Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών
Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Φωνητικές Διεπαφές Χρήστη-Τεχνολογίες Φωνής

Ενότητα 5: **Διαχείριση και Έλεγχος Φωνητικού Διαλόγου**

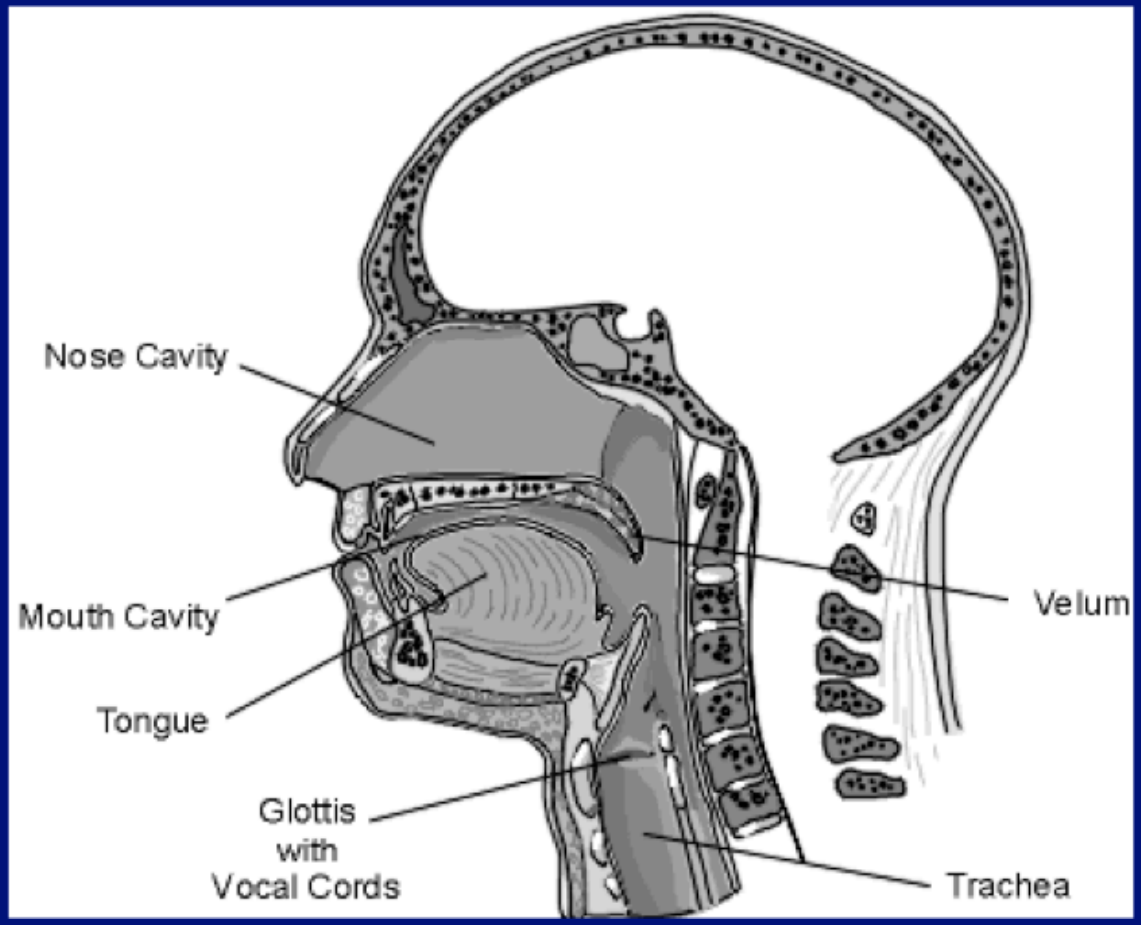
Γεώργιος Κουρουπέτρογλου
koupe@di.uoa.gr

Εργαστήριο Φωνής & Προσβ@σιμότητας
Πανεπιστήμιο Αθηνών
Τμήμα Πληροφορικής

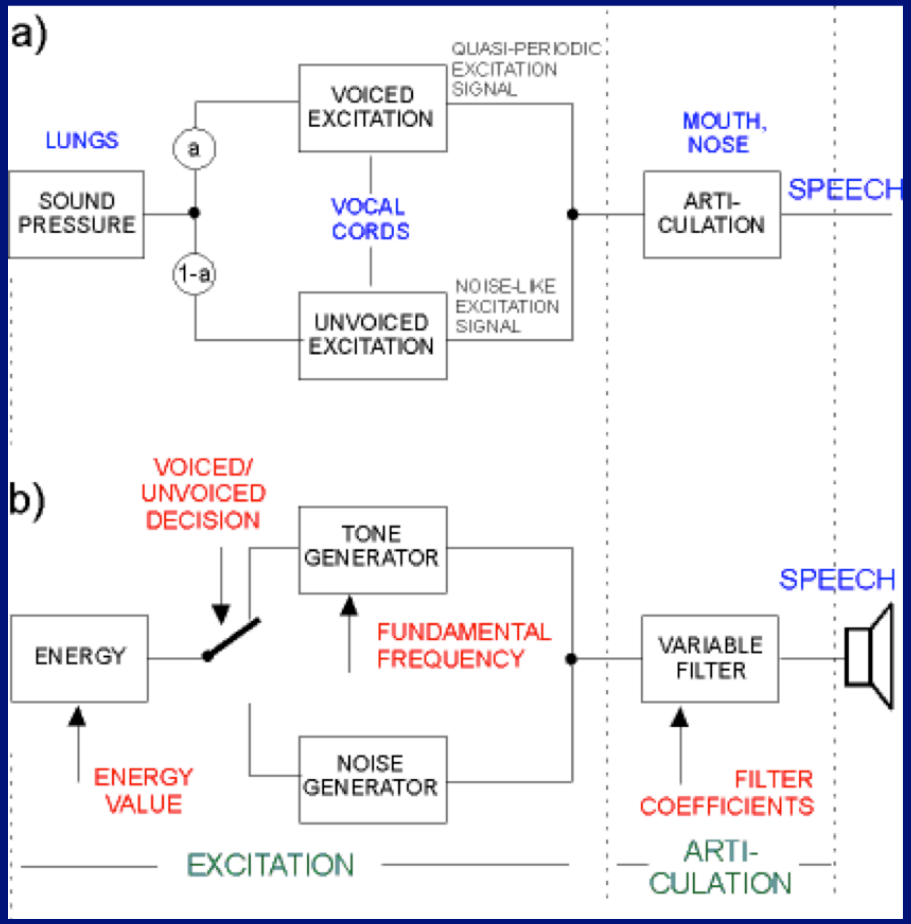
Περιεχόμενα ενότητας

Μοντέλα και εργαλεία διαχείρισης και ελέγχου φωνητικού διαλόγου

Μοντέλο παραγωγής ομιλίας (1 από 2)



Μοντέλο παραγωγής ομιλίας (2 από 2)



Σύνθεση φωνοσυντονισμών

- Ελέγχεται από έναν πίνακα 40 παραμέτρων που ανανεώνουν την συμπεριφορά πηγών και φίλτρων κάθε 5 msec.
- «Είμαι ο πρώτος συνθέτης ομιλίας του Πανεπιστημίου Αθηνών» (1998)
  - Μέτρια καταληπτότητα
  - Χαμηλή φυσικότητα

Σύνθεση συρραφής διφώνων

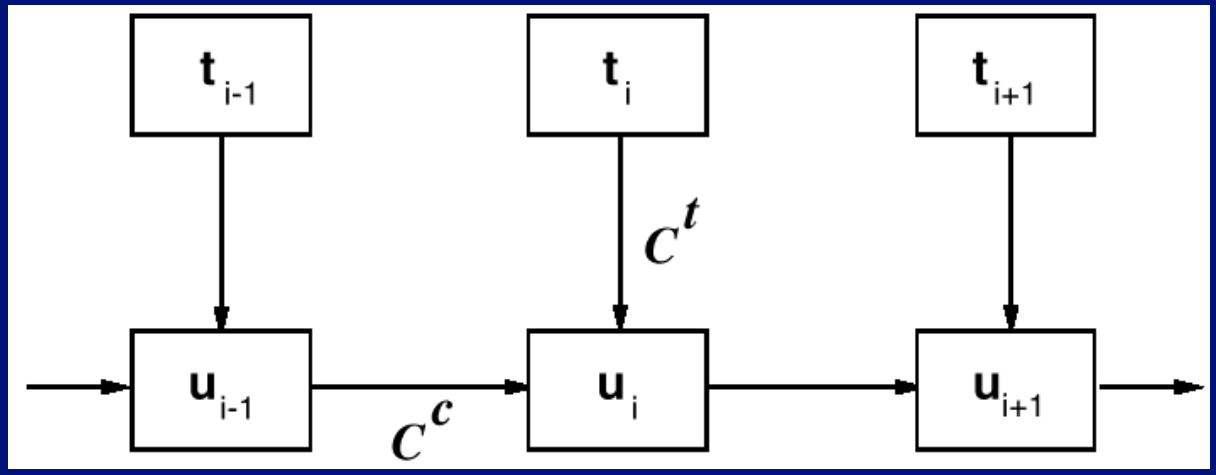
- Οι παράμετροι της ομιλίας εμπεριέχονται σε μικρά ηχογραφημένα τμήματα (φωνήματα, δίφωνα, λέξεις, φράσεις).
- «Είμαι η πρώτη φωνητική βάση διφώνων του Πανεπιστημίου Αθηνών» (2001)
  - Υψηλή καταληπτότητα
  - Μέτρια φυσικότητα

Προσωδία

- «Η θέλησή του Δημοσθένη ήταν τόσο μεγάλη, ώστε, όπως μας αναφέρει ο Πλούταρχος, έβαζε στο στόμα του μικρά χαλίκια την ώρα που απήγγειλε λόγους, προκειμένου να βελτιώσει την άρθρωσή του. Οι προσπάθειες του αυτές, απέδωσαν καρπούς και εξελίχθηκε σε σπουδαίο ρήτορα και πολιτικό.»
- Από εμπειρικά μοντέλα... σε εκπαιδευόμενα...

Σύνθεση από σώμα ηχογραφήσεων...

- Οι παράμετροι της ομιλίας εμπεριέχονται σε μικρά ηχογραφημένα τμήματα (φωνήματα, δίφωνα, λέξεις, φράσεις).
  - Υψηλή καταληπτότητα
  - Υψηλή φυσικότητα



...γενικών και περιορισμένων θεματικών πεδίων

- «Αυτό το έκθεμα είναι ένας αμφορέας και σήμερα βρίσκεται στο μουσείο της Αθήνας»
  - Rhetorical – rVoice 1 (γενικό)
  - Rhetorical – rVoice 2 (γενικό)
  - Loquendo – Actor 1 (γενικό)
  - ΔΗΜΟΣΘΈΝΗΣ – (περιορισμένο)

Τι είναι «κείμενο»;

- «Ο Νίκος παίζει πιάνο.»
- «Στις 21/12 ο Νίκος θα παίξει 3 από τις 4 μπαλάντες του στο ΚΨΜ.»
- «Ο Νίκος (ο αδελφός του κ. Πέτρου) δουλεύει στο ΚΨΜ.»
- «Ο Νίκος είναι:
  - Ψηλός
  - Έξυπνος
  - Αφηρημένος»
- Ο Νίκος έρχεται με τα πόδια, ενώ η Μαρία με αυτοκίνητο.
- <table><tr><td>Μοντέλο</td><td>Κυβικά</td></tr><tr><td>Ibiza</td><td>1600</td></tr><tr><td>Amazon</td><td>2000</td></tr>...</table>

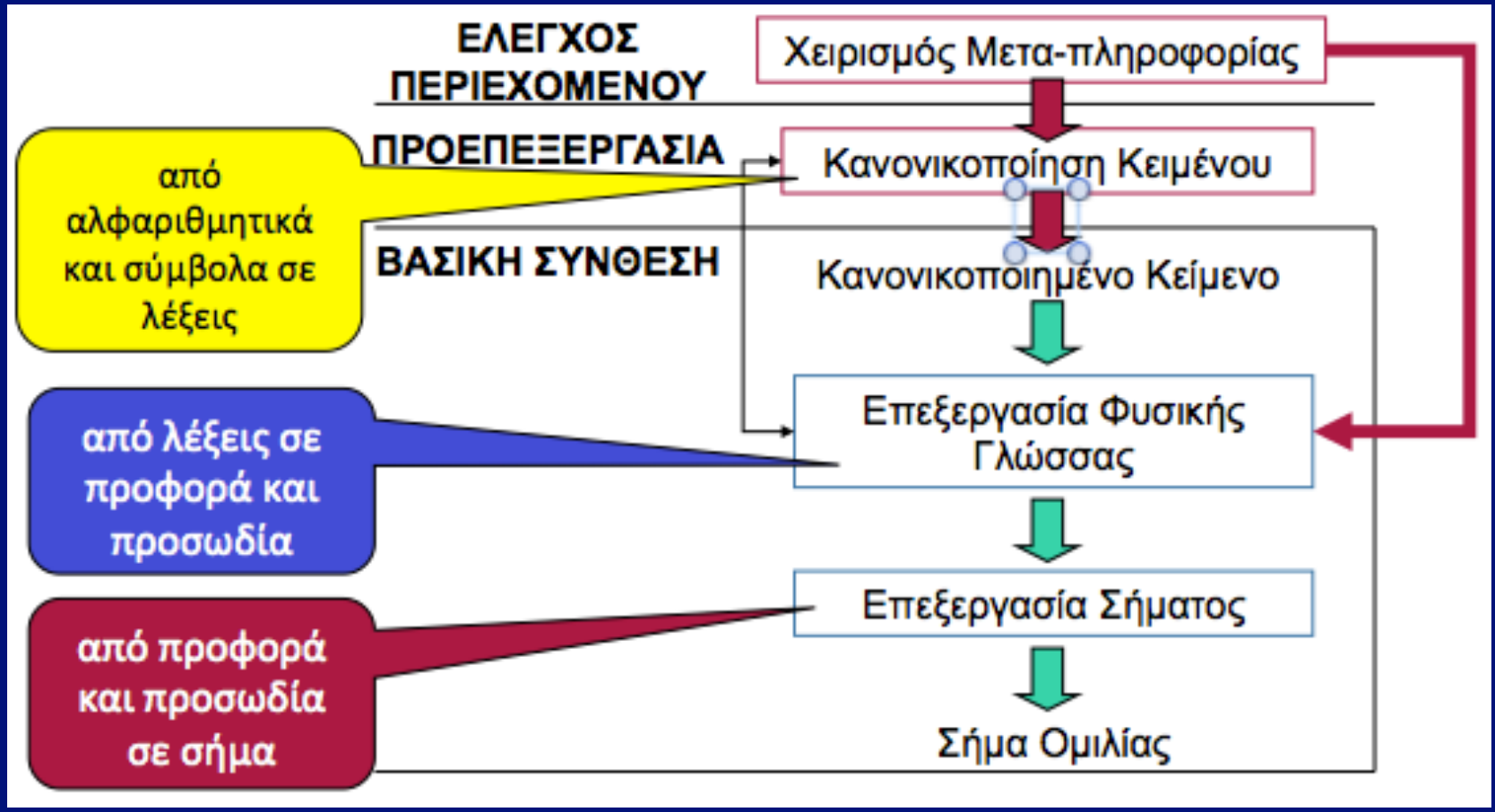
Τι είναι «συνθετική ομιλία»;

- Ομιλία που παράγεται με αυτόματες διαδικασίες από έναν ηλεκτρονικό υπολογιστή και **προσομοιάζει** την συμπεριφορά της ανθρώπινης ομιλίας.
- Σαν τεχνολογία υφίσταται πειραματικά από τα τέλη ’50 και εμπορικά από το ’70.
- Διεπαφή με τον χρήστη τελευταίας γενιάς.
- Από hardware (κουτάκια, χαμηλή ρομποτική ποιότητα) σε software (ευέλικτα, υψηλή ποιότητα)

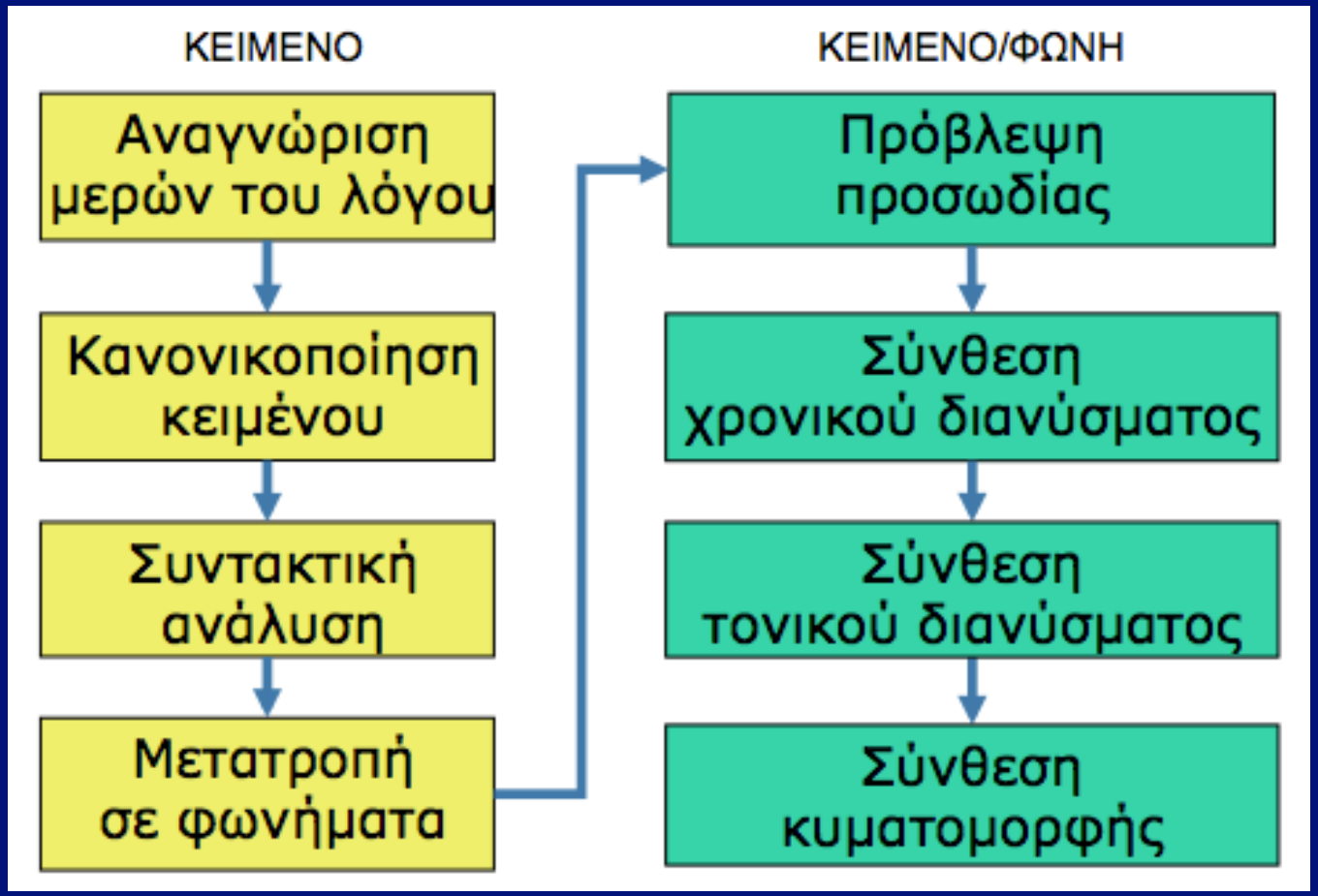
Τι είναι ένα «Σύστημα Μετατροπής Κειμένου σε Ομιλία»;

- Ένα σύστημα το οποίο δέχεται μία ακολουθία **συμβόλων** και με τεχνικές νοημοσύνης συνθέτει μία ή περισσότερες **εκδοχές** από αντίστοιχα **ακουστικά σήματα ομιλίας**.
- Ποιό είναι το πεδίο ορισμού των συμβόλων;
- Πόσο εξελιγμένες είναι οι τεχνικές νοημοσύνης;
- Πως επιλέγεται η κατάλληλη εκδοχή;
- Πόσο φυσικά είναι τα ακουστικά σήματα που παράγονται;

Διαδικασία σύνθεσης



Κλασσική Μετατροπή ΚΣΟ



Κανονικοποίηση κειμένου (1 από 3)

- «Μένω Πατησίων και 3ης Σεπτεμβρίου»
- «Πάρε με στο 210 7275320»
- «Ήταν 13 Ιανουαρίου»
- «Ήταν 13 του μηνός»
- «Ήταν 1 γυναίκα, 1 άντρας και 1 παιδί»
- «Εξετάστηκαν οι αιτήσεις 1654 φοιτητών»
- «Ελάτε μεταξύ 3 και 4 η ώρα το απόγευμα»
- «1500 λουλούδια ήταν οι υποψήφιες»
- «1500 ήταν οι υποψήφιες»

Κανονικοποίηση κειμένου (2 από 3)

	EXPN	abbreviation	*adv, N.Y, mph, gov't*
alpha	LSEQ	letter sequence	*CIA, D.C, CDs*
	ASWD	read as word	*CAT*, proper names
	MSPL	misspelling	*geogaphy*
	NUM	number (cardinal)	*12, 45, 1/2, 0.6*
	NORD	number (ordinal)	*May 7, 3rd, Bill Gates III*
	NTEL	telephone (or part of)	*212 555-4523*
	NDIG	number as digits	*Room 101*
N	NIDE	identifier	*747, 386, I5, pc110, 3A*
U	NADDR	number as street address	*5000 Pennsylvania, 4523 Forbes*
M	NZIP	zip code or PO Box	*91020*
B	NTIME	a (compound) time	*3.20, 11:45*
E	NDATE	a (compound) date	*2/2/99, 14/03/87* (or US) *03/14/87*
R	NYER	year(s)	*1998, 80s, 1900s, 2003*
S	MONEY	money (US or other)	*$3.45, HK$300, Y20,000, $200K*
	BMONEY	money tr/m/billions	*$3.45 billion*
	PRCT	percentage	*75%, 3.4%*
	SPLT	mixed or "split"	*WS99, x220, 2-car* (see also SLNT and PUNC examples)
	SLNT	not spoken,	word boundary or emphasis character:
M		word boundary	*M.bath, KENT*RLTY, _really_*
I	PUNC	not spoken,	non-standard punctuation: "***" in
S		phrase boundary	*$99,9K***Whites*, "..." in *DECIDE...Year*
C	FNSP	funny spelling	*sllooooowww, sh*t*
	URL	url, pathname or email	*http://apj.co.uk, /usr/local, phj@tpt.com*
	NONE	should be ignored	ascii art, formatting junk

Κανονικοποίηση κειμένου (3 από 3)

- «Μένω Πατησίων και 3ης Σεπτεμβρίου»
- «Ήταν 13 Ιανουαρίου»
- «Ήταν 13 του μηνός»
- «Πάρε με στο 210 7275320»
- «Ήταν 1 γυναίκα, 1 άντρας και 1 παιδί»
- «Εξετάστηκαν οι αιτήσεις 1654 φοιτητών»
- «Ελάτε μεταξύ 3 και 4 η ώρα το απόγευμα»
- «1500 λουλούδια ήταν οι υποψήφιες»
- «1500 ήταν οι υποψήφιες»

Συντακτικές συμφωνίες

- Οι κλιτές (inflected) γλώσσες έχουν επιπλέον το πρόβλημα της κλίσης των Non Standard Words (NSW).
- Για την αντιμετώπιση του προβλήματος χρησιμοποιούμε (ΔΗΜΟΣΘέΝΗΣ) γλωσσολογική επεξεργασία που στηρίζεται στη Γραμματική της Νέας Ελληνικής (Μπαμπινιώτη – Χρήστου).
- Φροντίζουν για τη γραμματική συνέπεια ανάμεσα σε αριθμητικά και ουσιαστικά.
- 36,34% των περιπτώσεων βασίζονται σε αυτές τις συμφωνίες για την ορθή κανονικοποίηση τους.
- Συμφωνία ονοματικών συνόλων: **1636 φοιτητών**
- Αντικείμενο στην αιτιατική: **Το μουσείο δέχεται καθημερινά 1500 επισκέπτες**
- Κατηγορούμενο στην ονομαστική: **Οι επιτυχόντες είναι 1501**
- ...

Μετατροπή σε φωνήματα

- Ταύτιση της προφορικής αναπαράστασης με τη γραπτή.
- Εξαρτάται από ιδιώματα κάθε περιοχής.

«Τα παιδιά είδαν για μιά στιγμή τον Κώστα και την Χαρά σε μία ταβέρνα.»

«ta pEDÉÜ ßDan Ãa MÉa stiãmß ton gýsta KE tií xarÜ sE mßa tavÝrna »

- «χαρά» «xarÜ»
- «χέρι» «÷Ýri»

Προσωδία (1 από 4)

- Τονικές (κινέζικα, αφρικάνικα) & επιτονικές γλώσσες (ευρωπαϊκές)
- Τονικές: διαφορετικά επίπεδα τόνου αλλάζουν το νόημα μίας λέξης.
- Επιτονικές γλώσσες: free-stress (π.χ. Αγγλικά) και fixed-stress (π.χ. Γαλλικά).
- Ο μελωδικός τονισμός και η καμπύλη επιτονισμού εξαρτώνται από την οργάνωση των λέξεων σε μονάδες υψηλότερου επιπέδου (προσωδιακές λέξεις, ονοματικά σύνολα, ενδιάμεσες φράσεις κλπ)

Προσωδία (2 από 4)

- Συντακτική ανάλυση σχετικά εύκολη σε σχέση με σημασιολογική και πραγματολογική
- Είναι ένα σημαντικό θέμα, σε ένα πρώτο επίπεδο, η απόδοση της προσωδίας σε συνάρτηση των συντακτικών σχέσεων ανάμεσα στις λέξεις των φράσεων.
  - Μελωδικές φράσεις συνήθως συναντώνται μέσα σε μορφο-συντακτικές ομάδες
  - Είναι σπάνια η εμφάνιση παύσεων ανάμεσα σε στενά συνδεδεμένες (συντακτικά) λέξεις.
- Είναι όμως δυνατόν να διαβαστεί ένα κείμενο χωρίς καταρχήν να γίνεται κατανοητό;

Προσωδία (3 από 4)

- Εστίαση (focus): «Ο Νίκος ήρθε στη Νάξο με πλοίο»
  - Σωστό μήκος συλλαβών και επιτονισμός
  - Λάθος εστίαση οδηγεί σε παρεξήγηση
  - Το TtS πρέπει να καταλαβαίνει όχι απομονωμένες προτάσεις αλλά όλο το κείμενο...
- Κοινός τόπος εκφωνητή (Ε) και ακροατή (Α):
  - Ένα προσωδιακό πρότυπο από τον Ε -> Α εξαρτάται από το τι γνωρίζει ο Α για την κατάσταση που του περιγράφεται
  - Σταμάτα να ρωτάς αν ο «Νίκος ήρθε στη Νάξο με πλοίο». Είναι η τρίτη φορά που σου λέω ότι ΔΕΝ ήρθε.
  - Η ιδέα ότι ο Νίκος θα ερχότανε στη Νάξο με πλοίο είναι αστεία: σιχαίνεται τα πλοία.

Προσωδία (4 από 4)

Τα βήματα που ακολουθεί ένα TtS προκειμένου να γεννήσει τα προσωδιακά διανύσματα μίας φράσης:

- Αναγνώριση και κατηγοριοποίηση φράσεων.
- Πρόβλεψη προσωδιακής δομής φράσεων
  - Θέση, διάρκεια και τύπος παύσεων (break indices)
  - Θέση και τύπος μελωδικών τόνων (pitch accents)
  - Θέση και τύπος τόνων στα όρια (endtones – prosodic phrase tones + boundary tones)
- Υπολογισμός διανύσματος διάρκειας
- Απόδοση καμπύλης F0.
- Υπολογισμός διανύσματος έντασης

Αναγνώριση μερών του λόγου

- Απαραίτητη διαδικασία για:
  - Κανονικοποίηση κειμένου σε κλιτές γλώσσες (π.χ. Ελληνικά)
  - Μετατροπή σε φωνήματα (π.χ. «χρόνια», «record»)
  - Πρόβλεψη προσωδιακής δομής μίας πρότασης:
    - Προσδιορισμός προσωδιακών φράσεων
    - Προσδιορισμός παύσεων στις προσωδιακές φράσεις
    - Προσδιορισμός μελωδικού τονισμού (pitch accent) σε φράσεις
    - Προσδιορισμός τόνων ορίων (boundary tones) σε φράσεις
- Ποσοστό επιτυχίας (ελληνικά): ~95%

Θεματικά πεδία (1 από 2)

- Πεδία εφαρμογής TtS στα οποία η θεματολογία είναι συγκεκριμένη.
  - Περιορισμένο λεξιλόγιο.
  - Περιορισμένα γλωσσολογικά φαινόμενα.
  - Περιορισμένα προσωδιακά φαινόμενα (από επαγγελματίες εκφωνητές του πεδίου).
  - Άρα, μικρότερο πεδίο προσομοίωσης και λιγότερα λάθη κατά την επεξεργασία
- Ευκολύνεται η δημιουργία στατιστικών μοντέλων από πραγματικές μετρήσεις (π.χ. Μορφοσυντακτική ανάλυση, προσωδία -> ~95%)
- Ακουστικό σήμα: γίνεται προσιτή η χρήση μεγαλύτερων ακουστικών μονάδων συρραφής -> μεγαλύτερη φυσικότητα

Θεματικά πεδία (2 από 2)

- Παραδείγματα:
  - Υπηρεσίες καταλόγου (131, σινεμά, βενζινάδικα κλπ)
  - Ανακοινώσεις (δελτίο καιρού, οικονομικό δελτίο κλπ)
  - Περιήγηση σε μουσεία
  - Τεχνικά ή νομικά κείμενα
  - Ειδήσεις
- Με βάση τους περιορισμούς που πρέπει να θέτει ένα θεματικό πεδίο, δεν θεωρούνται θεματικά πεδία:
  - Λογοτεχνικά κείμενα
  - Ποίηση

«Φωνή»

- Ο καθένας θέλει μία φωνή με προσωποποιημένα χαρακτηριστικά.
- Τα ομιλούντα προϊόντα (θα) θέλουν να έχουν την δική τους ξεχωριστή φωνή.
- Τι είναι φωνή;
  - Καταρχήν, ένα συγκεκριμένο ηχόχρωμα
  - Μία ακολουθία από διαδικασίες επεξεργασίας φυσικής γλώσσας: διαφορετικές τοπολαλιές, ιδιώματα, τρόποι προφοράς συνεπτηγμένων μορφών, προσωδιακή συμπεριφορά.

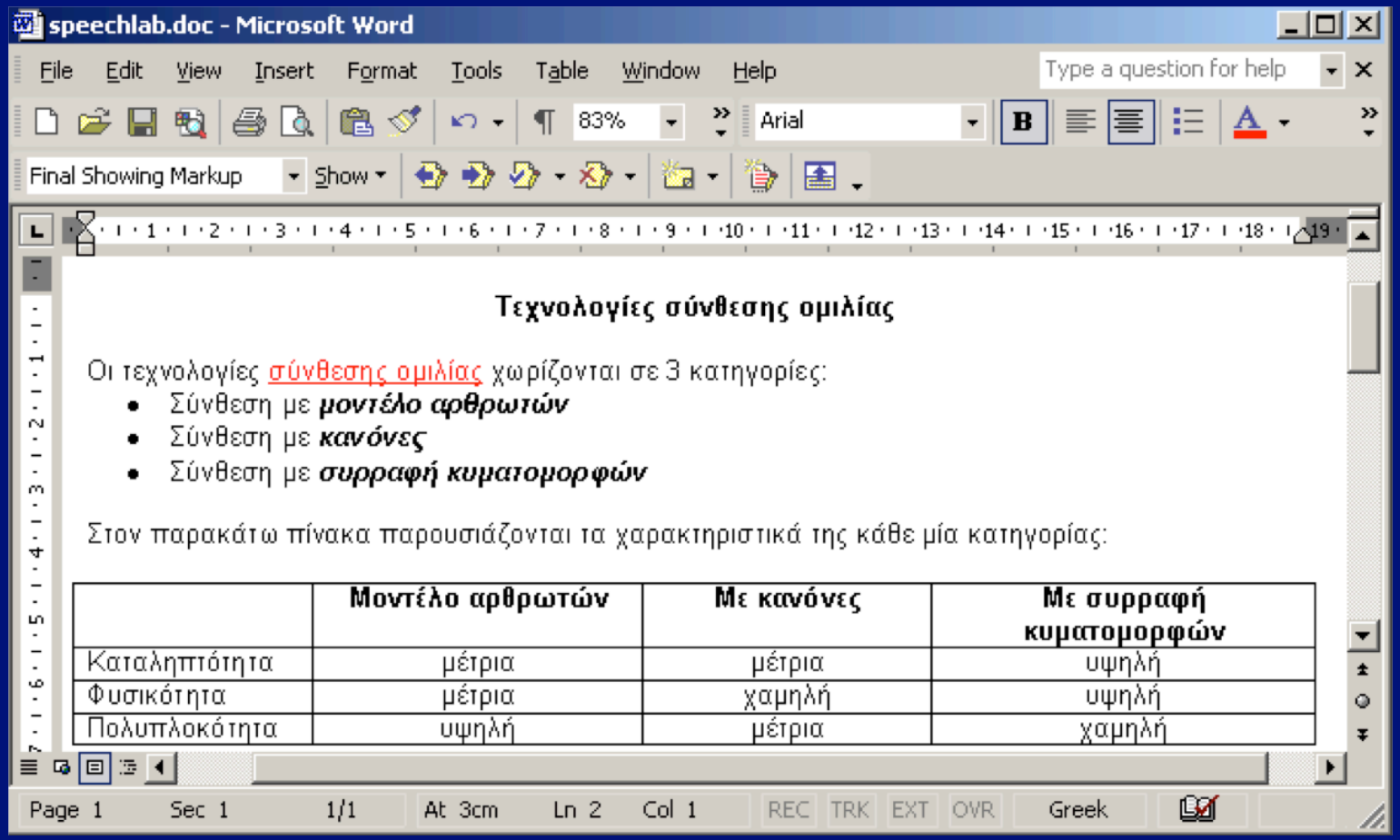
Ποιότητα συνθετικής ομιλίας

- **Καταληπτότητα:** ήταν κατανοητό το περιεχόμενο της ομιλίας;
- **Φυσικότητα:** πόσο κοντά στην φυσική χροιά ήταν η ομιλία;
- Επιπλέον, μεταδόθηκε σωστά η **πληροφορία**;

Μετα-πληροφορία (1 από 2)

- **Οδηγίες οπτικοποίησης**: bold, italics, tables, bullets, big-small letters κλπ (π.χ. HTML, MS-Word)
- **Οδηγίες δομής**: header, title, section, record content κλπ (π.χ. HTML, XML, SQL)
- **Δομή κειμένου**: παρενθέσεις, σημάδια, κλπ
- **Γλωσσολογική πληροφορία**: ρητορικές σχέσεις, σύνταξη, γραμματική, μορφολογία (π.χ. SOLE, plain κείμενο)
- **Οδηγίες ομιλίας**: prosody, emp, rate, pitch κλπ (π.χ. SABLE, VoiceXML, SSML, ACSS)

Μετα-πληροφορία (2 από 2)



	Μοντέλο αρθρωτών	Με κανόνες	Με συρραφή κυματομορφών
Καταληπτότητα	μέτρια	μέτρια	υψηλή
Φυσικότητα	μέτρια	χαμηλή	υψηλή
Πολυπλοκότητα	υψηλή	μέτρια	χαμηλή

Άσκηση 5.1

Α) Εγκαταστήστε στον υπολογιστή σας το σύστημα Μετατροπής Κειμένου σε Ομιλία ΔΗΜΟΣΘέΝΗΣ http://demosthenes.di.uoa.gr/

Β) Γράψτε 10 ελληνικές προτάσεις που να περιέχει κάθε μία τουλάχιστον δύο αριθμητικά από τις κατηγορίες: τηλέφωνο, ημερομηνία, ποσό (να παραδοθούν και οι προτάσεις). Δώστε τις προτάσεις αυτές να τις εκφωνήσει ο ΔΗΜΟΣΘέΝΗΣ και σημειώστε (και αναφέρετε) τα τυχόν λάθη στο τμήμα Κανονικοποίησής του.

Γ) Πάρετε ένα τυχαίο ελληνικό κείμενο από το διαδίκτυο τουλάχιστον 100 λέξεων (να παραδοθεί το κείμενο αυτό) και δώστε το να το εκφωνήσει ο ΔΗΜΟΣΘέΝΗΣ. Ελέγξατε αν το τμήμα Μετατροπής σε Φωνήματα του ΔΗΜΟΔΘέΝΗ λειτουργεί σωστά σε όλες τις περιπτώσεις φωνημάτων και αλλοφώνων. Σημειώστε (και αναφέρετε) τα προβλήματα που βρήκατε.

Τέλος Ενότητας

Διαχείριση και Έλεγχος Φωνητικού Διαλόγου

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Σημείωμα Αναφοράς

Σημείωμα Αδειοδότησης

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων (1 από 2)

Οι φωτογραφίες που περιέχονται στην παρουσίαση αποτελούν πνευματική ιδιοκτησία τρίτων. Απαγορεύεται η αναπαραγωγή, αναδημοσίευση και διάθεσή τους στο κοινό με οποιονδήποτε τρόπο χωρίς τη λήψη άδειας από τους δικαιούχους.

Σημείωμα Χρήσης Έργων Τρίτων (2 από 2)