια χαρακτηριστικά. Στα μοντέλα αυτά, οι ιδιότητες των λέξεων και οι παράμετροι της πιθανότητας να υπάρξει αριστερή η δεξιά διακλάδωση που αναπαριστούν τα στοχαστικά χαρακτηριστικά της έκφρασης, χρησιμοποιούνται ως παράμετροι εισόδου ενός τροφοδοτούμενου από μπροστά νευρωνικού δικτύου που χρησιμοποιείται για την πρόβλεψη. Για να βρούμε τις πιθανότητες, πρώτα, γίνεται εκμάθηση των κύριων χαρακτηριστικών των φράσεων και των χαρακτηριστικών των παύσεων, μέσω της ΣΑΑΣΓ εκπαίδευσης, χρησιμοποιώντας έναν *μέσα-έξω αλγόριθμο (inside-outside algorithm)*. Έπειτα, οι πιθανότητες κάθε παρενθετικής δομής (bracketing structure) υπολογίζονται χρησιμοποιώντας την ΣΑΑΣΓ. Για την επιβεβαίωση της αποτελεσματικότητας αυτών των στοχαστικών μοντέλων διεξήχθησαν πειράματα, σύμφωνα με τα οποία επιτεύχθηκε σωστή πρόβλεψη του 92.9% των κύριων ορίων των φράσεων, με σφάλμα 16.9%, και του 85.2% των ορίων των παύσεων, με σφάλμα 9.1%. Τα αποτελέσματα αυτά, δείχνουν ότι τα μοντέλα που χρησιμοποιούν ΣΑΑΣΓ είναι αποτελεσματικά.

Τα Μοντέλα Πρόβλεψης Σχηματικά.



Σχήμα 5. Ροή της παραγωγής των παραμέτρων για το μοντέλο.

Τα μοντέλα πρόβλεψης των θέσεων των κύριων ορίων και της εισαγωγής των παύσεων που παρουσιάζονται στην ενότητα αυτή, έχουν και οι δυο τον ίδιο σχεδιασμό και διαδικασία εκμάθησης, όπως εικονίζεται στο σχήμα 5. Οι δομικοί περιορισμοί που περιγράφονται με εξαρτήσεις φράσεων που διασκευάζονται με το χέρι χρησιμοποιώντας φωνητικά σύμβολα, γίνονται κατά προσέγγιση αντιληπτά από τις ΣΑΑΣΓ. Οι παράμετροι που αντικατοπτρίζουν τους δομικούς περιορισμούς, υπολογίζονται χρησιμοποιώντας την πιθανότητα των κανόνων παραγωγής σε κάθε ΣΑΑΣΓ, και μετά χρησιμοποιούνται σε μοντέλα για την πρόβλεψη των θέσεων των κύριων ορίων των φράσεων και τις θέσεις των παύσεων. Καθώς τα μέρη-του-λόγου (*Part of Speech – POSMTL*) πριν και μετά τα όρια έχουν συχνά χρησιμοποιηθεί για την πρόβλεψη των θέσεων των κύριων ορίων των φράσεων και των θέσεων των παύσεων, συμπεριελήφθησαν ως παράμετροι για την κατασκευή των μοντέλων. Κάθε μοντέλο προβλέπει μια θέση χρησιμοποιώντας αυτές τις παραμέτρους σε συνδυασμό με ένα τροφοδοτούμενο από μπροστά νευρωνικό δίκτυο, το οποίο περιγράφεται παρακάτω.

Τα Δεδομένα Ομιλίας.

503 προτάσεις με δεδομένα ομιλίας χρησιμοποιήθηκαν για την εκπαίδευση του ΣΑΑΣΓ και ένα νευρωνικό δίκτυο. Οι προτάσεις εκφωνήθηκαν από δέκα Ιάπωνες

επαγγελματίες εκφωνητές και αφηγητές, δίδοντας 5030 προτάσεις που περιελάμβαναν 70020 όρια ως δεδομένα εκπαίδευσης και δοκιμών. Οι προτάσεις των δεδομένων ομιλίας αναλύθηκαν μορφηματικά, τέθηκαν εντός παρενθέσεων, και τους δόθηκε κάποιο όνομα για το ΜΤΛ.

Εκμάθηση των Θέσεων των Κύριων Ορίων των Φράσεων και των Θέσεων των Παύσεων χρησιμοποιώντας μια ΣΑΑΣΓ.

Οι παράμετροι που αναπαριστούν τις δομές εξάρτησης των φράσεων χρειάζονται για την πρόβλεψη των θέσεων των κύριων ορίων των φράσεων και των θέσεων των παύσεων. Αναμένουμε πως η ΣΑΑΣΓ μπορεί να καταλάβει αυτές τις δομές μέσω της διαδικασίας εκπαίδευσης. Για την εκπαίδευση της ΣΑΑΣΓ, έχει προταθεί ένας αποτελεσματικός μέσα-έξω αλγόριθμος (Lary K., Young S.J., 1989) και μια επέκταση αυτού, έχει εφαρμοσθεί σε κείμενο που έχει μερικώς τεθεί εντός παρενθέσεων (Pereira F., Shabes Y., 1992). Επεκτείνουμε αυτόν τον αλγόριθμο ώστε να γίνει η εκμάθηση της δομής της προσωδιακής εξάρτησης των φράσεων (Sagisaka Y. and Kaiki N. 1992). Για την εκπαίδευση της ΣΑΑΣΓ χρησιμοποιώντας αυτή τη μέθοδο, τα δεδομένα των προτάσεων ονοματίστηκαν με δύο τύπους πληροφοριών για την εισαγωγή των παρενθέσεων, όπως εξηγείται παρακάτω:

(1) Εισαγωγή παρενθέσεων που υποδεικνύουν εξάρτηση των φράσεων.

Η δομή της εξάρτησης των φράσεων και ένα μέρος της ακολουθίας του λόγου σημαδεύτηκαν χειροκίνητα από εκπαιδευμένους στη διασκευή με φωνητικά σύμβολα. Καθώς αυτή η χειροκίνητα σημαδεμένη δομή της εξάρτησης των φράσεων προσδιορίζεται χρησιμοποιώντας και συντακτικές και σημασιολογικές σχέσεις, μπορούμε μόνο να περιμένουμε ότι η συντακτική πληροφορία που εκδηλώνεται σιωπηρώς κατά το χωρισμό αυτό με τις παρενθέσεις, θα μπορέσει να γίνει αντιληπτή μέσω της διαδικασίας εκπαίδευσης του ΣΑΑΣΓ.

(2) Εισαγωγή παρενθέσεων σχετιζόμενες με την προσωδία και παρενθέσεων σχετιζόμενες με τις παύσεις.

Τα τονικά όρια των φράσεων και τα όρια των παύσεων ελήφθησαν υπόψη σε αυτό το δεύτερο επίπεδο εισαγωγής παρενθέσεων. Ακούγοντας την ομιλία και παρατηρώντας την αναλυμένη F₀ καμπύλη, μονάδες που το μέγεθός τους εξαρτάται από τον τονισμό των φράσεων, τέθηκαν εντός παρενθέσεων με χειροκίνητο τρόπο. Για ένα μοντέλο θέσεων κύριων συντακτικών ορίων, η εισαγωγή παρενθέσεων που χωρίζουν τη φράση βάσει προσωδιακών χαρακτηριστικών, έγινε αυτόματα βρίσκοντας τα σημεία της F₀ που παρατηρούνται *resets* (ρήσετς). Ως *reset* ορίζεται μια αύξηση της F₀ κατά μέσον όρο υπολογιζόμενη κατά μήκος δυο διαδοχικών τονικών φράσεων. Και για ένα μοντέλο εισαγωγής παύσεων, η εισαγωγή παρενθέσεων που σχετίζονται με τις παύσεις επιτεύχθηκε ομαδοποιώντας όλα τα συστατικά που κατατμώνται από τις παύσεις.

Υλικό κειμένου με παρενθέσεις εξάρτησης φράσεων, αρχικά χρησιμοποιήθηκαν για να εκπαιδεύσουν την ΣΑΑΣΓ από το μηδέν. Έπειτα αυτή η ΣΑΑΣΓ επανεκπαιδεύτηκε χρησιμοποιώντας το ίδιο υλικό κειμένου με προσωδιακές παρενθέσεις και μετά με παρενθέσεις παύσεων.

Ο προσδιορισμός ενός αποτελεσματικού συνόλου τερματικών συμβόλων κι ένας σωστός αριθμός μη-τερματικών συμβόλων είναι απαραίτητος για να πάρουμε μια ΣΑΑΣΓ. Λαμβάνοντας υπόψη τους περιορισμούς του μεγέθους των δεδομένων και το υπολογιστικό κόστος, τα ΜΤΛ και τα μεταθετικά μόρια επιλέχθηκαν ως τερματικά σύμβολα. Το γεγονός ότι συμπεριελήφθησαν τα μεταθετικά μόρια είχε ως αποτέλεσμα την αύξηση της ακρίβειας μιας ΣΑΑΣΓ, γιατί αυτά τα μόρια προσδιορίζουν τις συντακτικές ιδιότητες μιας φράσης. Τα μεταθετικά μόρια που προέκυψαν περισσότερες από 50 φορές στα δεδομένα της ομιλίας χρησιμοποιήθηκαν ως τερματικά σύμβολα. Το αποτέλεσμα ήταν να επιλεγθούν τα παρακάτω τέσσερα σετ τερματικών συμβόλων:

(1) Χαρακτηρισμοί ΜΤΛ μόνοι τους ($n = 23$):

13 είδη λέξεων που εκφράζουν νόημα (π.χ. επίθετο, επίρρημα, ρήμα, βοηθητικό ρήμα, ονομαστικό ουσιαστικό, ρηματικό ουσιαστικό, κύριο όνομα, επιθετικό ουσιαστικό, ποσοτικοποιητής, και αντωνυμία) και δέκα είδη λειτουργικών λέξεων (π.χ. βοηθητικό ρήμα, μόριο ζεύγους, συζευκτικό μόριο, εγκλιτικό μόριο, επιρρηματικό μόριο, τελικό μόριο, ομοιόβαθμο μόριο και τεταγμένο μόριο) χρησιμοποιούνται. Τρεις από τις λέξεις που εκφράζουν νόημα και μια από τις λειτουργικές κλίνονται.

(2) Χαρακτηρισμοί των ΜΤΛ ($n = 22$) συν οι ακόλουθοι γραμματικοί χαρακτηρισμοί:

(α) Επτά κλάσεις από μόρια ζεύγους:
“ga”, “no”, “ni”, “wo”, “de”, “to” και άλλα.

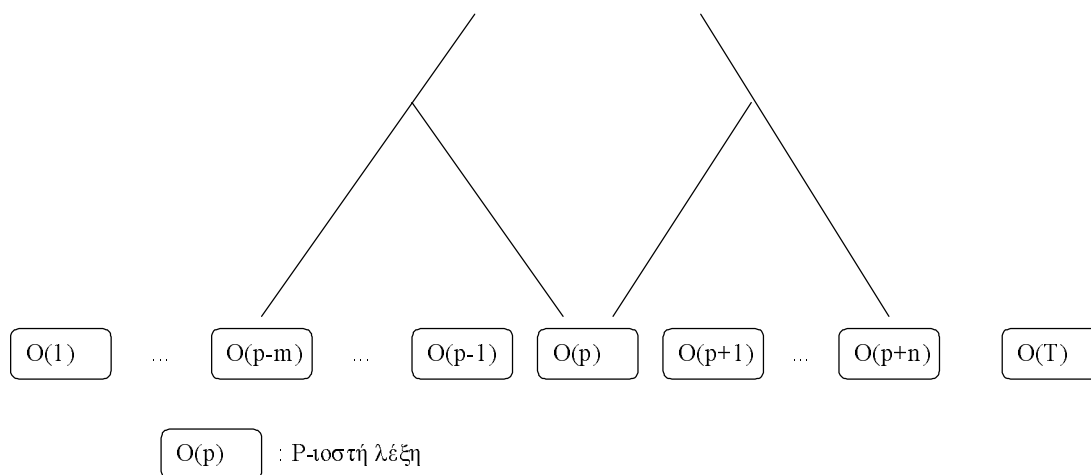
(β) Δυο κλάσεις από συνδετικά μόρια:
“te” και άλλα.

(γ) Δυο κλάσεις από εγκλιτικά μόρια:
“wa” και άλλα.

Υπολογισμός των Παραμέτρων για την Πρόβλεψη χρησιμοποιώντας μια ΣΑΑΣΓ.

Θεωρούμε δυο παραμέτρους P_m και Q_n οι οποίες παριστάνουν τις εξαρτήσεις των φράσεων που μπορούν να γίνουν αντιληπτοί από την ΣΑΑΣΓ. Στο σχήμα 6, η πιθανότητα της αριστερής διακλάδωσης P_m παριστάνει την πιθανότητα μιας λέξης να είναι μέρος μιας δομής αριστερής διακλάδωσης, η οποία περιλαμβάνει τις προηγούμενες m λέξεις. Παρόμοια, η πιθανότητα της δεξιάς διακλάδωσης Q_n παριστάνει την πιθανότητα, η λέξη να είναι μέρος μιας δομής δεξιάς διακλάδωσης, η οποία

περιλαμβάνει τις n επόμενες λέξεις. Αυτές οι πιθανότητες, παριστάνουν τις δομές εξάρτησης των φράσεων και μπορούν να υπολογισθούν χρησιμοποιώντας τις εσωτερικές/εξωτερικές πιθανότητες οι οποίες ορίζονται στον μέσα-έξω αλγόριθμο που περιγράφεται στην παράγραφο που ακολουθεί.



Σχήμα 6. Συντακτική δομή με αριστερή/δεξιά δομή διακλάδωσης.

Έστω $a(i, j, k)$ η πιθανότητα το μη τερματικό σύμβολο i να παράγει το ζεύγος των μη τερματικών συμβόλων j και k . Έστω $b(i, m)$ η πιθανότητα το μη τερματικό σύμβολο i να παράγει ένα τερματικό σύμβολο m . Στον μέσα-έξω αλγόριθμο (Lari K. and Young S. J. 1989), η εσωτερική πιθανότητα $e(s, t, i)$ ορίζεται ως η πιθανότητα ενός μη τερματικού συμβόλου i να παράγει την παρατήρηση $O(s), \dots, O(t)$ και μπορεί να εκφρασθεί ως εξής:

ΠΕΡΙΠΤΩΣΗ 1: $s = t$:

$$e(s, s, i) = b[i, O(s)], \quad (1)$$

ΠΕΡΙΠΤΩΣΗ 2: $s \neq t$:

$$e(s, t, i) = \sum_{j,k} \sum_{r=s}^{t-1} \alpha[i, j, k] \cdot e(s, r, j) \cdot e(r+1, t, k). \quad (2)$$

Η εξωτερική πιθανότητα $f(s, t, i)$ είναι η πιθανότητα, κατά τη διαδικασία επανεγγραφής, το i να παράγεται και τα αλφαριθμητικά στα οποία δεν κυριαρχεί να είναι $O(1), \dots, O(s-1)$ στα αριστερά και $O(t+1), \dots, O(T)$ στα δεξιά. Άρα:

$$f(s, t, i) = \sum_{j,k} \left[\sum_{r=1}^{s-1} f(r, t, j) \cdot \alpha(j, k, i) \cdot e(r, s-1, k) + \sum_{r=t+1}^T f(s, r, j) \cdot \alpha(j, i, k) \cdot e(t+1, r, k) \right] \quad (3)$$

$$\text{και } f(1, T, i) = \begin{cases} 1, & i = S(\text{start_symbol}) \\ 0, & \text{αλλιώς} \end{cases} \quad (4)$$

Το μη τερματικό σύμβολο i μπορεί να ακολουθεί δυο δυνατούς κανόνες $j \rightarrow ik$ ή $j \rightarrow ki$ $f_i(s, t, i)$ είναι η πιθανότητα όταν μόνο οι κανόνες $j \rightarrow ik$ λαμβάνονται υπόψη και

$f_r(s, t, i)$ είναι η πιθανότητα όταν μόνο οι κανόνες $j \rightarrow ki$ λαμβάνονται υπόψη. Αυτές, εκφράζονται ως εξής:

$$f_1(s, t, i) = \sum_{j,k} \sum_{r=t+1}^T f(s, r, j) \cdot \alpha(j, i, k) \cdot e(t+1, r, k) \quad (5)$$

$$f(s, t, i) = \sum_{j,k} \sum_{r=1}^{s-1} f(r, t, j) \cdot \alpha(j, k, i) \cdot e(r, s-1, k) \quad (6)$$

Η πιθανότητα η παρατήρηση $O(1), \dots, O(T)$ να έχει μια δομή αριστερής διακλάδωσης που περιλαμβάνει την παρατήρηση $O(s), \dots, O(t)$, και η πιθανότητα η παρατήρηση $O(1), \dots, O(T)$ να έχει μια δομή δεξιάς διακλάδωσης που περιλαμβάνει την παρατήρηση $O(s), \dots, O(t)$ δίνονται, αντίστοιχα, από τις παρακάτω σχέσεις:

$$\sum_i e(s, t, i) f_1(s, t, i), \quad (7)$$

$$\sum_i e(s, t, i) f_r(s, t, i). \quad (8)$$

Η πιθανότητα που παράγεται για ολόκληρη την παρατήρηση $O(1), \dots, O(s), \dots, O(t), \dots, O(T)$ είναι $e(1, T, S)$.

Επομένως, τα P_m και Q_n στην p -ιοστή λέξη, έχουν ως εξής:

$$P_m = \frac{\sum_i e(p-m, p, i) f_1(p-m, p, i)}{e(1, T, S)}, \quad (9)$$

$$Q_n = \frac{\sum_i e(p, p+n, i) f_r(p-m, p, i)}{e(1, T, S)}. \quad (10)$$

Μοντέλο πρόβλεψης με τη χρήση Νευρωνικού Δικτύου.

Ένα τροφοδοτούμενο από μπροστά νευρωνικό δίκτυο (feed-forward neural network) χρησιμοποιήθηκε για να προβλέπει τις θέσεις των κύριων ορίων των φράσεων και των θέσεων των παύσεων. Για την εκπαίδευση του νευρωνικού αυτού δικτύου, χρησιμοποιήθηκαν γρήγορες μέθοδοι διάδοσης προς τα πίσω (back-propagation) (Haffner P., Sawai H., Waibel A. and Shikano K. Mar. 1989). Το νευρωνικό αυτό δίκτυο έχει τέσσερα επίπεδα: ένα επίπεδο εισόδου με 50 μονάδες, δύο κρυφά επίπεδα με 25 μονάδες, και ένα επίπεδο εξόδου με 2 μονάδες.

Οι παράμετροι εισόδου έχουν ως εξής:

(1) P_m και Q_n στις ακόλουθες λέξεις (όπου $m, n = 1, 2, 3, 4$ και 5 ή μεγαλύτερα):

(α) η λέξη που εκφράζει νόημα και προηγείται της λέξης πριν το όριο,

(β) η λέξη πριν το όριο,

(γ) η λέξη μετά το όριο, και

(δ) η λέξη που εκφράζει νόημα και ακολουθεί τη λέξη μετά το όριο.

(2) Η κλάση των τερματικών συμβόλων από τα παρακάτω:

(α) οι πέντε λέξεις που προηγούνται του ορίου και

(β) οι πέντε λέξεις που ακολουθούν το όριο.

Οι παράμετροι εξόδου τέθηκαν στο 1 ή στο 0 υποδηλώνοντας την παρουσία ή την απουσία κύριου ορίου φράσης ή ορίου παύσης.

Το Μοντέλο Campbell

Η αρχή της ελαστικότητας (Campbell W. 1992) στην ισχυρότερή της μορφή λειπει ότι όλες οι διάρκειες των τμημάτων σε ένα συλλαβικό πλαίσιο λαμβάνονται από τον ίδιο παράγοντα z – ο επονομαζόμενος z -score (ζη-σκόρ) ή κανονικοποιημένη διάρκεια (Wightman C., Shattuck-Hufnagel S., Ostendorf M. and Price P.) – ως εξής:

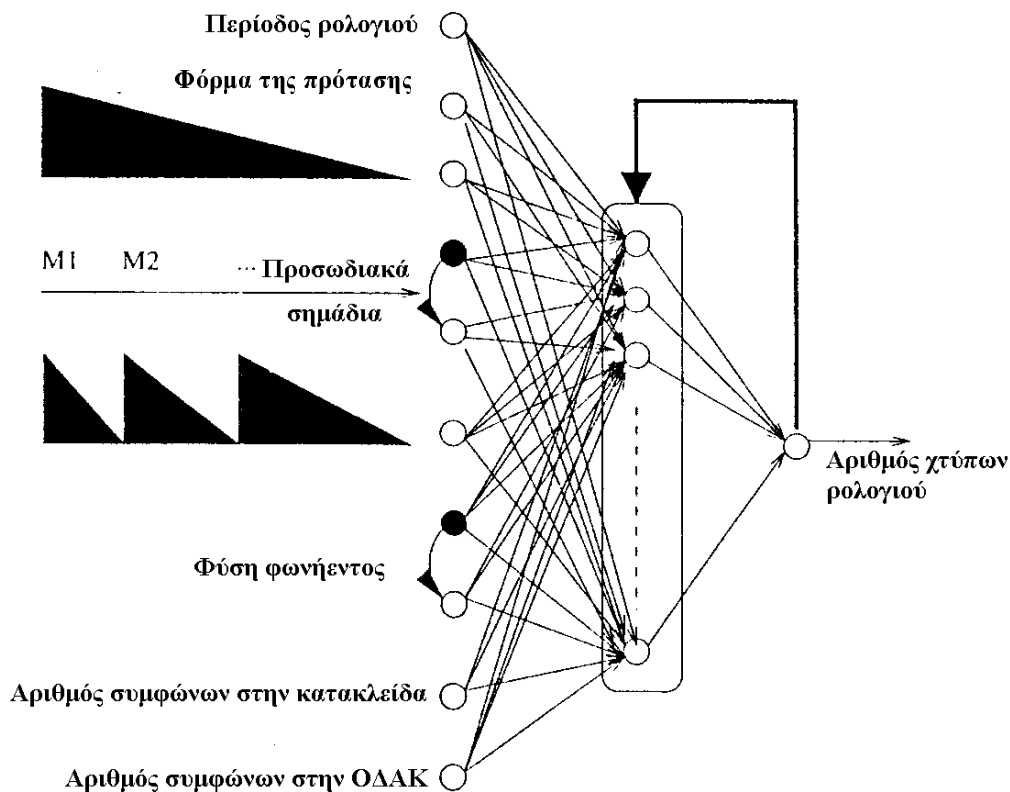
$$Dur_i = \exp(\mu_i + z\sigma_i), \quad (11)$$

όπου μ_i και σ_i είναι η μέση τιμή και τυπική απόκλιση του λογαρίθμου των διαρκειών (σε ms) των πραγματοποιήσεων του φωνήματος i . Αυτά τα z -scores υπολογίζονται στη συλλαβή από τον τύπο:

$$\sum_i Dur_i = \text{διάρκεια συλλαβής}. \quad (12)$$

Η διαδοχική χρήση των δυο παραπάνω εξισώσεων αναφέρεται ως αλγόριθμος ανακατανομής. Το μοντέλο του Campbell εξελίσσεται σε δυο στάδια:

- πρόβλεψη της διάρκειας της συλλαβής από φωνοτακτικές και φωνολογικές πληροφορίες από κάποιο στατιστικό μοντέλο (ένα πολυεπίπεδο perceptron)
- χρήση του αλγορίθμου ανακατανομής (Campbell W. 1992).



Σχήμα 7. Το σειριακό δίκτυο για την παραγωγή των θέσεων των ΚΑ σε μονάδες εσωτερικού ρολογιού. Το δίκτυο μετατρέπει απλές συναρτήσεις ράμπας που δηλώνουν το μήκος και τη συνάρτηση κάθε γλωσσολογικής μονάδας της ειπωθείσας φράσης, σε ρυθμικές καμπύλες ανάλογα με το ρυθμό ομιλίας και φωνοτακτικούς περιορισμούς. Οι κυψέλες εισόδου που παρουσιάζονται μαυρισμένες, αποθηκεύουν τις τιμές των συνδεδεμένων κυψελών με καθυστέρηση -1 , δηλαδή που ακολουθούν τα προσωδιακά σημάδια και τον επόμενο πυρήνα.

Στην είσοδο του δικτύου, περιγράφονται σειριακά, οι προσωδιακές και φωνολογικές πληροφορίες που φαίνονται να είναι σχετικές για την πρόβλεψη των διαρκειών του τρέχοντος ΔΔΑΚ. Δυο κυψέλες εισόδου παραμένουν σταθερές: (1) η συχνότητα του εσωτερικού ρολογιού και (2) η εγκλιτικότητα της πρότασης. Δυο άλλες κυψέλες αντιστοιχούν σε πιθανά βασικά ρυθμικά πρότυπα. Οι ράμπες δηλώνουν την έκταση των δυο βασικών προσωδιακών μονάδων: (3) η πρόταση, και (4) η προσωδιακή ομάδα⁵. Αυτές οι γραμμικές ράμπες, τίθενται αρχικά σε μια τιμή ίση με τον αριθμό των ΟΔΑΚ της μονάδας και φτάνουν στο μηδέν στο τέλος της μονάδας. Οι έξι κυψέλες που απομένουν τροφοδοτούνται με: (5) το τρέχον προσωδιακό σημάδι, (6) το επόμενο προσωδιακό σημάδι, (7) τη φύση του τρέχοντος φωνήεντος, (8) τη φύση του επόμενου φωνήεντος, (9) τον αριθμό των συμφώνων στο ΟΔΑΚ, και (10) την κατακλείδα της πρότασης – το τέλος της εκφώνησης της (ώστε να είναι σε θέση να διορθώσει την εκτίμηση του ΚΑ σε περίπτωση που εμφανιστεί παύση).

⁵ Η προσωδιακή ομάδα ορίζεται σε μια γλωσσολογική βάση: Αποτελείται από κάθε λέξη που εκφράζει νόημα και κάθε εξαρτώμενη λειτουργική λέξη. Οι προσωδιακές ομάδες είναι συνδεδεμένες από προσωδιακά σημάδια (Baillly G. 1989). Αυτά τα σημάδια κατατάσσονται ανάλογα με το βαθμό της συνοχής μεταξύ των γειτονικών προσωδιακών μονάδων.

Οι παράμετροι του μοντέλου ανακατανομής για τα Γαλλικά δίδονται στον πίνακα 5, για δύο ομιλητές του συστήματος μετατροπής κειμένου σε ομιλία για το οποίο γίνεται λόγος. Οι παράμετροι αυτοί προέκυψαν από στατιστική ανάλυση των διαρκειών των τμημάτων των λογάτομα που παράγονται απομονωμένα (ομιλητής FB) ή σε μια φέρουσα πρόταση (ομιλητής EV). Χρησιμοποιούνται μόνο δεδομένα από τα δυσλλαβικά και τρισυλλαβικά λογάτομα. Μολονότι ο FB είναι γρηγορότερος από τον EV, γίνονται αντιληπτά πολύ γνωστά χαρακτηριστικά των εγγενών διαρκειών των φωνημάτων: τα ρινικά φωνήεντα είναι μακρύτερα από τα στοματικά, τα φωνούμενα σύμφωνα (voiced consonants) είναι μακρύτερα από τα μη φωνούμενα κοκ.

phon	EV	FB	phon	EV	FB	phon	EV	FB
a	4.58 ± .30	4.74 ± .34	œ	4.61 ± .34	4.68 ± .34	ɔ	4.62 ± .38	4.69 ± .30
ɛ	4.43 ± .32	4.68 ± .27	Φ	4.65 ± .31	4.75 ± .27	ο	4.67 ± .32	4.81 ± .27
e	4.49 ± .33	4.73 ± .33	y	4.59 ± .40	4.61 ± .29	u	4.64 ± .40	4.69 ± .31
i	4.49 ± .38	4.55 ± .33	ʔ	4.81 ± .29	4.89 ± .31	˘ε	4.75 ± .28	4.93 ± .31
˘a	4.80 ± .30	4.89 ± .29	j	4.28 ± .23	4.47 ± .29	ɥ	4.33 ± .26	4.48 ± .23
w	4.36 ± .21	4.51 ± .25	∇	4.25 ± .19	4.48 ± .32	ə	4.16 ± .28	4.86 ± .43
l	4.16 ± .19	4.48 ± .26	t	4.51 ± .26	4.54 ± .55	k	4.60 ± .23	4.56 ± .43
p	4.60 ± .19	4.58 ± .53	d	4.28 ± .21	4.48 ± .23	g	4.32 ± .18	4.42 ± .24
b	4.34 ± .19	4.48 ± .24	z	4.39 ± .23	4.71 ± .20	ʒ	4.43 ± .23	4.62 ± .21
v	4.34 ± .24	4.59 ± .28	s	4.87 ± .24	5.26 ± .37	ʃ	4.81 ± .23	4.99 ± .25
f	4.73 ± .25	5.04 ± .38	n	4.35 ± .26	4.59 ± .24	η	4.75 ± .24	4.68 ± .23
m	4.46 ± .24	4.61 ± .23						

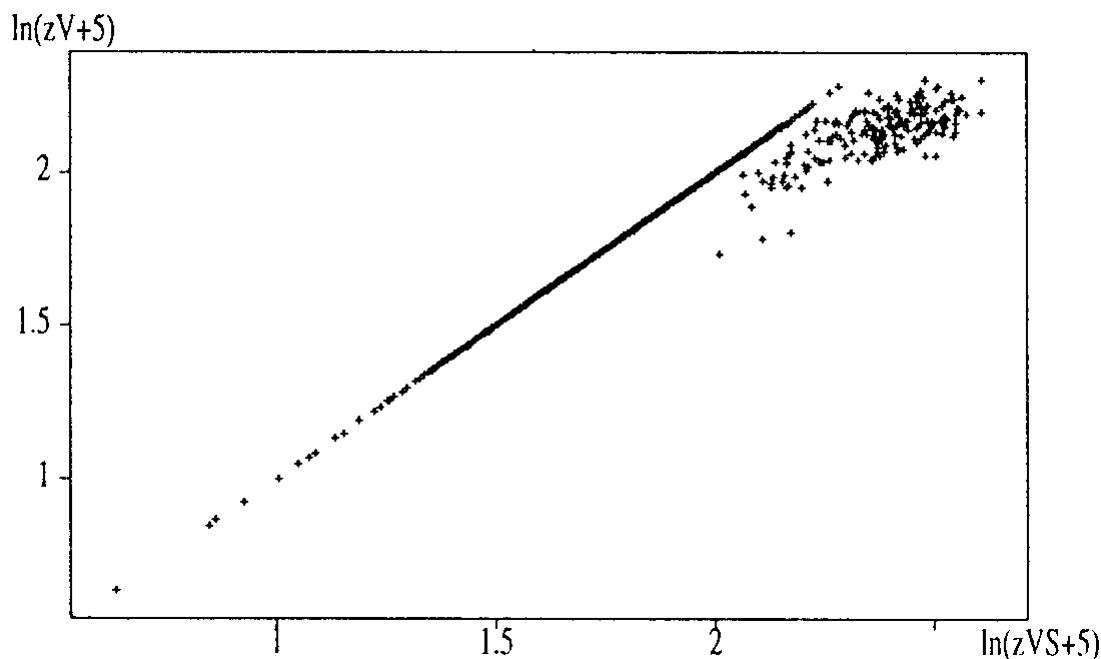
Πίνακας 5. Παράμετροι του z-score μοντέλου για δύο ομιλητές, τους EV και FB υπολογισμένους στα λογάτομα. Μέσες τιμές ± τυπικές αποκλίσεις των λογαρίθμων των διαρκειών των τμημάτων δίδονται.

Ο αλγόριθμος ανακατανομής τροποποιήθηκε ώστε να περιλαμβάνει την εμφάνιση των παύσεων. Γι' αυτό δοκιμάστηκε μια νέα διαδικασία για τις διάφορες εγγραφές της ίδιας πρότασης σε διαφορετικούς ρυθμούς ομιλίας.

Τα Δεδομένα Ομιλίας

Χρησιμοποιήθηκαν 20 προτάσεις οι οποίες ειπώθηκαν με πέντε διαφορετικούς ρυθμούς ομιλίας, ώστε να μελετηθεί η επιρροή της εμφάνισης των παύσεων στη συνολική ρυθμική δομή. Προκειμένου να απλοποιηθεί το πρόβλημα εύρεσης των ΚΑ, όλες οι προτάσεις είναι σειρές ΣΦ μεταπτώσεων (όλες οι παύσεις είχαν έτσι ένα φωνήεν πριν). Αποτελεσματικά διαχωρίσιμοι ρυθμοί επιτεύχθηκαν ελέγχοντας το ρυθμό ομιλίας του ομιλητή με συνθετικές ερωτήσεις. Στον ομιλητή δόθηκε η οδηγία να απαντά τις ερωτήσεις αυτές, με προτάσεις οι οποίες εμφανίζονταν σε μια οθόνη με τον ίδιο ρυθμό ομιλίας όπως των ερωτήσεων. Ο ρυθμός ομιλίας των συνθετικών ερωτήσεων ελέγχονταν πολλαπλασιάζοντας τους συντελεστές μ_i και σ_i κάθε φωνήματος με έναν παράγοντα, όπως προτάθηκε από τον Wightman και τους συνεργάτες του (Wightman C., Shattuck-Hufnagel S., Ostendorf M. and Price P.). Αυτή η διαδικασία ήταν ιδιαίτερος αποτελεσματική στη διαφοροποίηση των πέντε ρυθμών ομιλίας από πολύ αργό ρυθμό σε

πολύ γρήγορο κατά κάποιο συνεχή τρόπο. Οι μέσες τιμές των πέντε ρυθμών ομιλίας δίδονται στον πίνακα 6.



Σχήμα 8. Γραφική απεικόνιση του λογαρίθμου των πραγματικών z-score z_v κάθε φωνήεντος από τα δεδομένα ομιλίας, συναρτήσει του λογαρίθμου των φανταστικών z-score z_{vs} για όλους τους ρυθμούς ομιλίας.

Συγχωνεύοντας το φαινόμενο των Παύσεων στον Αλγόριθμο Ανακατανομής.

Προκειμένου να υπολογίσουμε τη διάρκεια της σιωπής που πρέπει να αποδοθεί σε κάθε ΔΔΑΚ στο στάδιο της παραγωγής του, έχει μελετηθεί η σχέση μεταξύ των πραγματικών z-scores των φωνηέντων $z_v = (\ln(\text{Dur}_v) - \mu_v) / \sigma_v$ και φανταστικών z-scores των ίδιων φωνηέντων. Αυτά τα φανταστικά z-scores υπολογίζονται προσθέτοντας στην πραγματική διάρκεια του φωνηέντος Dur_v τη διάρκεια του γειτονικού διαστήματος σιωπής που μπορούμε να επιλέξουμε, δηλαδή $z_{vs} = (\ln(\text{Dur}_{\text{φωνηέντος}} + \text{Dur}_{\text{σιωπής}}) - \mu_v) / \sigma_v$.

Όπως παρουσιάζεται στο σχήμα 8, $z_v = z_{vs}$ έως το κρίσιμο σημείο $z_{vs} \cong 2.4$. Έπειτα, μεταξύ των τιμών $z_{vs} \cong 2.4$ και $z_{vs} \cong 4.5$, ο ομιλητής ακολουθεί δυο στρατηγικές:

- (1) Ο ομιλητής συνεχίζει να επιμηκύνει το φωνήεν χωρίς παύση, το οποίο σημαίνει ότι παράγεται μια υποκειμενική παύση. Αυτό, αντιστοιχεί στο τέλος του τμήματος $z_v = z_{vs}$ που προαναφέρθηκε.
- (2) Ο ομιλητής εισάγει μια σιωπηρή παύση που αντιστοιχεί στα διάσπαρτα σημεία, και το οποίο γίνεται όλο και πιο πυκνό όσο το z_{vs} αυξάνει.

Τελικά, για μεγάλα z_{vs} , μόνο η δεύτερη στρατηγική παραμένει, και η παύση πάντοτε εισάγεται.

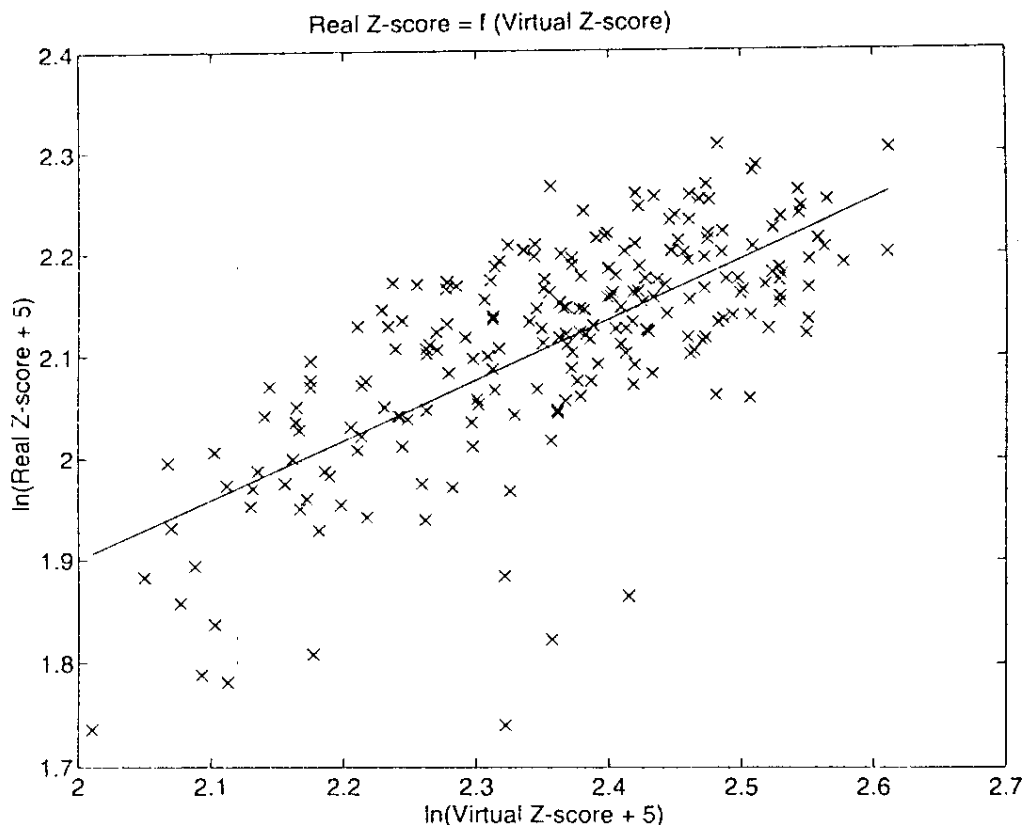
Το σχήμα 9, παρουσιάζει μια μεγέθυνση του τμήματος των διάσπαρτων σημείων του σχήματος 8. Το σχήμα 3 δείχνει ότι ανεξαρτήτως της εμφάνισης παύσεως, ο ομιλητής και πάλι επιμηκύνει το φωνηεντικό μέρος του ΟΔΑΚ. Το μοντέλο της εμφάνισης των παύσεων χρησιμοποιεί την ευθεία (βλ. παρακάτω εξίσωση) για $z_v \neq z_{vs}$ ως το υπολογιστικό μοντέλο της διάρκειας των παύσεων.

$$(z_v + 5) = (z_{vs} + 5)^{0.59} \exp(0.72) \quad (13)$$

rate	Vowels			Consonants			Silences		
	SY	VO	PC	SY	VO	PC	SY	VO	PC
v. slow	34 ± 32	35 ± 32	34 ± 29	43 ± 46	42 ± 43	37 ± 46	67 ± 58	76 ± 62	55 ± 53
slow	38 ± 40	42 ± 41	36 ± 32	32 ± 32	34 ± 31	28 ± 26	73 ± 52	73 ± 52	73 ± 56
normal	26 ± 26	30 ± 38	29 ± 23	27 ± 26	28 ± 26	25 ± 22	54 ± 37	64 ± 36	56 ± 35
fast	17 ± 19	18 ± 19	19 ± 20	21 ± 25	21 ± 27	19 ± 21	75 ± 77	68 ± 99	49 ± 78
v. fast	12 ± 12	11 ± 13	12 ± 14	13 ± 13	12 ± 17	11 ± 12	59 ± 33	93 ± 28	44 ± 20

Πίνακας 6. Μέσες τιμές ± τυπικές αποκλίσεις των απόλυτων λαθών πρόβλεψης για τρεις ηχητικές κλάσεις (φωνήεντα [vowels], σύμφωνα [consonants], διαστήματα σιωπής [silences]) που χρησιμοποιούν τον τροποποιημένο αλγόριθμο ανακατανομής στα αρχικά δεδομένα. Η κατανομή των λαθών συγκρίνεται με τρεις διαφορετικές ρυθμικές μονάδες: τη συλλαβή (SY), την ομάδα διαφωνηεντικής έναρξης (VO), και την ΟΔΑΚ (PC). Ο ρυθμός (rate) κυμαίνεται από πολύ αργός (v. slow) έως πολύ γρήγορος (v. fast), πέντε διαβαθμίσεις συνολικά.

Μια παύση μπορεί θεωρητικά λοιπόν να “εμφανισθεί” όταν αυτή η γραμμή διασταυρωθεί με τη γραμμή $z_v = z_{vs}$. Το κρίσιμο σημείο έτσι θα είναι το $z_{vsc} = 0.79$. Δεν παράγεται καμία παύση μεταξύ του z_{vsc} και του ελάχιστου $z_{vs} \cong 2.4$ γιατί ένα ελάχιστο διάστημα σιωπής απαιτείται ώστε να αναπαυθούν οι αρθρωτές της ομιλίας. Μια ελάχιστη διάρκεια σιωπής τέθηκε για κάθε ρυθμό ομιλίας ως αποτέλεσμα μιας ανάλυσης που έγινε πάνω στα δείγματα ομιλίας που χρησιμοποιήθηκαν (πίνακας 6). Η ελάχιστη αυτή διάρκεια δεν μεταβάλλεται ιδιαίτερα συναρτήσει του ρυθμού ομιλίας και παραμένει γύρω στα $Dur_{\text{σιωπής}} \cong 60\text{ms}$.



Σχήμα 9. Γραφική απεικόνιση του λογαρίθμου των πραγματικών (real) z-score z_v , συναρτήσει του λογαρίθμου των φανταστικών (virtual) z-score z_{vs} για $z_{vs} = z_v$.

Ο τροποποιημένος αλγόριθμος ανακατανομής εξελίσσεται ως εξής:

- Υπολογισμός του z-score για ένα δεδομένο ΟΔΑΚ (z_g)
- Αν το z_g είναι μεγαλύτερο ή ίσο από το κρίσιμο z_{vsc} , η διαδικασία τελειώνει: οι διάρκειες των τμημάτων λαμβάνονται χρησιμοποιώντας την εξίσωση (11).
- Αν το z_g είναι μεγαλύτερο από z_{vsc} , το z-score του φωνήεντος z_v λαμβάνεται από την εξίσωση (11), θέτοντας $z_{vs} = z_g$.
- Οι διάρκειες των τμημάτων υπολογίζονται με τον αλγόριθμο ανακατανομής με $z_g = z_v$ και προστίθενται. Η διαφορά μεταξύ αυτού του αποτελέσματος και της γνήσιας διάρκειας της ΟΔΑΚ, δίνει τη διάρκεια της σιωπής.
- Αν η διάρκεια της σιωπής είναι μεγαλύτερη από την ελάχιστη η διαδικασία τελειώνει.

Αν όχι, δεν εισάγεται καμία σιωπή και το z-score της ΟΔΑΚ διατηρείται ίσο με z_g .

Είναι σημαντικό να σημειωθεί ότι κανένας περιορισμός στη θέση – όπως η θέση τονισμού – επιβάλλεται στα διαστήματα σιωπής. Τοποθετούνται με βάση το αποτέλεσμα του τροποποιημένου αλγόριθμου ανακατανομής, του οποίου η περιγραφή έγινε παραπάνω. Ο τροποποιημένος αυτός αλγόριθμος, εφαρμόστηκε στα καθαρά δεδομένα για πέντε διαφορετικούς ρυθμούς ομιλίας. Η κατανομή των λαθών δίδεται στον πίνακα 7. Ο πίνακας αυτός, δείχνει ότι η ΟΔΑΚ προβλέπει τις διάρκειες των συμφώνων και των παύσεων με περισσότερη ακρίβεια απ'ότι άλλες ρυθμικές μονάδες όπως η συλλαβή ή η ομάδα διαφωνηεντικής έναρξης (*inter-vocalic onset group*).

Ρυθμός	Ρολόι (ms)	Ελάχιστη παύση (ms)	Λάθη στις θέσεις	Πραγματικά διαστήματα σιωπής
πολύ αργά	360	75	15	99
αργά	325	67	13	64
φυσιολογικά	270	57	11	32
γρήγορα	210	51	-	26
πολύ γρήγορα	165	59	1	5

Πίνακας 7. Αριθμός λαθών που προκύπτουν από τον αλγόριθμο ανακατανομής στην τοποθέτηση των παύσεων. Ο αριθμός των πραγματικών διαστημάτων σιωπής, ανάλογα με το ρυθμό ομιλίας στις φυσικές φράσεις επίσης δίδεται (υπάρχουν 285 ΟΔΑΚ στα δεδομένα που διαθέτουμε).

Χρησιμοποιώντας τις αρχικές διάρκειες των ΟΔΑΚ, μόνο μερικά από τα διαστήματα σιωπής τοποθετήθηκαν σε θέσεις που δεν επέλεξε ο ομιλητής (πίνακας 7). Όλες όμως οι θέσεις, αντιστοιχίστηκαν σε κάποιο λανθάνων σημείο για την πραγματοποίηση του τονισμού (προσωδιακό σημάδι).

Αυτόματη Εκμάθηση.

Όπως περιγράφηκε και προηγουμένως, ένα σειριακό δίκτυο εκπαιδεύτηκε προκειμένου να παράγει τα διαδοχικά ΔΔΑΚ για κάθε πρόταση των δεδομένων ομιλίας που διαθέτουμε. Πενήντα προτάσεις (δέκα προτάσεις σε πέντε ρυθμούς η κάθε μια) από τα δεδομένα, επιλέχθηκαν για τη φράση εκμάθησης. Το δίκτυο γενικεύει με ιδιαίτερη επιτυχία τις αντίστοιχες σειρές των ΔΔΑΚ για όλες τις προτάσεις των δεδομένων μας (συγκρίνετε τις τυπικές αποκλίσεις των λαθών μεταξύ των προτάσεων εκμάθησης και δοκιμών στον πίνακα 8). Συνθετικές διάρκειες τμημάτων και σιωπής, ελήφθησαν εφαρμόζοντας τον τροποποιημένο αλγόριθμο ανακατανομής που περιγράφηκε προηγουμένως.

Τα λάθη μεταξύ των πραγματικών διαρκειών των τμημάτων και αυτών που προκύπτουν από το μοντέλο που μελετάμε, υπολογίστηκαν για κάθε ρυθμό ομιλίας και για τους δυο τύπους μονάδων: τμήματα και διαστήματα σιωπής (πίνακας 8). Τα ιστογράμματα των λαθών αυτών, για όλες τις προτάσεις του συνόλου εκμάθησης, συσχετίζονταν ισχυρώς με την κανονική κατανομή (ελάχιστο 95%).

Διεξήχθη ένα πείραμα για να δοκιμασθεί η αντιληπτική συμπεριφορά της οργάνωσης των ΟΔΑΚ. Το ερέθισμα λήφθηκε από μια τεχνική ανάλυσης κι

επανασύνθεσης που επονομάζεται PSOLA (Charpentier F. and Moulines E. 1990, Bailly G., Barbe T. and Wand H. 1992), στην οποία τροποποιήθηκαν⁶ μόνο οι διάρκειες των τμημάτων. Δύο εκδόσεις από κάθε ειπωθείσα φράση συγκρίθηκαν: η πρώτη (σημειωμένη ως *μοντέλο*), από το μοντέλο που μελετάμε και η δεύτερη (σημειωμένη ως *τυχαία*) λαμβάνεται προσθέτοντας Γκαουσιανό θόρυβο στις πραγματικές διάρκειες των τμημάτων. Η κατανομή του θορύβου ισούται με την κατανομή της πρόβλεψης των λαθών που προκύπτει από το μοντέλο που μελετάμε. Τρεις διαφορετικές κατανομές χρησιμοποιούνται για τα φωνήεντα, τα σύμφωνα, και τα διαστήματα σιωπής (πίνακας 6). Οι δύο εκδόσεις διαφέρουν μόνο στη χρονική δομή των λαθών: Η έκδοση του *μοντέλου* αναμένεται να προβλέπει με περισσότερη ακρίβεια τη σειρά των ΚΑ.

Ρυθμός	Σύνολο εκμάθησης		Σύνολο δοκιμής	
	Διαστήματα σιγής	Τμήματα	Διαστήματα σιγής	Τμήματα
πολύ αργά	-41(137)	-1(49)	3(207)	4(56)
αργά	70(116)	-2(40)	-82(147)	6(46)
φυσιολογικά	67(122)	0(37)	-105(113)	5(43)
γρήγορα	-189(84)	3(30)	64(144)	0(28)
πολύ γρήγορα	-4(170)	2(25)	-49(193)	-1(23)

Πίνακας 8. Μέσες τιμές (και τυπικές αποκλίσεις) σε milliseconds από τα ιστογράμματα των λαθών μεταξύ των χωρισμένων σε τμήματα και των παραγομένων διαρκειών για τα σύνολα των προτάσεων εκμάθησης και δοκιμής.

Κατακλείδα

Υπολογίζεται πως η σύνθεση του λόγου θα φαίνεται πιο ομαλή, θα είναι πιο ευχάριστη ακουστικά, και πιθανότατα θα είναι πιο κατανοητή όταν οι σιωπηρές και οι γεμάτες παύσεις θα συναρτώνται συστηματικά στη ροή του λόγου. Καθώς οι παύσεις αυτές περικλείουν και προσωδιακά τμήματα - τα οποία θεωρήσαμε ισοδύναμα με τις εκτελεστικές δομές - οι παύσεις μπορούν να προβλεφθούν με τους ίδιους μηχανισμούς που μας επιτρέπουν να προβλέψουμε τις εκτελεστικές δομές. Έχουν αναπτυχθεί οι αλγόριθμοι εκείνοι που μπορούν να δοκιμάσουν τις προβλέψεις αυτές, και οι οποίοι δοκιμάζονται τώρα σύμφωνα με διάφορους τύπους υλικού ομιλίας.

Οι εξελίξεις είναι μέρος μιας μεγαλύτερης προσπάθειας να αναπτυχθούν μοντέλα ικανά να ρυθμίσουν διάφορους τομείς στην ευφράδεια του λόγου στα συστήματα από κείμενο-σε-ομιλία. Ζητήματα που αφορούν τις παύσεις καθώς και τις διορθώσεις στα λάθη ομιλίας είναι εξίσου σημαντικά με την αυτόματη αναγνώριση της ομιλίας, μια και τα μελλοντικά συστήματα για την κατανόηση του αυθορμήτου λόγου θα πρέπει να αποδείξουν πως έχουν την επαρκή νοημοσύνη για να χειριστούν τη διόρθωση του

⁶ Η εισαγωγή/διαγραφή βραχύβιων σημάτων κυριαρχείται από την εμφάνιση λειτουργιών που υπολογίζονται με χρονική αποσύνθεση (Bailly G., Marteau P.F., Arby C. 1989). Οι τιμές της θεμελιώδους συχνότητας καταγράφονται μόνο με τρεις τιμές ανά τμήμα: αρχική, μέση και τελική, και η ενέργεια καταγράφεται από μια μέση τιμή.

λόγου, και μπορούν να ωφεληθούν από τα σχήματα κανονικότητας της διάρκειας των παύσεων και των συλλαβών που βρίσκονται στη χρονική δομή της ομιλίας.

ΑΝΑΦΟΡΕΣ - ΒΙΒΛΙΟΓΡΑΦΙΑ

- Allen, G.D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics*, 3, 75-86.
- Allen, J., Hunnicutt, M.S., & Klatt, D. (1987). From text to speech. *The MITalk system*. Cambridge, England: Cambridge University Press.
- Bachenko J. & Fitzpatric E. (1990). Acomputational grammar of discourse-neutral prosodic phrasing in English. *Computational Linguistics*, 16, 155-170.
- Barbosa ,P. & Bailly, G. (1993). Generation and evaluation of rythmic patterns for text-to-speech synthesis. *Proceedings of an ESCA Workshop on Prosody (pp.66-99)*. Lund, Sweden
- Barbosa ,P. & Bailly, G. (1994). Characterization of rythmic patterns for text-to-speech synthesis. *Speech Comm.* 15:127-137, 1994.
- Bailly G. (1989). Integration of rythmic and syntactic constraints in a model of generation of French prosody. *Speech Comm.* 8:137-146.
- Bailly G., Marteau P.F., and Arby C. (1989). A new algorithm for temporal decomposition of speech. Application to a numerical model of coarticulation. In *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, 508-511.
- Bailly G., T. Barbe and Wang H. (1992). Automatic labelling of large prosodic databases: tools, methodology and links with a text-to-speech system. In *Talking Machines: Theories, Models and Designs*, Bailly G. and Benoit C., eds. Elsevier B.V., North Holland, Amsterdam, 323-333.
- Bloodstein, O. (1981). *Handbook on stuttering*. Chicago: National Easter Seal Society for Crippled Children and Adults.
- Boomer, D.S. & Ditmann, A.T. (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech*, 5 ,215.
- Botte, M.C., Canevet, G., Demany, L., & Sorin, C. (1989). Psychoacoustique et perception auditive. Serie audition. Paris: Inserm / Sfa / CNET.
- Butcher, A. (1980). Pause and syntactic structure. In W.Deichert & M.Raupach (Eds.), *Temporal variables in speech* (pp. 86-90). Mouton.
- Caelen – Haumont, G. (1991). Strategies des locuteurs et consignes de lecture d'un texte: *analyse des interactions entre modeles syntaxiques, semantiques, pragmatique et parametres prosodiques*. Theses d'Etat, Aix-en-Provence.

- Campbell W. (1992). *Multi-level Timing in Speech*. Ph.D. thesis, University of Sussex, Sussex UK.
- Charpentier F. and Moulines E. (1990). Pitch-synchronous waveform processing techniques for text-to-speech using diphones. *Speech Comm.* 9(5-6):453-467.
- Cutler A. and Norris D. (1988). The role of strong syllables in segmentation for lexical access. *J. Experimental Psychology: Human Perception and Performance* 14:113-121.
- Chafe, W. (1980). Some reasons for hesitating. In W.Deichert & M.Raupach, (Eds.), *Temporal variables in speech* (pp. 169-180). Mouton.
- Cook, M., Smith, J. & Lalljee, M. (1974). Filled pauses and syntactic complexity. *Language and Speech*, 17, 11-16.
- Cooper, W., & Patricia Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.
- Dauer, R.M. (1983). Stress-timing and syllable timing reanalyzed. *Journal of Phonetics*, 11, 51-62.
- Deichert, W., & Raupach, M. (Eds.), *Temporal variables in speech*. Mouton.
- Fraisse, P. (1974). *La psychologie du rythme*. PUF, Paris.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 2, 233-253.
- Fonagy, I. (1992). Fonctions de la duree vocalique. In P.Martin (Ed.), *Melanges Leon* (pp.141-164). Editions Melodie-Toronto.
- Fujimura, O. (1981). Temporal organization of articulatory movements as a multidimensional phrasal structure. *Phonetica*, 38, 66-83.
- Gee, J.P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411-458.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.
- Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, 15, 103-113.

- Gregoire, A. (1899). Variation de la duree de la syllable en francais. *La Parole*, 1, 161-176.
- Grosjean, F. (1980a). Linguistic structures and performance structures: Studies in pause distribution. In W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp.91-106). Mouton.
- Grosjean, F. (1980b). Comparative studies of temporal variables in spoken and sign languages: A short review. In W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp.307-312). Mouton.
- Grosjean, F. & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du francais. *Phonetica*, 31, 144-184.
- Grosjean, F. & Dommergues, J.Y. (1983). Les structures de performance en psycholinguistique. *L'Annee psychologique*, 83, 513-536.
- Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11, 58-81.
- Hirst, D., & Di Cristo, A. (in press). *Intonation systems: A survey of twenty languages*.
- Hirose K., Fujisaki H., Kawai H., and Yamaguchi (1990). Manifestation of linguistic and para-linguistic information in the voice fundamental frequency contours of spoken Japanese. In Proc. ICSLP, pp.485-488.
- Hakota K. and Sato H. (1980). Prosodic rules in connected speech synthesis. Trans. IECE Japan, J63-D:715-722 (στα Ιαπωνικά).
- Haffner P., Sawai H., Waibel A., and Shikano K. (March 1989). Fast backpropagation learning methods for large phonemic neural networks. In Rec. Spring Meeting, Acoust. Soc. Jpn, pp.27-28.
- Jordan M.I. (1989). Serial order: a parallel, distributed processing approach. In *Advances in Connectionist Theory: Speech*, Elman J.L. and Rumelhart D.E., eds. Lawrence Erlbaum, Hillsdale, NJ.
- Kaiki N. and Sagisaka Y. (1992). Pause characteristics and local phrase dependency structure in Japanese. In Proc. ICSLP, pp.357-360.
- Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208-1221.
- Keller, E. (1990). Speech motor timing. In W.J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp.343-364). Kluwer Academic Publishers.

- Keller, E., Zellner, B., Werner, S., & Blanchoud, N. (1993). The prediction of prosodic timing: Rules for final syllable lengthening in French. *Proceedings ESCA Workshop on Prosody* (pp.212-215). September 27-29. Lund, Sweden.
- Keller, E. (1994). Fundamentals of speech synthesis and speech recognition, Zellner B.: *Pauses and the temporal structures of speech*, John Willey & Sons.
- Lacheret-Dujour, A. (1991). Le debit de la parole: un filtre utilise pour la generation des variantes de prononciation en francais parisien. *Actes du XIIeme Congres International des Sciences Phonetiques* (pp.194-197). Aix en Provence.
- Lari K. and Young S. J. (1989). The estimation of stochastic context-free grammars using the inside-outside algorithm. *Computer speech and Language*, 4:35-56.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018-2024.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Martin, Ph. (1980). L'intonation est-elle une structure congruente a la syntaxe? In M.Rossi et.al. (Ed.), *L'intonation: de l'acoustique a la semantique* (pp.234-271). Klincksieck.
- Martin, Ph. (1992). Il etait deux fois l'intonation. In P.Martin (Ed.), *Melanges Leon* (pp.293-304). Editions Melodie-Toronto.
- Miller, J.Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, 41, 215-225.
- Monnin, P, & Grosjean, F. (1993). Les structures de performance en francais: caracterisation et prediction. *L'Annee Psychologique*, 93, 9-30.
- Morlec Y., Auberge V. and Bailly G. (1995). Evaluation of automatic generation of prosody with a superposition model. International Congress of Phonetic Sciences. Stockholm, Sweden.
- Morton J., Marcus S. and Frankish C. (1976). Perceptual Centers (p-centers). *Psychological Revue* 83(5):405-408.
- O'Shaughnessy, D. (1981). A study of French vowel and consonant durations. *Journal of Phonetics*, 9, 385-406.
- O'Shaughnessy, D. (1984). A multispeaker analysis of durations in read French paragraphs. *Journal of the Acoustic Society of America*. 76, 1664-1672.

- Pereira F. and Schabes Y. (1992). Inside-outside reestimation from partially bracketed corpora. In Proc. ACL, pp.128-135.
- Pfauel, M.C. (1986). *Entre begue*. Paris: Retz.
- Sagisaka Y. and Kaiki N. (1992). Optimization of intonation control using statistical Fo resetting characteristics. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processes, pp.49-52.
- Sagisaka Y. and Pareira F. (March 1994). Inductive learning of prosodic phrasing characteristics using stochastic context-free grammar. In Rec. Spring Meeting, Acoust. Soc. Jpn., pp.225-226.
- Suzuki K. and Saito T. (February 1995). N-phrase parsing method for Japanese text-to-speech conversion and assignment of prosodic features based on N-phrase structures. Trans. IEICE Japan, J78-D-II:177-187 (στα Ιαπωνικά).
- Sagisaka Y., Takeda K., Abe M., Katagiri S. Umeda T. and Kuwabara H. (1990). A large-scale Japanese speech database. In Proceedings of the International Conference on Spoken Language Processing, Kobe, Japan, pp.1089-1092
- Selkirk, E.O. (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press, Cambridge, MA.
- Starkweather, W.C. (1987). *Fluency and stuttering*. Prentice Hall.
- Wightman C., Shuttuck-Hufnagel S., Ostendorf M. and Price P. (1992). Segmental durations in the vicinity of prosodic boundaries. *J. Acoust. Soc. Amer.* 91(3):1707-1717.
- Zellner-Bechel, B. (1992). Le be begayage et euh... l'hesitation en francais spontane. *Actes des 19eme J.E.P.* (pp.481-487). Bruxelles.
- Zwicker, E., & Feldkeller, R. (1981). *Psychoacoustique: l'oreille recepteur d'information. Collection technique et scientifique des telecommunications*. Paris: Masson.

Για την ανάλυση, τον εμπουτισμό και τη μετάφραση των πρωτοτύπων κειμένων στην Ελληνική γλώσσα χρησιμοποιήθηκε επίσης η ακόλουθη βιβλιογραφία ...

Gimson, A.C., revised by Susan Ramsaran (1989). An introduction to the pronunciation of English. Edward Arnold. Fourth Edition.

Μπαμπινιώτης, Γ. (1980). Θεωρητική Γλωσσολογία, Εισαγωγή στη σύγχρονη γλωσσολογία. Αθήνα.

Readers Digest (1987). Universal Dictionary. Scotland.

Readers Digest Oxford (1993). Complete word finder, a unique and powerful combination of dictionary and thesaurus. Oxford.

Sagisaka Y., Campbell N., Higuchi N. (1997). Computing prosody. Shigeru F., Yoshinori S., Higushi N.: Prediction of Major Boundary Location and Pause Insertion Using a Stochastic Context-free Grammar. Springer-Verlag. N.Y.

Van Santen Jan P.H., Sproat R.W., Olive J.P., Hirshberg J. (1997). Progress in speech synthesis. Barbosa P.A., Bailly G.: *Generation of pauses within the z-score model*. Springer-Verlag. N.Y.

Yule, G. (1993). The Study of Language. Cambridge university press. Great Britain.

Yule, G., & Brown G. (1996). Discourse analysis. *Pause defined units* 160-164. Cambridge university press. Great Britain.

Χρήσιμες σχετικές πηγές (ενημέρωση, software) στο παγκόσμιο διαδίκτυο (Internet) μπορούν να βρεθούν στις ακόλουθες διευθύνσεις (World Wide Web) ...

<http://www-speech.sri.com/people/stolcke/papers/icassp96/paper.html>

<http://www.sil.org/linguistics>

<http://www.cisnet.or.jp/home/rlrose/pause/links.html>

<http://www.elsenet.org>

<http://www.hd.uib.no/Acohum/cl>

<http://www.cs.columbia.edu/~radev/u/db/acl>

ΠΙΝΑΚΑΣ ΑΓΓΛΙΚΩΝ ΟΡΩΝ

accent	1) προφορά 2) τονισμός	1) τοπική παραλλαγή προφοράς μιας γλώσσας που οφείλεται στη χρήση διαφορετικών τύπων φωνηέντων και συμφώνων καθώς και στα φωνολογικά και προσωδιακά πρότυπα που παράγει ο ομιλητής. Διακρίνεται από τη διάλεκτο. 2) το μέρος της λέξης, φράσης ή πρότασης στο οποίο δίδεται αυξημένη ακουστότητα καθώς και μια αλλαγή στο πρότυπο του επιτονισμού.
adverbial particle	επιρρηματικό ουσιαστικό	
adnominal particle	ομοιόβαθμο μόριο	
adjective	επίθετο	
adjectival noun	επιθετικό ουσιαστικό	
allophone	αλλόφωνο	ο συστηματικός διαφορετικός τρόπος εκφοράς ενός φωνήματος σε μια γλώσσα.
anchor points	ορμητήρια / σημεία εκκίνησης	
articulation	άρθρωση	
auditory	ακουστικός	
auxiliary noun	βοηθητικό ρήμα	
back-propagation	διάδοση προς τα πίσω	
bisection compounds	διχοτομημένες συνθέσεις	
bracketing structure	παρενθετική δομή	
case particle	μόριο ζεύγους	
co-articulation	συνάρθρωση	
coda	κατακλείδα	το τέλος της εκφώνησης μιας πρότασης
conjunctive particle	συζευκτικό μόριο	
cognitive	γνωστικός	
cohesion	ενότητα	
consonant	σύμφωνο	
constituents	συστατικά	
continuous speech	συνεχής ομιλία	
content word	λέξεις που εκφράζουν νόημα	
convergent	συγκλίνων	
co-occurrence	συνεμφάνιση	
coordinate	τεταγμένο μόριο	
corpus	σώμα / υλικό κειμένου	
corpora	πληθυντικός του corpus	
correlation	συσχέτιση	στατιστικός όρος.
determiner	άρθρο	
discourse	λόγος, ομιλία	
discourse analysis	ανάλυση λόγου	

dysfluent	δύστοκος	όχι ευφράδης
emphasis	έμφαση	
extralinguistic	εξωγλωσσικός	
feed-forward	τροφοδοτούμενο από εμπρός	
extrinsic	εξωγενής	
fluent	ευφράδης	
freestanding adjectives	ανεξάρτητα επίθετα	
final particle	τελικό μόριο	
foot	1) πόδι 2) μέτρο (όρος της μουσικής που έχει να κάνει με τον ρυθμό)	
generative syntax	γενετική μετασχηματιστική θεωρία	δες υποσημείωση στο τέλος του πίνακα.
hand-tagged	χειροκίνητο σημάδεμα	δηλαδή όχι με κάποιο αλγόριθμο
homomorphic	ομομορφικός	
human-like	που μοιάζει με την ανθρώπινη	
inspiration	εισπνοή	
intelligible	κατανοητός	
inter-lexical	διαλεξικός	μεταξύ των λέξεων
inter-perceptual center group	ομάδα δι-αντιληπτικού κέντρου	μια γνήσια ρυθμική μονάδα η οποία περιορίζεται από προεξέχοντα ακουστικά γεγονότα.
inter-perceptual center interval	διάστημα δι-αντιληπτικού κέντρου	
intervening	εμβόλιμος	
inter-vocalic onset group	ομάδα διαφωνηεντικής έναρξης	
intonation	επιτονισμός	
intrasegmental	ενδοτμηματικός	
intrinsic	ενδογενής	
isomorphic	ισομορφικός	
laryngo-phonatory reflex	λαρυγγοφωνητικό αντανακλαστικό	
lax vowel	χαλαρό φωνήεν	
lexemes	λεξήματα	δες morphemes
lexical	λεξικός	ο χαρακτηρίζει μια λέξη
lexical words	λεξικές λέξεις /κυρίως λέξεις	οι λέξεις που εκφράζουν κάποιο νόημα (αυτόνομα) και από τις οποίες εξαρτώνται άλλα δευτερεύοντα τμήματα του λόγου.
linguistic	1) γλωσσικός 2) γλωσσολογικός	
logatoms	λογάτομα	λόγια από διαφορετικά άτομα

model	μοντέλο	
modality of a sentence	φόρμα μιας πρότασης	
modal particle	εγκλιτικό μόριο	
morphemes	μορφήματα	το μορφολογικό εκείνο τμήμα της γλώσσας που αποδίδει κάποιο νόημα και που δεν μπορεί να χωριστεί σε περαιτέρω τμήματα. Ένα μόρφημα διαχωρίζεται βάσει των λειτουργιών του μέσα στο γλωσσολογικό σύστημα.
muscular	μυϊκός	
natural sounding speech	ομιλία που μοιάζει στο άκουσμά της φυσική	
nasal vowel	ρινικό φωνήεν	
neural network	νευρωνικό δίκτυο	
neural pulsation	νευρωνικός παλμός	
neutral phrasing	ουδέτερος τρόπος διατύπωσης	
nominal noun	ονομαστικό ουσιαστικό	
noun	ουσιαστικό	
noun phrase	ουσιαστική φράση	που έχει ως κύριο συστατικό το ουσιαστικό και γύρω από αυτό δομούνται τα υπόλοιπα τμήματα του λόγου π.χ. αντωνυμίες, επίθετα κ.τ.λ.
nucleus / nuclei	πυρήνας / πυρήνες	
onset	εκκίνηση, έναρξη	
oral vowel	στοματικό φωνήεν	
particle	μόριο	
part of speech	μέρος του λόγου	
parsing	γραμματική ανάλυση	
pattern	πρότυπο, μορφή	
perceptual center	κέντρα αντίληψης	
perception	αντίληψη	
perceptron	αλγοριθμική υλοποίηση βιολογικού νευρώνα	
performance	εκτέλεση	
phoneme	φώνημα	το φώνημα αποτελεί έναν ήχο αντίθεσης ή διάκρισης μεταξύ άλλων ήχων σε μια γλώσσα, ή φωνήματα είναι οι ήχοι που χρησιμοποιούνται σε μια γλώσσα για τη δημιουργία διαφορετικών λέξεων.
phonetic context	φωνητικό πλαίσιο	
phonetics	φωνητική	
phonological phrases	φωνολογικές φράσεις	
phonology	φωνολογία	
phonotactic	φωνοτακτικός	
plausible	παραδεκτό	

post-posed	που δομικά έπονται	
post-positional	μεταθετικά μόρια	
post-verbal	μεταρθρωτικός	
prehead modifiers	προσδιορισμός πριν από την κεφαλή της φράσης	<i>Head</i> : κάθε γλωσσική φράση περιέχει ένα ειδικό χαρακτηριστικό που καλείται "κεφαλή" και η οποία αποτελεί το κεντρικό, χαρακτηριστικό και υποχρεωτικό συστατικό της φράσης αυτής (π.χ. μια σύνθεση λέξεων μπορεί να θεωρηθεί ως ρηματική φράση μόνο αν περιέχει και ένα ρήμα που είναι η κεφαλή της φράσης αυτής). <i>Modifiers</i> : το συστατικό εκείνο που περιορίζει το πού αναφέρεται η κεφαλή.
prepausal	προπαυσικός	που βρίσκεται πριν από την παύση
pronoun	αντωνυμία	
proper noun	κύριο όνομα	
prosodic	προσωδιακός	
prosodic constituents	προσωδιακά χαρακτηριστικά	Τα στοιχεία εκείνα που χαρακτηρίζουν την προσωδία (δες και prosodic).
prosody	προσωδία	μαζί, τα εξής χαρακτηριστικά του ήχου: μουσικός τόνος, διάρκεια, ένταση.
proximal syntax	προσεγγιστική σύνταξη	
psychoacoustician	ψυχοακουστικός επιστήμονας	
psycholinguistic	ψυχογλωσσολογικός	
pulse	παλμός	
quantifier	ποσοτικοποιητής	
regularity	κανονικότητα	
resets	ειδικός όρος (ρήσεντ)	ως reset ορίζεται μια αύξηση της Fo κατά μέσον όρο υπολογιζόμενη κατά μήκος δυο διαδοχικών τονικών φράσεων.
respiration	αναπνοή	
schwa		Το φώνημα [θ]. Χρησιμοποιείται για να δηλώσει γενικά ένα φωνήεν χωρίς τόνο όταν γίνεται τονούμενο.
salient acoustic event	προεξέχοντα ακουστικά γεγονότα	
segmental content	περιεχόμενο γλωσσικών τμημάτων	
segmentation	κατάτμηση	διαδικασία που καθορίζει, που αρχίζει και που τελειώνει ένα φώνημα.
semantic	σημασιολογικός	
shwa	δες schwa	
speaking disability	ανικανότητα ομιλίας	περιορισμός ή έλλειψη της ικανότητας ομιλίας σε σχέση με τη συνήθη ικανότητα ή στα όρια που αυτή θεωρείται κανονική.

speech motor	μηχανισμός ομιλίας	
speech recognition	αναγνώριση ομιλίας	
standard deviation	τυπική απόκλιση	στατιστικός όρος.
stress	δυναμικός τόνος	
stress-timed languages	γλώσσες δυναμικού τονισμού	γλώσσες στις οποίες αλλάζει το νόημα, όταν αλλάζει ο τονισμός.
stochastic context-free grammar	στοχαστική ανεπηρέαστη από συμφραζόμενα γραμματική	
stuttering	τραυλισμός	
supersegmental	υπερτμηματικός	εκτείνεται σε έναν αριθμό τμημάτων – δες και segmentation.
suprasegmental	δες και prosodic	υποδηλώνει τα φωνητικά χαρακτηριστικά που αποτελούν το υπόβαθρο, και όχι τα ξεχωριστά τμήματα μιας λέξης ή μιας πρότασης όπως π.χ. δυναμικός τόνος (stress), επιτονισμός (intonation).
syllable-timed languages	γλώσσες με χρονισμό στη συλλαβή	γλώσσες στις οποίες το νόημα αλλάζει, όταν αλλάζει η συλλαβή.
tag	γραμματικός χαρακτηρισμός / σημάδι	
tense vowel	τεταμένο φωνήεν	
temporal	χρονικός	
text-to-speech synthesis	σύνθεση ομιλίας από κείμενο	
theoretical tradition	θεωρητική προσέγγιση	
tone	τόνος	
tradeoffs	συμβιβαστικές ανταλλαγές	
transitions	μεταπτώσεις	CV transitions: μεταπτώσεις από σύμφωνο σε φωνήεν.
utterance	εκστόμιση, λεγόμενο	
verb	ρήμα	
verb phrase	ρηματική φράση	που έχει ως κύριο συστατικό το ρήμα και γύρω από αυτό δομούνται τα υπόλοιπα τμήματα του λόγου π.χ. αντωνυμίες, επίθετα κ.τ.λ.
verbal noun	ρηματικό ουσιαστικό	
vocal inflection	διαμόρφωση της άρθρωσης	
voice onset time	χρόνος εκκίνησης φωνής	
voiced section	φωνούμενο τμήμα	οι φωνητικές χορδές πάλλονται κατά την άρθρωση του, π.χ. voiced consonants είναι τα σύμφωνα: /p/, /j/, /g/
vowel	φωνήεν	
vowel reduction	πτώση των φωνηέντων	Η πτώση των φωνηέντων εξαιτίας άτονης προφοράς τους είτε στην αρχή, είτε στο τέλος μιας λέξης (π.χ. Απ' όλα, παρόλα αυτά). Σε κάποιες άλλες γλώσσες η φωνηεντική πτώση μπορεί να επέλθει ακόμα και στη μέση μιας λέξης

		(όπως στην Αγγλική γλώσσα, ιδιαίτερα σε σχέση με το φώνημα schwa [ə], όταν υπάρχει ακολουθία συμφώνου +/ð/ +/r/ +αδύναμου φωνήεντος.
word	1) λέξη / 2) σημείο	

Γενετική Μετασχηματιστική Θεωρία: από το 1950, επεξεργαζόμενοι την προεργασία του γλωσσολόγου Noam Chomsky, έγιναν πολλές προσπάθειες να παραχθεί ένα είδος θεωρίας της σύνταξης, που θα είχε ένα ξεκάθαρο σύστημα κανόνων που θα μπορούσε να προσδιορίσει ποιοι συνδυασμοί των γλωσσικών στοιχείων θα απέδιδαν καλοσχηματισμένες προτάσεις.

Η θεωρία αυτή, ονομάζεται *γενετική*, από το γεγονός ότι μπορεί να παράγει έναν άπειρο αριθμό κανόνων (όπως στα μαθηματικά $- 3x+2y$, όπου τα x και y μπορούν να αντικατασταθούν από διαφορετικές τιμές και να παράγουν διαφορετικά αποτελέσματα) που δίδουν τη δυνατότητα παραγωγής απείρων προτάσεων.

Η έννοια του μετασχηματισμού από την άλλη, είναι απαραίτητη στη θεωρία αυτή. Μία και οι γενετικοί κανόνες μπορούν να αποδώσουν μια συγκεκριμένη μόνο σειρά των λέξεων μιας φράσης, απαραίτητοι είναι οι *μετασχηματιστικοί κανόνες* (*transformational rules*), οι οποίοι, βοηθούν στη μετακίνηση ορισμένων συστατικών μέσα στη δομή μιας φράσης.

ΠΙΝΑΚΑΣ ΣΥΝΤΥΜΗΣΕΩΝ

CV ΣΦ	Consonant – Vowel
IPCG ΟΔΑΚ	Inter-Perceptual Center Group
IPCI ΔΔΑΚ	Inter-Perceptual Center Interval
PC ΚΑ	Perceptual Center
POS ΜΤΛ	Part Of Speech
TtS	Text to Speech
VOT	Voice Onset Time
SCFG ΣΑΑΣΓ	Stochastic Context-free Grammar