



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ  
ΣΧΟΛΗ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΚΑΙ  
ΠΟΛΙΤΙΣΜΙΚΩΝ ΣΠΟΥΔΩΝ  
Τμήμα Φιλολογίας

# ΕΠΙΠΕΔΑ ΑΝΑΛΥΣΗΣ ΤΗΣ ΓΛΩΣΣΑΣ

10<sup>η</sup> ΔΙΑΛΕΞΗ  
Π. ΓΑΚΗΣ  
gakis@sch.gr

# Ποσοτική Γλωσσολογία

# Ποσοτική γλωσσολογία

- Ποσοτική Γλωσσολογία (ΠΓ) είναι ο κλάδος εκείνος της Γλωσσολογίας που ασχολείται με την **ποσοτική ανάλυση της γλωσσικής δομής** και τη γλωσσολογική ερμηνεία της.
- Η ποσοτική ανάλυση χρησιμοποιείται για να ολοκληρωθεί η ποιοτική ανάλυση που διεξάγει ο κλάδος της θεωρητικής γλωσσολογίας. Γενικότερα, η χρήση ποσοτικών μεθόδων όπως αυτές που θα περιγραφούν παρακάτω **λειτουργούν συμπληρωματικά** με τις ποιοτικές θεωρήσεις ως προς την κατανόηση του γλωσσικού φαινομένου.

# Ηλεκτρονικά Σώματα Κειμένων

- Η ποσοτική αντιμετώπιση της γλωσσικής χρήσης **θα ήταν αδύνατη** εάν δεν υπήρχαν τα ΗΣΚ
- Ορισμοί
  - είναι η **συλλογή** τμημάτων γλώσσας τα οποία επιλέγονται και **διατάσσονται** σύμφωνα με συγκεκριμένα γλωσσολογικά **κριτήρια** έτσι ώστε να χρησιμοποιηθούν ως αντιπροσωπευτικό δείγμα μιας συγκεκριμένης γλώσσας (EAGLES 1996).
  - είναι μια **συλλογή** κειμένων η οποία είναι **κωδικοποιημένη** για **τυποποιημένες** (standardized) και **ομοιογενείς** εργασίες ανάκτησης γλωσσικής πληροφορίας.

## Χαρακτηριστικά της έρευνας που βασίζεται στη χρήση ΗΣΚ

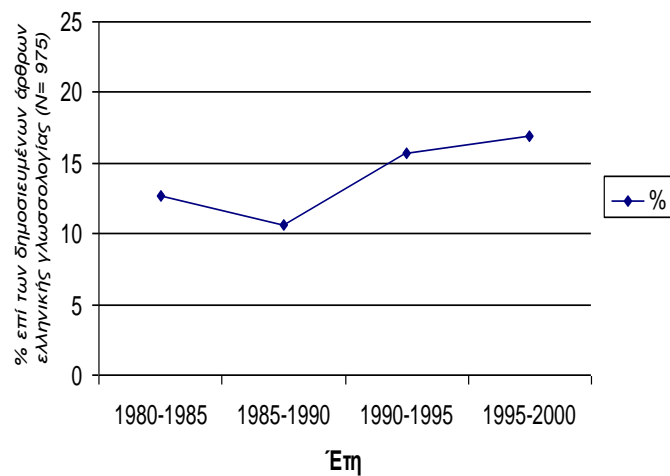
- **Άμεση εξάρτηση** από τα κείμενα που περιέχονται στο ΗΣΚ
  - Η ποσοτική (μέγεθος) και η ποιοτική (κειμενική ποικιλία) σύσταση του ΗΣΚ διαμορφώνει τα αποτελέσματα που παίρνουμε από αυτό
- **Αξιοποίηση των Η/Υ** στην επεξεργασία του γλωσσικού υλικού
  - **Ταχύτητα** και **αξιοπιστία** κατά την εκτέλεση τυποποιημένων εργασιών γλωσσικής ανάλυσης
- **Ποσοτική και ποιοτική προσέγγιση** της γλωσσικής χρήσης
  - Η **γλωσσική χρήση αντιμετωπίζεται ολιστικά** και παρέχονται πληροφορίες τόσο για την ποσοτική **δομή**, όσο και για την λειτουργική **αλληλεπίδραση** των γλωσσικών στοιχείων

# Ο ρόλος των Η/Υ στην ανάλυση των ΗΣΚ

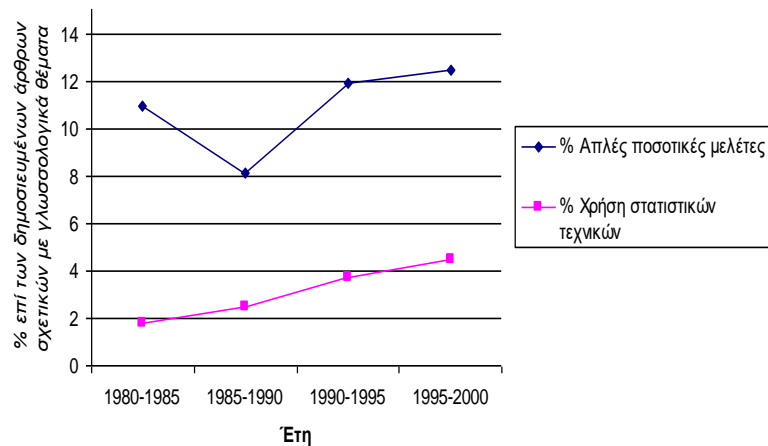
- Δυνατότητα τεράστιας αποθήκευσης κειμενικών δεδομένων
- Μεγάλη ταχύτητα επεξεργασίας γλωσσικών δεδομένων (Wordsmith: 15.000 λέξεις το δ/λπτο)
- Συνεπής και «αλάνθαστη» απόδοση σε επαναληπτικές διαδικασίες

# Ποσοτικές μελέτες στην ελληνική γλώσσα

Χρονολογική αύξηση των ποσοπικών μελέτων στην ελληνική γλώσσα



Συγκριτική αύξηση των ποσοπικών μελετών σχετικών με την ελληνική γλώσσα ανάλογα με το αν χρησιμοποιούν απλές ποσοπικές μεθόδους ή εξειδικευμένες στατιστικές τεχνικές



# Ποσοτική ανάλυση της Νέας Ελληνικής

- Ανάλυση της γλωσσικής χρήσης
  - Λίστες συχνότητας λέξεων
  - Λεξικά συνάψεις
  - Γλωσσική ποικιλία
  - Στατιστικά χαρακτηριστικά της γλώσσας
- Υφομετρική ανάλυση κειμένων



# Ανάλυση Γλωσσικής Χρήσης: Λίστες Συχνότητας Λεξιλογίου

N	Word	Freq.	%
1	και	52.675	2,99
2	το	37.936	2,15
3	του	36.708	2,08
4	να	34.539	1,96
5	της	29.760	1,69
6	η	27.909	1,58
7	που	25.922	1,47
8	την	24.985	1,42
9	με	21.951	1,25
10	από	20.384	1,16
11	για	18.205	1,03
12	τα	18.145	1,03
13	ο	17.423	0,99
14	των	17.366	0,99
15	είναι	15.038	0,85
16	δεν	13.869	0,79
17	τη	12.922	0,73
18	σε	12.819	0,73
19	οι	12.792	0,73
20	τον	12.278	0,70
21	στο	12.223	0,69
22	θα	11.805	0,67
23	τους	11.138	0,63
24	στην	10.573	0,60
25	τις	9.245	0,52
26	ότι	8.629	0,49

N	Word	Freq.
3545	αα	8
3546	ααα	3
3547	άαα	1
3548	άααα	1
3549	άαααα	1
3550	αβ	2
3551	αβαθείς	1
3552	αβαθή	2
3553	άβακα	2
3554	αβακούμ	1
3555	αβάνα	1
3556	αβαντάς	3
3557	αβαπτιστα	2
3558	αβαρίες	1
3559	αβασάνιστα	4
3560	αβασάνιστη	1
3561	αβασάνιστης	1
3562	αβασάνιστο	1
3563	αβάσιμη	1
3564	αβασκαμένους	1
3565	αβάσκαντη	1
3566	αβάστακτη	2
3567	αβάσταχτες	1
3568	αβασταχτη	1
3569	αβάσταχτη	4
3570	αβάσταχτο	5

# Κατάλογοι συχνότητας (στατιστικά στοιχεία)

Σειρά	Λέξη	Συχνότητα	Ποσοστό (%)	Κάλυψη (%)
191	φαίνεται	9849	0,04	16,43
192	χώρα	9846	0,04	16,97
193	περιοχή	9837	0,04	11,81
194	πάντα	9807	0,04	15,04
195	χώρο	9760	0,04	15,37
196	εκ	9734	0,04	15,89
197	οποίος	9709	0,04	18,03
198	επειδή	9705	0,04	14,12
199	μέρος	9647	0,04	15,54
200	προβλήματα	9636	0,04	16,38

- Σε μια χαμηλότερη ζώνη συχνότητας αρχίζουν να εμφανίζονται λεξικά στοιχεία ή πλήρεις τύποι, και μάλιστα **γενικού είδους και χωρίς αναφορά σε συγκεκριμένο πεδίο** της γλώσσας.
- Έτσι, στη σειρά συχνότητας 191-200, μαζί με τα γραμματικά στοιχεία *εκ, οποίος, επειδή*, εμφανίζονται λεξικοί τύποι όπως *φαίνεται, χώρα, περιοχή, προβλήματα κ.ά.*, στοιχεία που θα λέγαμε ότι ανήκουν στο βασικό λεξιλόγιο μιας γλώσσας (Γούτσος 2006)
- Κι εδώ ένας τύπος όπως ο τύπος *πάντα* καλύπτει **πολλά λήμματα**, μεταξύ των οποίων το επίρρημα *πάντα* και το επίθετο *τα πάντα*

# Κατάλογοι συχνότητας (στατιστικά στοιχεία)

12.001	γκαραζ	184		0,36
12.002	διδασκικές	184		0,40
12.003	εβραίων	184		0,29
12.004	εγγραφές	184		0,45
12.005	εγκληματική	184		0,51
12.006	εισήγαγε	184		0,50
12.007	εργοτάξιο	184		0,32
12.008	ιδιωτικός	184		0,50
12.009	κέρδισαν	184		0,58
12.010	κλίσεις	184		0,42

- Καθώς μικραίνει ακόμη περισσότερο η συχνότητα (π.χ. από το 12.001 και μετά), **εμφανίζονται πολλές λέξεις με την ίδια συχνότητα** (184), που χρησιμοποιούνται σε πολύ μικρό ποσοστό κειμένων (από 0,29 έως 0,58%, δηλαδή 1 φορά σε κάθε 200 με 300 κείμενα).
- Στη ζώνη αυτή **δεν εμφανίζονται πλέον γραμματικοί τύποι.**

# Κατάλογοι συχνότητας (στατιστικά στοιχεία)

330.441	Αβινιόν	1		
330.442	Αβιονίζ	1		
330.443	αβιοτικά	1		
330.444	αβιοτικού	1		
330.445	αβιοτικούς	1		
330.446	αβιταμινώσεις	1		
330.447	αβιάτων	1		
330.448	αβλάφεια	1		
330.449	αβλαβός	1		
330.450	άβλαφη	1		

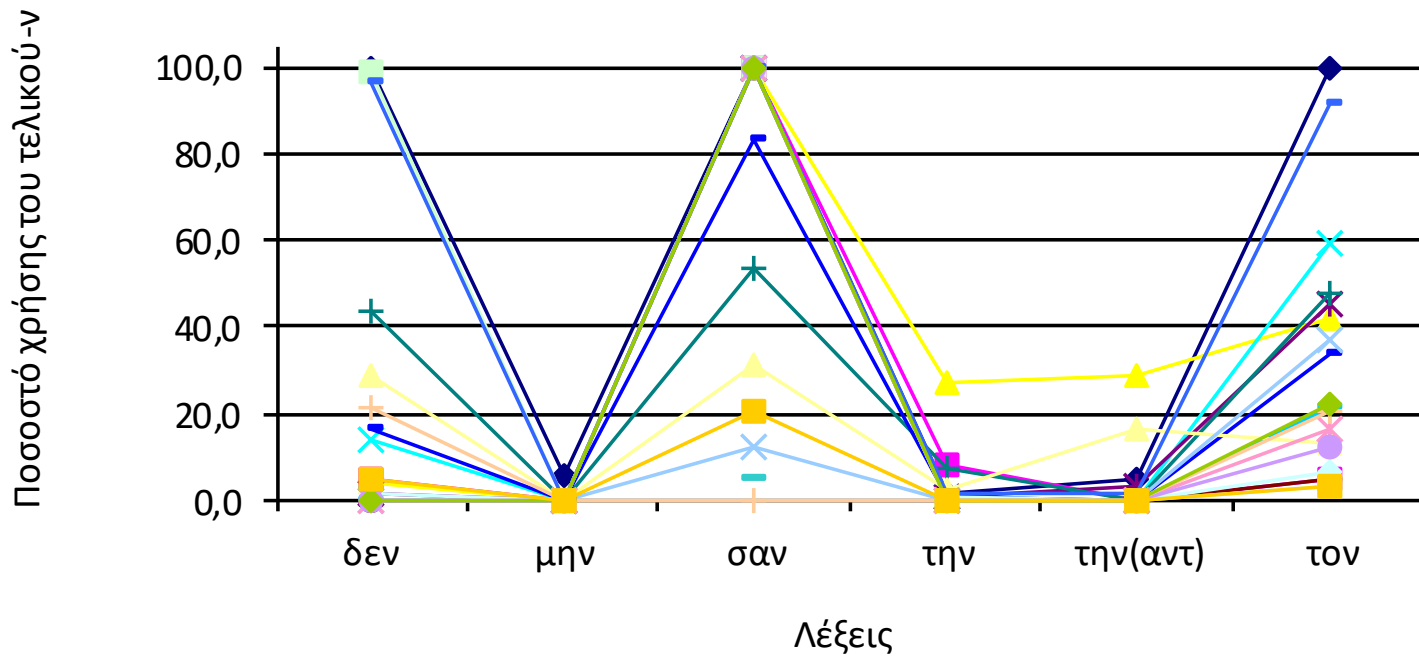
- Στο τέλος του καταλόγου συχνότητας υπάρχει ένας τεράστιος αριθμός λέξεων που απαντούν στο σώμα κειμένων **μόνο μία φορά**.
- Πρόκειται για τα **άπαξ λεγόμενα** (και στα αγγλικά χρησιμοποιείται ο ελληνικός όρος *hapax legomena*), τα οποία, όπως βλέπουμε, αποτελούν **ασυνήθιστες λέξεις**, για παράδειγμα ειδικευμένους όρους, κύρια ονόματα, παλαιότερους τύπους κ.ά.

## Ανάλυση Γλωσσικής Χρήσης: Τα συχνότερα κύρια ονόματα

A/A	Λέξη	Εμφανίσεις	Συχνότητα (τοίς χιλίοις)
1	Ελλάδα	30032	0,8754 %
2	Αθήνα	13020	0,3795 %
3	Τουρκία	11944	0,3481 %
4	Ευρώπη	8976	0,2616 %
5	Παπανδρέου	8855	0,2581 %
6	Έλληνας	7324	0,2135 %
7	Σημίτης	6226	0,1815 %
8	Θεσσαλονίκη	5888	0,1716 %
9	Ρέππας	4946	0,1442 %
10	Γιώργος	4837	0,1410 %

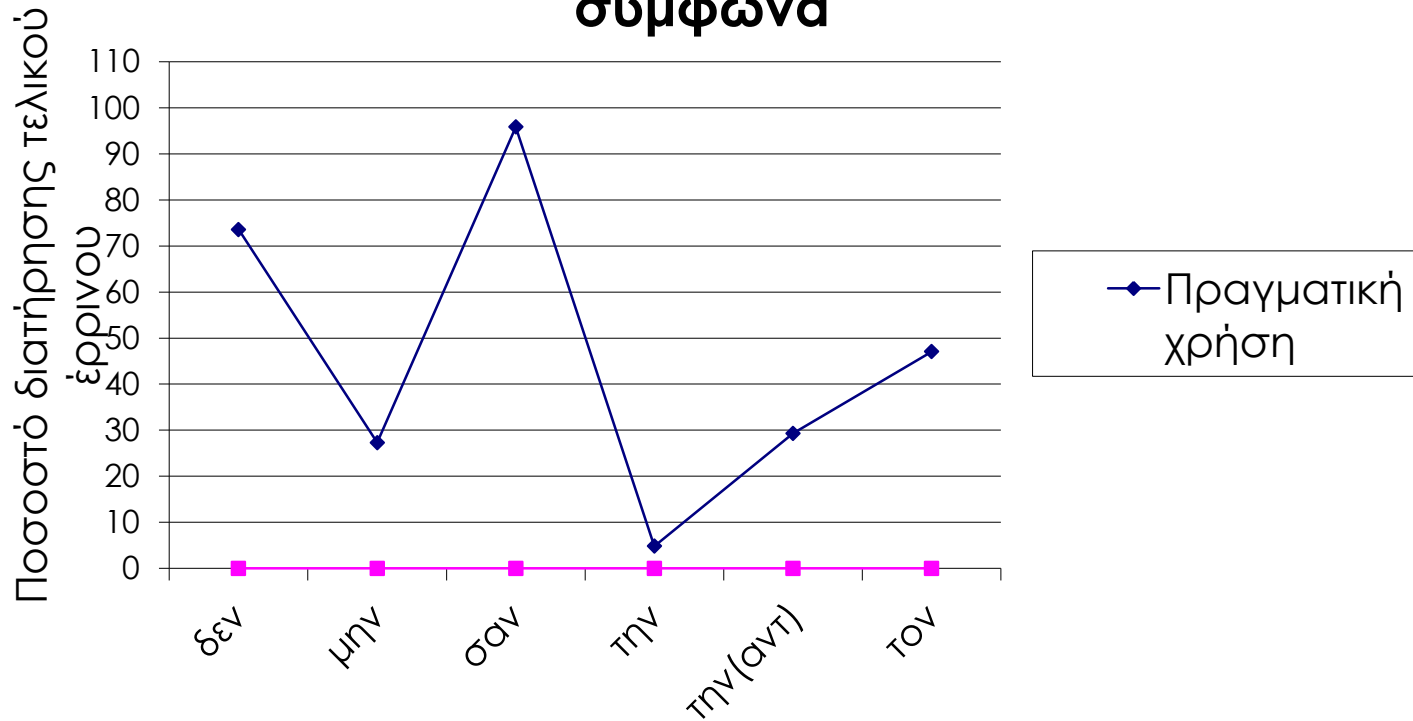
## Ανάλυση Γλωσσικής Χρήσης:

Ποσοστό χρήσης του τελικού -ν πριν από εξακολουθητικά ανάλογα με τον συγγραφέα (Εκδόσεις Καστανιώτη)



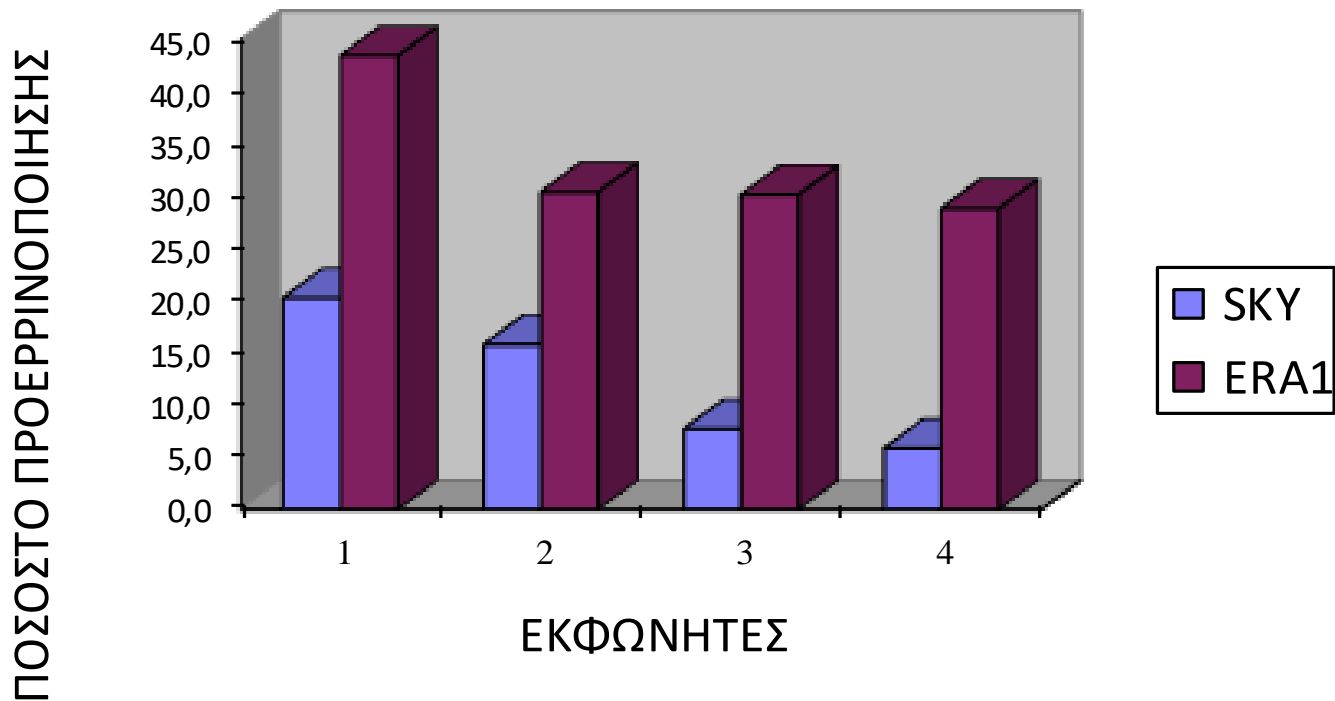
# Ανάλυση Γλωσσικής Χρήσης:

## Σύγκριση Πραγματικής και Ιδανικής Χρήσης του Κανόνα πριν από Εξακολουθητικά σύμφωνα



# Ανάλυση Γλωσσικής Χρήσης:

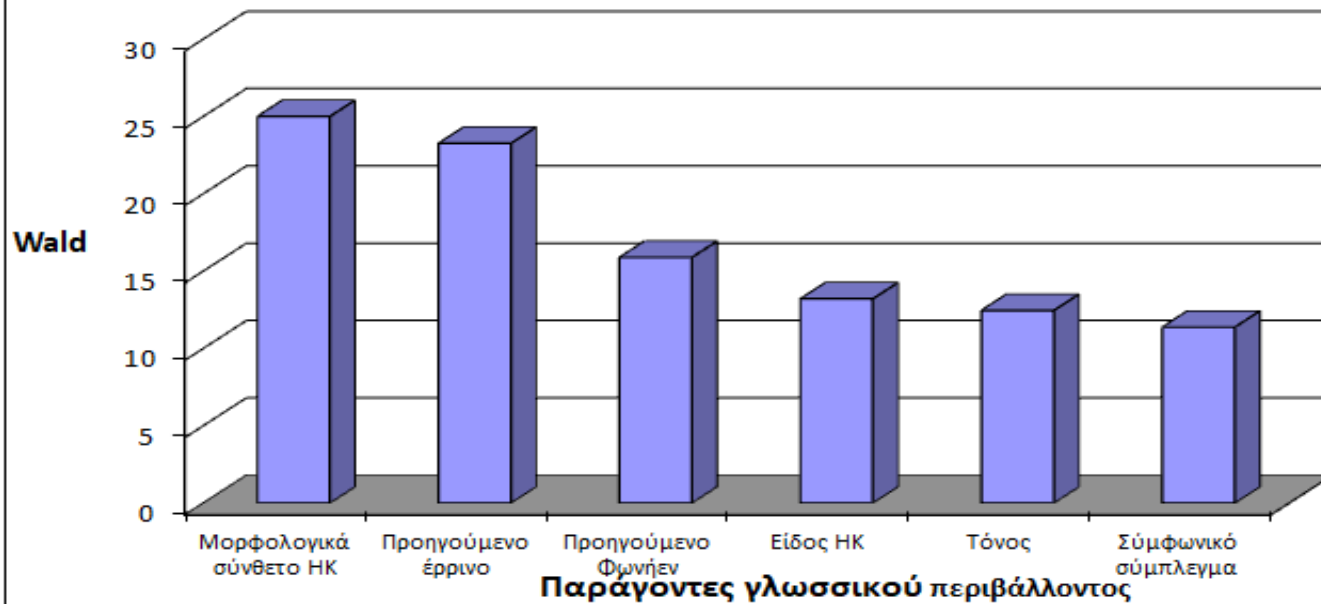
## ΠΟΣΟΣΤΑ ΠΡΟΕΡΡΙΝΟΠΟΙΗΣΗΣ ΑΝΑ ΕΚΦΩΝΗΤΗ ΣΤΗΝ ERA1 ΚΑΙ SKY 100,4



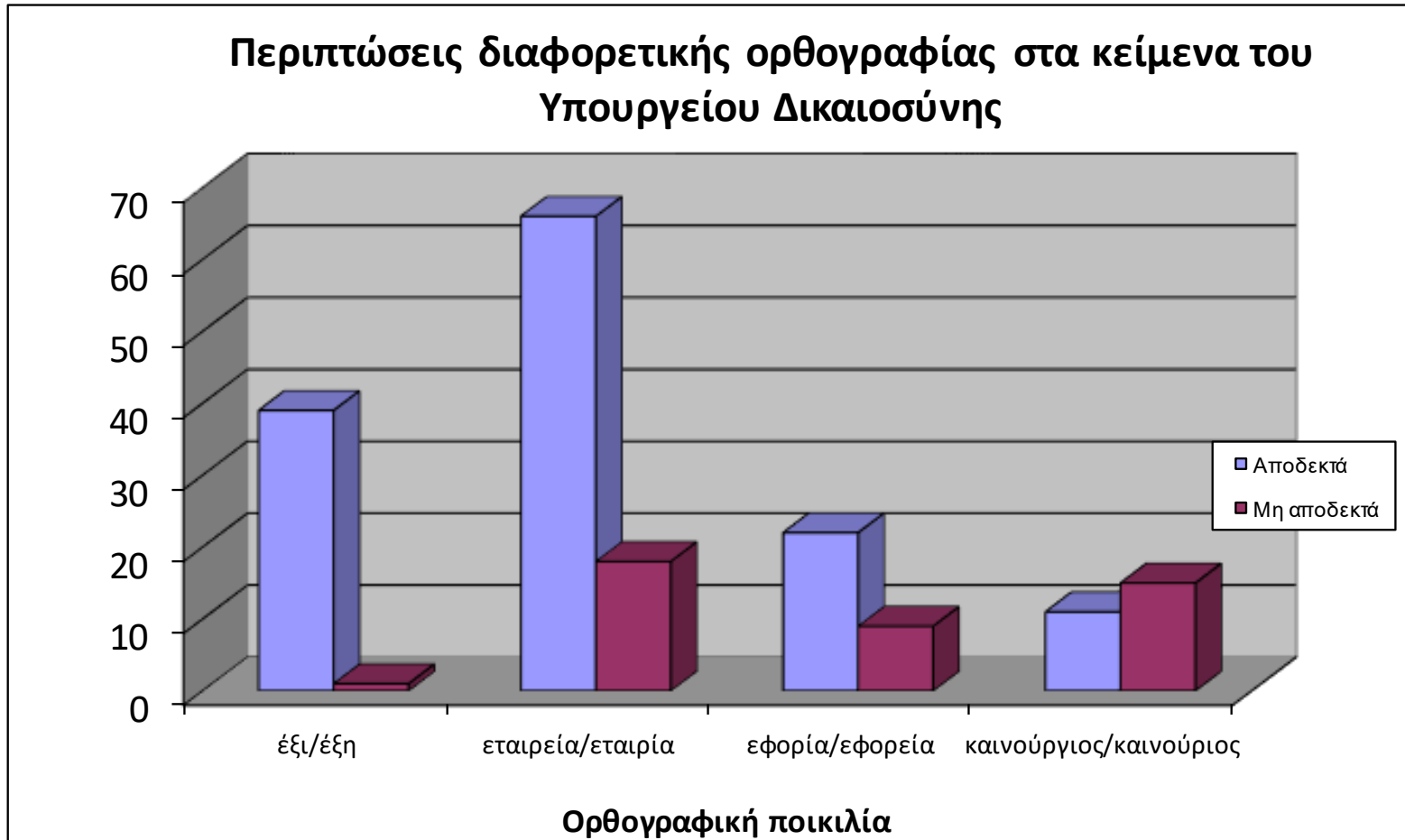


## Οι γλωσσικοί παράγοντες που σχετίζονται με την προερρινοποίηση

Η δύναμη της επίδρασης των παραγόντων στην προερρινοποίηση των ΗΚ

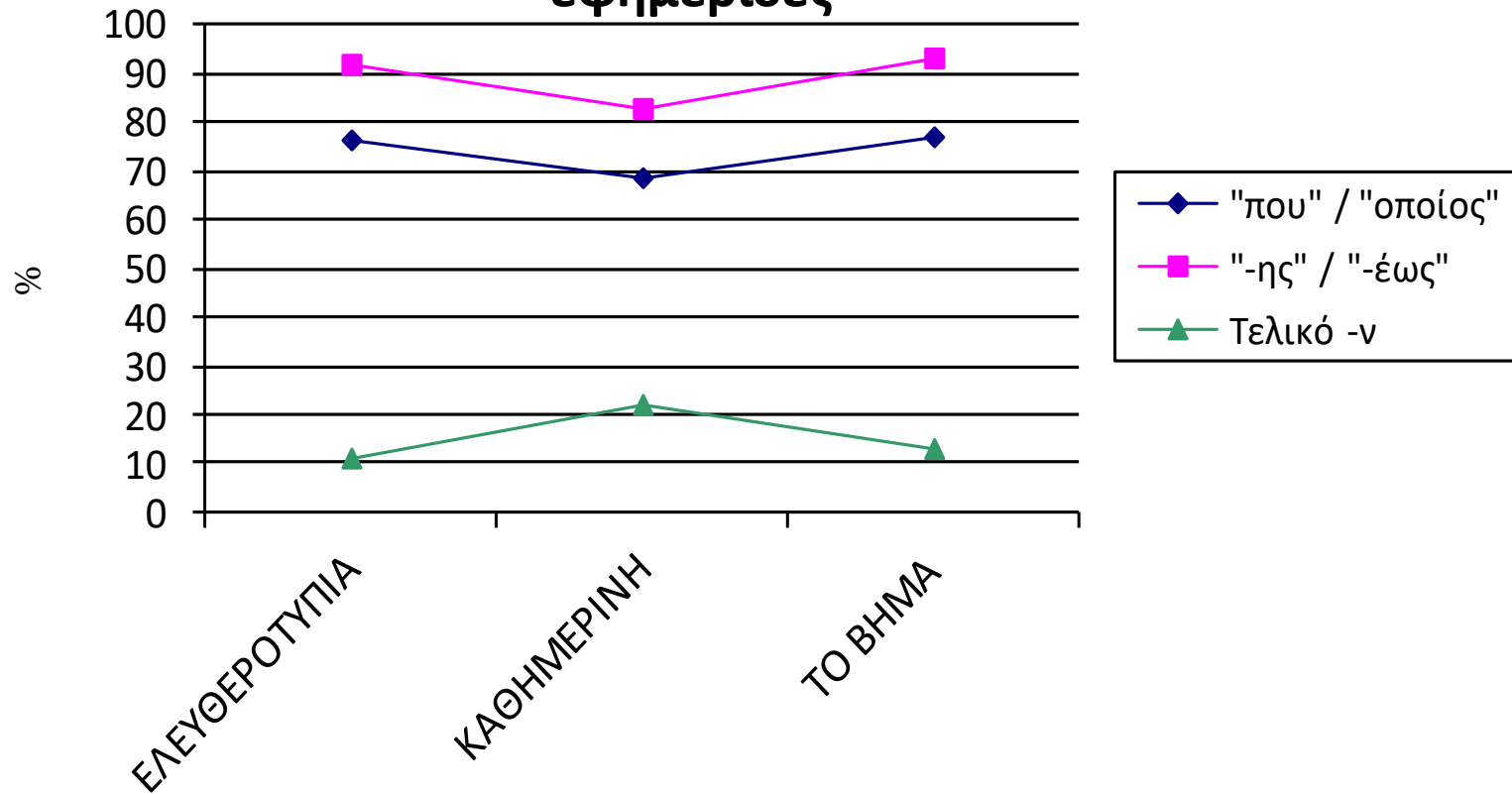


## Ανάλυση Γλωσσικής Χρήσης:



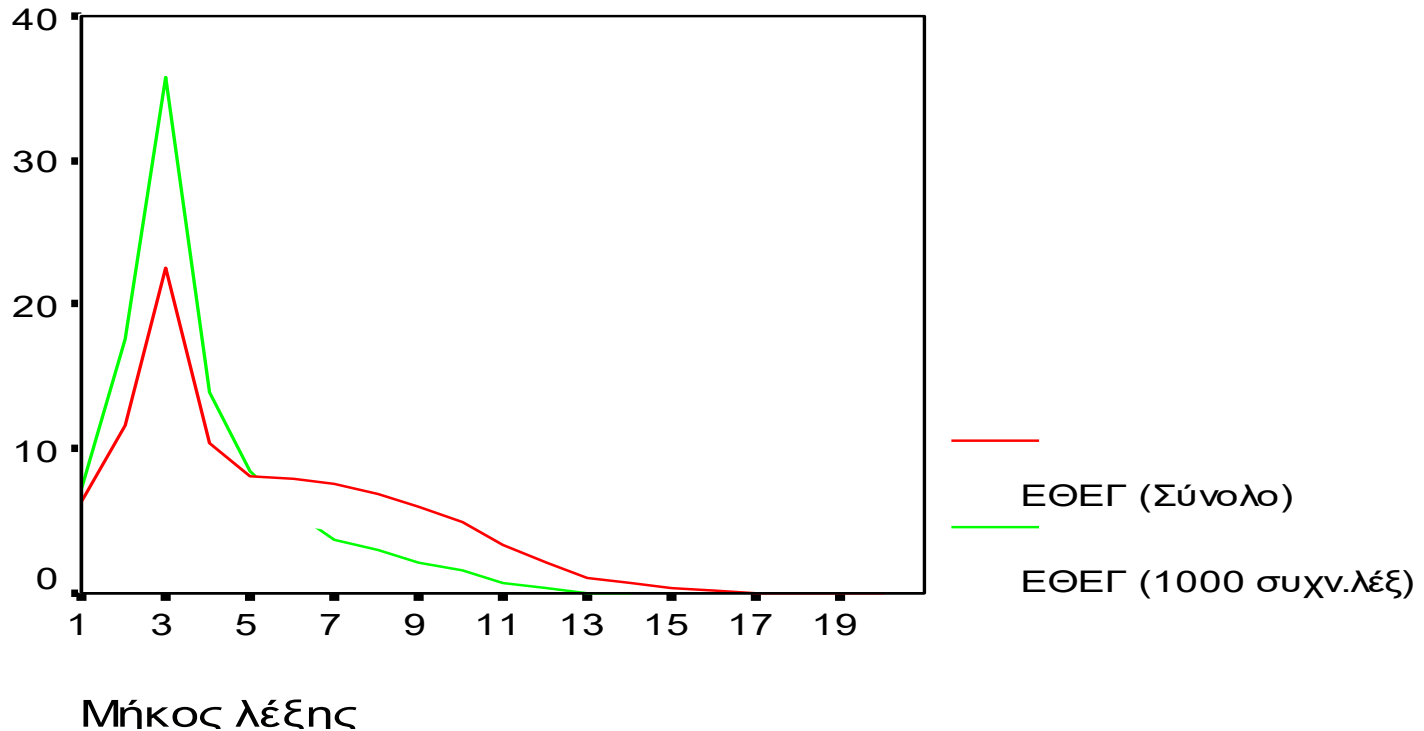
# Ανάλυση Γλωσσικής Χρήσης:

## Γλωσσική ποικιλία και λόγια στοιχεία στις εφημερίδες

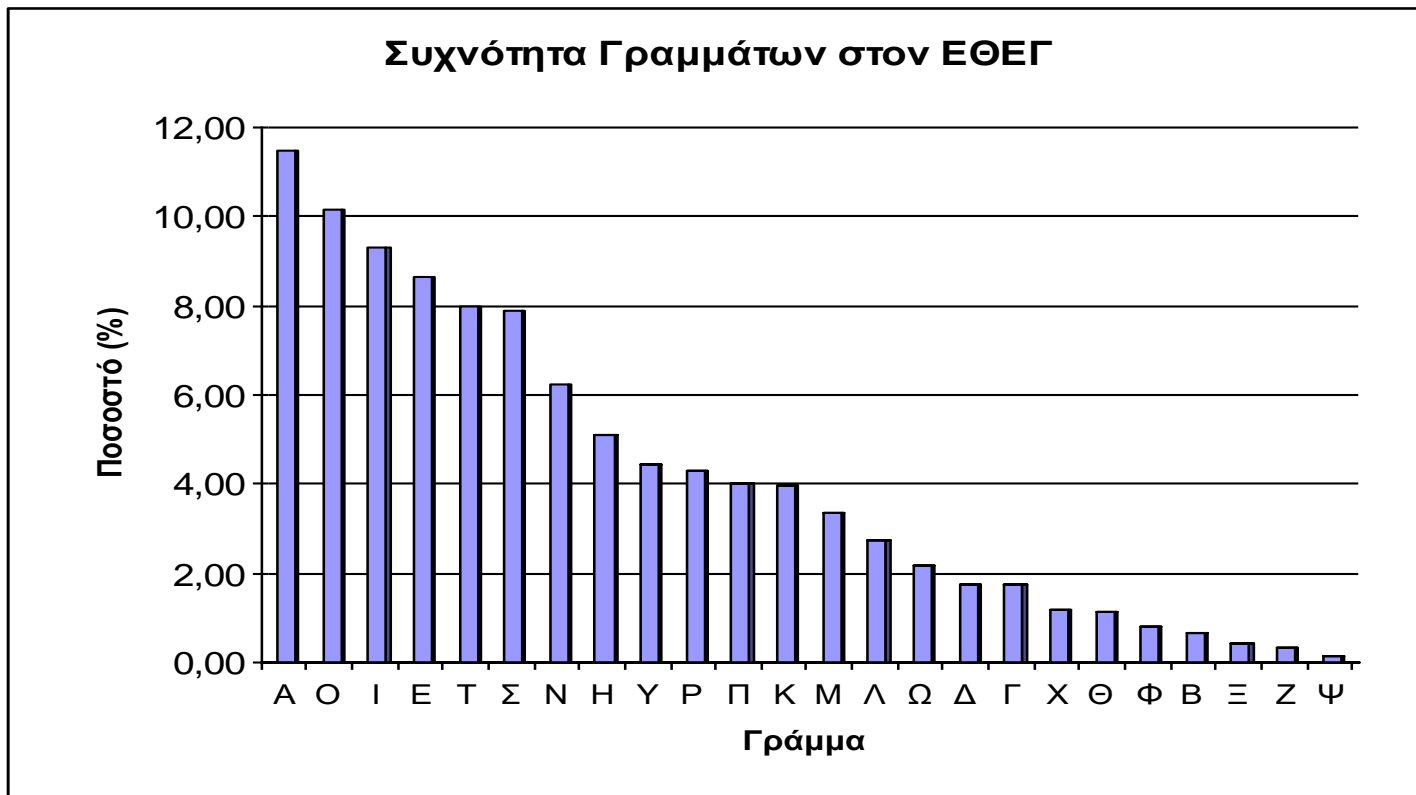


# Στατιστικά χαρακτηριστικά της γλώσσας

Σύγκριση του μήκους λέξης στις 1000  
συχν. λέξεις και στο σύνολο του ΕΘΕΓ

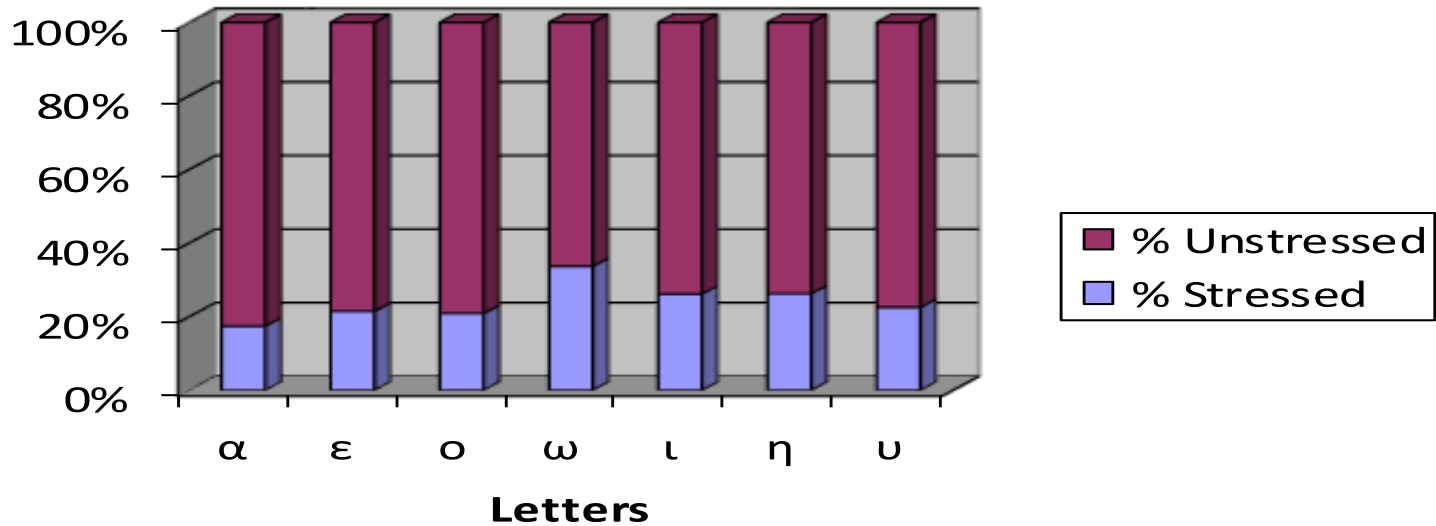


# Στατιστικά χαρακτηριστικά της γλώσσας Ν: Συχνότητες Γραμμάτων



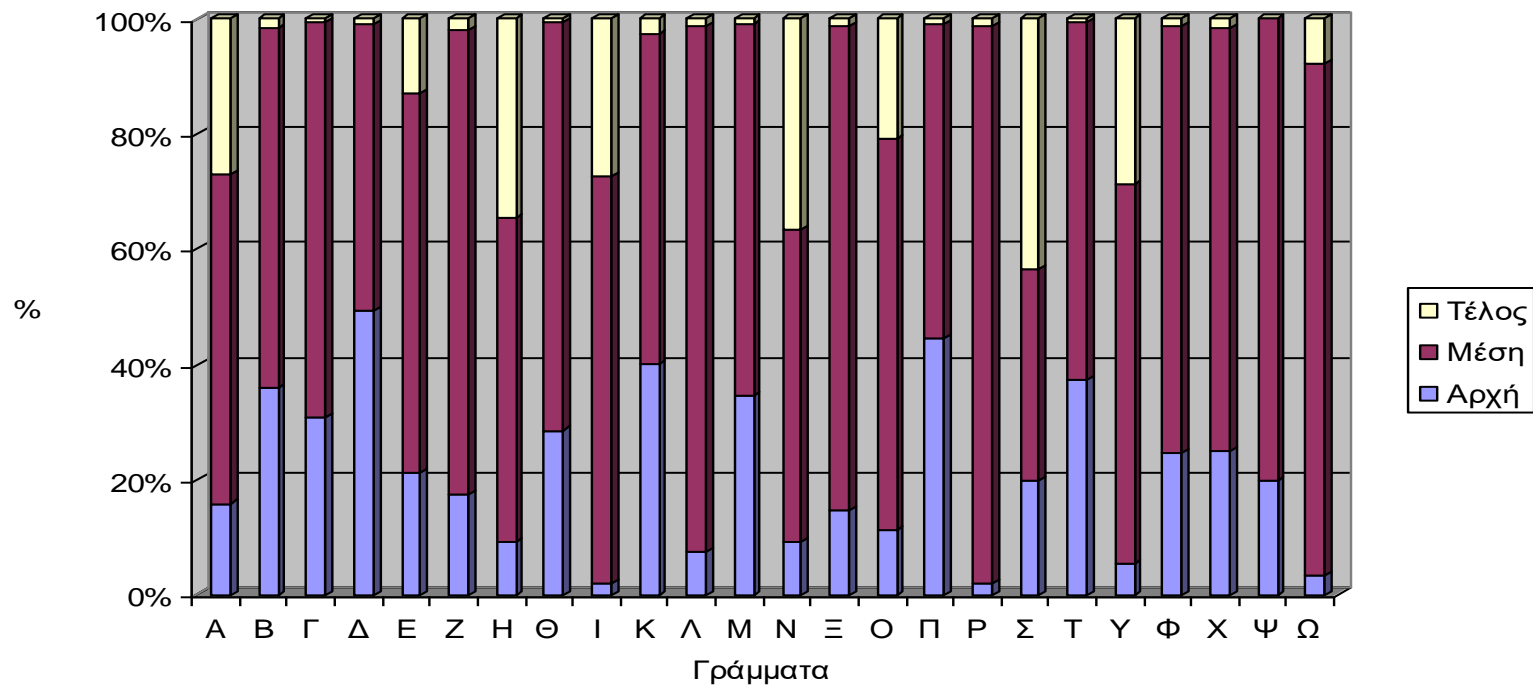
# Στατιστικά χαρακτηριστικά της γλώσσας Ν: Συχνότητες Γραμμάτων

**Distribution of stressed/unstressed letters  
in HNC**



# Στατιστικά χαρακτηριστικά της γλώσσας Ν: Συχνότητες Γραμμάτων

Κατανομή της θέσης των γραμμάτων μέσα στην λέξη



# Στατιστικά στοιχεία

Γενικά στοιχεία	
Λέξεις:	96,476,161
Μοναδικές λέξεις:	1,026,778
Λήμματα:	335,577
Προτάσεις:	5,253,120
Κείμενα:	115,340
Μέσος όρος αριθμού λέξεων στην πρόταση:	18
Μέσος όρος αριθμού χαρακτήρων στη λέξη:	9

Συχνότερες λέξεις			
	Λέξη	Εμφανίσεις	Συχνότητα
1.	και	3,449,019	35.7500 %
2.	να	2,014,868	20.8846 %
3.	της	1,664,968	17.2578 %
4.	την	1,569,974	16.2732 %
5.	η	1,568,265	16.2555 %
6.	με	1,299,564	13.4703 %
7.	για	1,254,956	13.0079 %
8.	από	1,214,189	12.5854 %
9.	το	1,116,165	11.5693 %
10.	ο	1,056,987	10.9559 %



# Στατιστικά στοιχεία

Συχνότερα λήμματα		
	Λέξη	Εμφανίσεις
1.	ο	12,403,735
2.	και	3,597,794
3.	σου	2,654,598
4.	να	2,022,788
5.	μου	2,003,903
6.	που	1,311,820
7.	με	1,308,618
8.	για	1,290,735
9.	από	1,267,682
10.	είμαι	1,175,565

Συνεμφανιζόμενα ζεύγη λέξεων			
	Συνδυασμός	Εμφανίσεις	Συχνότητα
1.	για την	244,654	2.5359 %
2.	από την	208,687	2.1631 %
3.	για να	188,625	1.9551 %
4.	με την	186,570	1.9338 %
5.	από το	163,191	1.6915 %
6.	πρέπει να	133,863	1.3875 %
7.	και η	131,204	1.3600 %
8.	και να	120,660	1.2507 %
9.	με το	118,753	1.2309 %
10.	για το	110,822	1.1487 %

# Στατιστικά στοιχεία

## Συμφωνιζόμενες τριάδες λέξεων

	Συνδυασμός	Εμφανίσεις	Συχνότητα
1.	θα πρέπει να	32,923	0.3413 ‰
2.	δεν μπορεί να	20,103	0.2084 ‰
3.	σε σχέση με	14,730	0.1527 ‰
4.	κατά τη διάρκεια	13,799	0.1430 ‰
5.	με την οποία	13,320	0.1381 ‰
6.	της νέας δημοκρατίας	12,822	0.1329 ‰
7.	θα μπορούσε να	12,162	0.1261 ‰
8.	σύμφωνα με την	11,446	0.1186 ‰
9.	για πρώτη φορά	11,061	0.1147 ‰
10.	και κύριοι συναδέλφοι	10,514	0.1090 ‰

## Συμφωνιζόμενες τετράδες λέξεων

	Συνδυασμός	Εμφανίσεις	Συχνότητα
1.	κατά τη διάρκεια της	6,669	0.0691 ‰
2.	κυρίες και κύριοι συναδέλφοι	6,610	0.0685 ‰
3.	χειροκροτήματα από την πτέρυγα	5,469	0.0567 ‰
4.	αναφορά με την οποία	4,621	0.0479 ‰
5.	σε ό τι αφορά	4,006	0.0415 ‰
6.	κατέθεσε αναφορά με την	3,832	0.0397 ‰
7.	κυρίες και κύριοι συναδέλφοι	3,522	0.0365 ‰
8.	από την πτέρυγα της	3,316	0.0344 ‰
9.	σε σχέση με το	3,280	0.0340 ‰
10.	πτέρυγα της νέας δημοκρατίας	2,931	0.0304 ‰

# Στατιστικά στοιχεία

Μέσο δημοσίευσης	
Άλλο:	1,647
Βιβλίο:	252
Διαδίκτυο:	64,647
Εφημερίδα:	46,649
Περιοδικό:	2,127

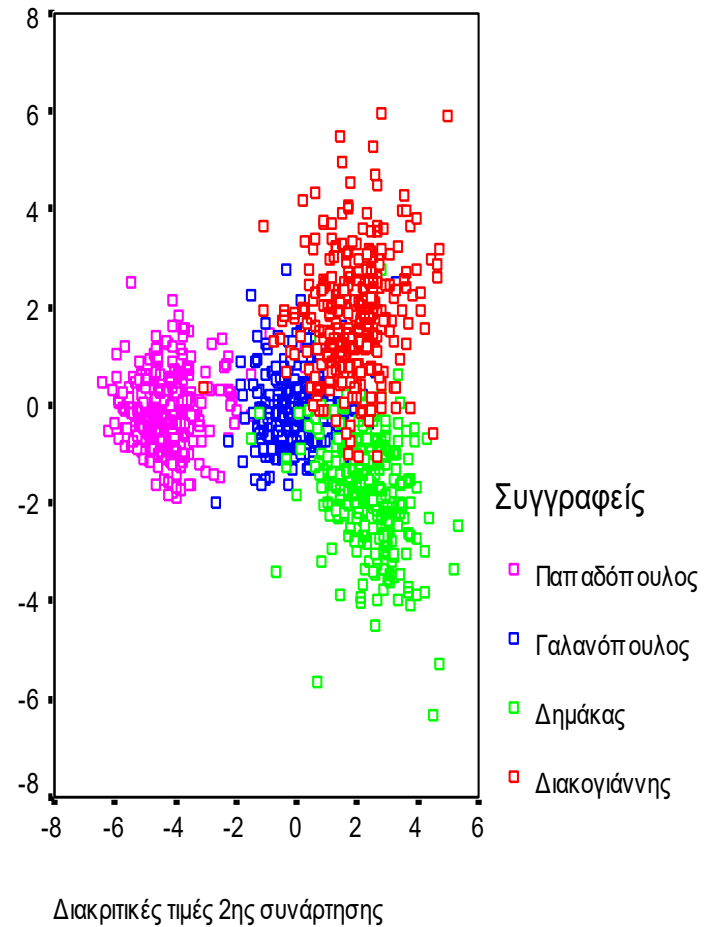
# Μερικά βασικά συμπεράσματα

- Οι 1000 συχνότερες λέξεις μένουν σταθερές ανεξαρτήτως του μεγέθους του ΗΣΚ που τις εξετάζουμε (89,5% ομοιότητα)
- Τα πιο συχνά φωνήεντα είναι τα «Α, Ο, Ι» και τα πιο συχνά σύμφωνα είναι τα «Τ, Σ, Ν».
- 22% των φωνηέντων είναι τονισμένα.

# Αυτόματη Αναγνώριση Συγγραφέα

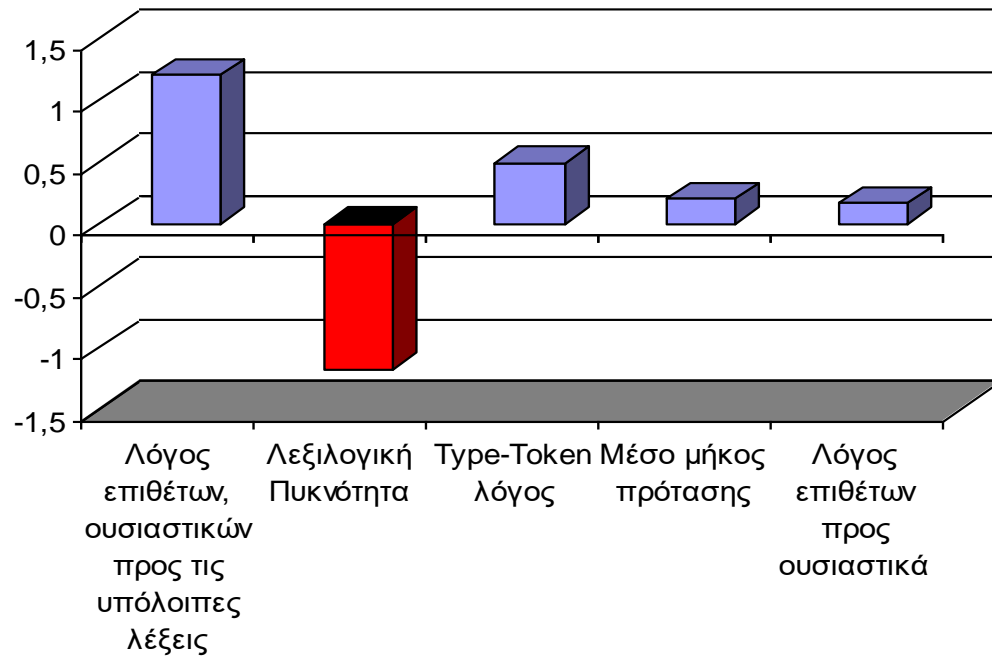
- **Λεξιλογικός πλούτος**
  - Yule's K
  - Λεξιλογική πυκνότητα
  - TTR
- **Μετρήσεις σε επίπεδο λέξης**
  - Μέσο μήκος λέξεων
  - Κατανομή του μήκους λέξεων
- **Μετρήσεις σε επίπεδο πρότασης**
  - Μέσο μήκος της πρότασης (σε λέξεις)
- **Σημεία στίξης**
- **Μεταβλητές «Διγλωσσίας»**
  - Ποικιλία των τριτόκλιτων καταλήξεων σε -ης και -εως
  - Ποικιλία στη χρήση των αναφορικών αντωνυμιών «που» και «οποίος»
- **80 συχνότερες λειτουργικές λέξεις**

**Αποτέλεσμα: Σε σύνολο 1200 κειμένων 93,6% ακρίβεια αναγνώρισης του συγγραφέα**



# Η σχέση υφομετρικών χαρακτηριστικών ενός κειμένου και της δυσκολίας κατανόησής του

Υφομετρικοί παράγοντες που επηρεάζουν την δυσκολία κατανόησης ενός κειμένου



# Συμπεράσματα

- Η ποσοτικές μέθοδοι επιτρέπουν την θέαση της γλωσσικής χρήσης τόσο σε μικροσκοπικό όσο και σε μακροσκοπικό επίπεδο.
- Προσφέρουν ακριβείς ποσοτικές πληροφορίες που επιτρέπουν την μαθηματική περιγραφή των γλωσσικών φαινομένων.
- Επιτρέπουν τον εμπειρικό έλεγχο των γλωσσικών θεωριών.
- Λειτουργούν υποστηρικτικά σε ένα ευρύ φάσμα εφαρμογών γλωσσικής τεχνολογίας.

## **Αλλά...**

- Τα αποτελέσματα εξαρτώνται άμεσα από την ποσοτική και ποιοτική σύσταση των ΗΣΚ που χρησιμοποιούμε.
- Η φύση των πληροφοριών που αντλούνται είναι πιθανοτική και όχι κατηγορική.
- Δεν καλύπτεται το σύνολο των γλωσσικών φαινομένων.

## Επίπεδα λεξιλογίου, συχνότητα και κάλυψη

Επίπεδο λεξιλογίου	Συχνότητα	Αριθμός λέξεων	Κάλυψη
Βασικό	Πολύ υψηλής συχνότητας	έως 2.000	80%
Επάρκειας	Υψηλής συχνότητας	έως 3.000	85%
Επιστημονικό	Μεσαίας συχνότητας	800	8% (σε ακαδημαϊκά κείμενα)
Ορολογίας	Μεσαίας συχνότητας	1.000-2.000 (σε κάθε αντικείμενο)	3% (σε ειδικευμένα κείμενα)
Απαιτητικό	Χαμηλής συχνότητας	100.000	2%



# Αντίστροφος κατάλογος συχνότητας

Σειρά	Λέξη	Συχνότητα	Κάλυψη %
90640	ευφυής	46	0,13
90641	αυτοφυής	3	0,01
90642	μεγαλοφυής	6	0,02
90643	ιδιοφυής	16	0,05
90644	συμφυής	4	0,02
90645	διφυής	6	0,02
90646	αφυής	1	
90647	Καρλσρούης	3	0,01
90648	Γενούης	4	0,01
90649	Λούης	227	0,10
90650	παδούης	1	
90651	πολύβουης	2	
90652	βουής	1	
90653	Βιζύης	51	0,06
90654	Γενεύης	122	0,28
90655	χλεύης	5	0,02

# Αντίστροφος κατάλογος συχνότητας (αξία)

- Ο αντίστροφος κατάλογος συχνότητας είναι ιδιαίτερα χρήσιμος για γλώσσες που διαθέτουν πλούσια μορφολογία όπως η ελληνική, ειδικά όταν το σώμα κειμένων δεν είναι επισημειωμένο.
- Για παράδειγμα, από δεδομένα όπως του Πίνακα μπορούμε με ευκολία να εντοπίσουμε τις λέξεις που έχουν κοινή κατάληξη (-ης) και να μελετήσουμε αν **οι καταλήξεις αυτές αποτελούν επιθήματα, πόσες λέξεις σχηματίζουν αυτά, πόσο συχνές είναι αυτές οι λέξεις και γενικά πόσο παραγωγικά είναι..**
- τα στοιχεία αυτά δεν είναι προβλέψιμα από τη διαίσθηση των ομιλητών: όσο ανεπτυγμένο λεξιλόγιο κι αν έχει **κανείς, δεν θα μπορούσε να φανταστεί όλες τις λέξεις που έχουν δημιουργηθεί με αυτά τα επιθήματα.**
- Τα σώματα κειμένων μάς επιτρέπουν να διαπιστώσουμε ποιες λέξεις πραγματικά έχουν δημιουργηθεί και πώς χρησιμοποιούνται αντί να διατυπώνουμε εικασίες για το ποιες θα μπορούσαν να δημιουργηθούν

## Κατάλογος συχνότητας των προφορικών δεδομένων συνομιλίας του ΣΕΚ

- οι κατάλογοι συχνότητας προσφέρουν στοιχεία για τις ιδιαιτερότητες των κειμενικών ειδών.
- Ας συγκρίνουμε, για παράδειγμα, τον Πίνακα με τους 10 πιο συχνούς τύπους στο σύνολο του ΣΕΚ με τον Πίνακα με τους 10 πιο συχνούς τύπους σε 300.000 περίπου λέξεις προφορικών δεδομένων από συνομιλίες του ΣΕΚ

Σειρά	Λέξη	Συχνότητα	Ποσοστό (%)	Κάλυψη (%)
1	και	10.839	3,47	100,00
2	το	9.842	3,15	100,00
3	να	9.098	2,91	100,00
4	ναι	5.891	1,88	100,00
5	είναι	5.854	1,87	100,00
6	δεν	5.724	1,83	100,00
7	θα	4.320	1,38	100,00
8	που	3.740	1,20	100,00
9	τα	3.719	1,19	100,00
10	ε	3.511	1,12	100,00

## Κατάλογος συχνότητας των προφορικών δεδομένων συνομιλίας του ΣΕΚ

- οι γραμματικοί τύποι μονοπωλούν και στα προφορικά κείμενα συνομιλίας τις πρώτες θέσεις του καταλόγου συχνότητας, όπως συμβαίνει σε όλα τα κειμενικά είδη.
- Παρ' όλα αυτά, έχει ιδιαίτερο ενδιαφέρον η υψηλή θέση του **να**, αλλά και των αρνητικών **δεν** και **δε** και η σχετικά χαμηλή θέση γενικών όπως **του** κ.ά., η παρουσία πολλών αντωνυμιών, ιδίως α' προσώπου (*μου, με, εγώ, σου*), η ύπαρξη μορίων όπως **ε** και **α** των δεικτών αλληλεπίδρασης **ναι** και **όχι**, αλλά και του φατικού **ρε** κ.λπ.
- Σημαντικό ότι οι περισσότεροι από τους 30 πιο συχνούς λεξικούς τύπους στο κειμενικό είδος της συνομιλίας απαντούν σε όλα τα κείμενα (κάλυψη 100%), κάτι που σημαίνει **ότι το λεξιλόγιο του προφορικού λόγου είναι πιο περιορισμένο σε σχέση με εκείνο του συνόλου των κειμένων του ΣΕΚ και σταθερά επαναλαμβανόμενο**

## Λεξικά συμπλέγματα στο υποσώμα συνομιλιών του ΣΕΚ.

- τα πιο συχνά συμπλέγματα στο υποσώμα προφορικών συνομιλιών του ΣΕΚ (700.000 λέξεις περίπου)

	Λεξικό σύμπλεγμα	Συχνότητα
1	ρε παιδί μου	423
2	να σου πω	372
3	ναι ναι ναι	165
4	δεν μπορώ να	88
5	σου πω κάτι	83
6	και μου λέει	82
7	να σου πω κάτι	79
8	και τα λοιπά	78
9	εν τω μεταξύ	66
10	δεν ξέρω αν	64
11	αυτό το πράγμα	63
12	πώς το λένε	62

13	δεν ξέρω τι	59
14	τι να πω	53
15	θέλω να πάω	50
16	είναι πάρα πολύ	47
17	δεν μπορείς να	46
18	τι να σου	46
19	το θέμα είναι	46
20	με το που	44
24	αυτό είναι το	42
25	δε μ' αρέσει	42

## Λεξικά συμπλέγματα στο υποσώμα συνομιλιών του ΣΕΚ. (στατιστικά στοιχεία)

- Όπως μπορούμε να παρατηρήσουμε, στα πιο συχνά λεξικά συμπλέγματα περιλαμβάνονται φράσεις που συνδέονται:
  1. **με την οικειότητα** της αλληλεπίδρασης στα συγκεκριμένα δεδομένα (ρε παιδί μου),
  2. **οργανωτικά στοιχεία** του κειμένου όπως μεταγλωσσικές εκφράσεις (να σου πω (κάτι), πώς το λένε, τι να πω),
  3. **εισαγωγικές φράσεις** ευθέος λόγου (και μου λέει) και αναφορικές φράσεις (αυτό το πράγμα, το θέμα είναι)
  4. **συμπλέγματα στάσης**, κυρίως με αρνητικά μόρια (δεν μπορώ να, δεν ξέρω αν/τι, δεν μπορείς να, δε μ' αρέσει)
  5. αλλά και **στερεότυπες εκφράσεις** που λειτουργούν ως ένα λέξημα, γίνονται δηλαδή αντιληπτές στα ελληνικά ως μία λέξη (και τα λοιπά, εν τω μεταξύ)

Λέξη-κλειδί	Συχνότητα (Αντόναρος)	%	Συχνότητα (Πεταλωτής)	%	Σημαντικότητα	P value
επίσης	25	0,11	1		27,09	0,0000001917
πυροσβεστικά	18	0,08	0		24,59	0,0000007062
άλλωστε	23	0,10	1		24,52	0,0000007342
πέρα	8	0,04	40	0,18	-23,98	0,0000009713
ώστε	20	0,09	63	0,29	-24,34	0,0000008045
έτσι	15	0,07	55	0,25	-25,17	0,0000005218

- Λέξεις-κλειδιά από τη σύγκριση δύο σωμάτων κειμένων κυβερνητικών εκπροσώπων

# Λέξεις – Κλειδιά

- Όπως φαίνεται στον Πίνακα , οι λέξεις-κλειδιά μπορεί να είναι λεξικοί (π.χ. πυροσβεστικά), αλλά και γραμματικοί τύποι (π.χ. επίσης, άλλωστε).
- Οι **λέξεις-κλειδιά** με μεγαλύτερο **λεξικό περιεχόμενο** συνήθως προσφέρουν στοιχεία για το θέμα των κειμένων που μελετώνται (λ.χ. εδώ την περίπτωση των πυρκαγιών του 2008), ενώ όσες έχουν **γραμματική λειτουργία** αποτελούν συχνά ένδειξη ύφους, γιατί ακριβώς δεν μπορούν να ελεγχθούν από τους ομιλητές, και επομένως μπορούν να μελετηθούν ως πιθανά στοιχεία της ιδιολέκτου τους.
- Αξίζει επίσης να σημειωθεί ότι οι λέξεις-κλειδιά είναι δυνατόν να είναι θετικές ή αρνητικές.
- Σημαντική είναι και η χρήση λέξεων-κλειδιών στον εντοπισμό του ειδικευμένου λεξιλογίου μιας θεματικής περιοχής ή ενός κειμενικού είδους



## Λέξεις-κλειδιά σε προφορικά κείμενα του ΣΕΚ με βάση το φύλο των συμμετεχόντων

Λέξη-κλειδί	Συχνότητα σε κείμενα μόνο με γυναίκες	Κάλυψη (αριθμός κειμένων)	Συχνότητα σε κείμενα με γυναίκες και άντρες	Σημαντικότητα
μου	2984	89	2742	77,20
είχαμε	295	68	189	40,96
Μαρία	111	27	43	39,98
μετά	704	88	582	36,21
σχολή	209	55	123	36,17
μαθήματα	145	52	74	33,97
κοπέλα	116	43	57	29,01
τους	521	82	855	-39,60
εδώ	403	82	705	-44,04
μαλάκα	6	4	83	-54,57
ο	1437	89	2257	-82,70

Συχνότητα σε άνδρες

# Στοιχεία

- Τα στοιχεία που παρέχει ο Πίνακας είναι ενδεικτικά για τις λεξικές επιλογές των φύλων στις συνομιλίες του ΣΕΚ. Έτσι, η σημαντικότερη λέξη-κλειδί στις συνομιλίες μεταξύ γυναικών είναι ο τύπος **μου**, που αντιστοιχεί τόσο στην προσωπική όσο και στην κτητική αντωνυμία α' προσώπου και αποτελεί χαρακτηριστική ένδειξη εμπλοκής.
- Οι **λέξεις-κλειδιά** που είναι λεξικοί τύποι αντανakλούν τη θεματική των συνομιλιών, καθώς οι συμμετέχουσες σε αυτές είναι κυρίως φοιτήτριες. (Για τη λέξη κοπέλα).
- Αντίθετα, η συμβολή των **αντρών** στις συνομιλίες **φαίνεται** να έγκειται σε πολύ περισσότερες αναφορές σε **ουσιαστικά αρσενικού** γένους, όπως φαίνεται από τους τύπους **ο** και **τους** του οριστικού άρθρου, αλλά και την κοινή **βωμολοχία μαλάκα**, που εμφανίζεται σπάνια στα κείμενα μόνο με γυναίκες (γύρω στο 2% των συνομιλιών).
- Τέλος, χρήζει περαιτέρω αναζήτησης η παρουσία του χρονικού **μετά** στις λέξεις-κλειδιά σε συνομιλίες μόνο με γυναίκες, σε αντίθεση με το τοπικό **εδώ** σε συνομιλίες στις οποίες συμμετέχουν και άντρες.

# Λέξεις-κλειδιά σε προφορικά κείμενα του ΣΕΚ με βάση την ηλικία.

Λέξη-κλειδί	Συχνότητα σε κείμενα μόνο με νέους	Κάλυψη (αριθμός κειμένων)	Συχνότητα σε κείμενα με μεικτή ηλικία συμμετεχόντων	Σημαντικότητα
ρε	2519	118	465	195,57
ντάξει	924	94	156	85,01
γώ	700	100	113	69,53
μου	4510	124	1326	52,73
βασικά	346	92	46	45,22
φίλε	234	23	25	38,41
ναι	7680	124	2461	36,91
παιδί	888	113	208	34,82
σχολή	297	80	46	31,16
εντάξει	1380	119	370	29,51
συ	267	83	40	29,43
άνθρωποι	65	41	62	-29,86
λοιπά	82	40	74	-32,35
προς	87	48	77	-32,44
παιδάκια	18	14	33	-35,12
γιαγιά	58	25	62	-36,11
μικρά	34	21	47	-38,27
θείος	23	16	41	-43,01
τους	884	117	476	-44,83
οι	971	123	522	-48,81
εδώ	718	114	409	-49,93
θεία	46	19	79	-81,81
τα	4143	124	1976	-90,57
μαμά	115	48	167	-147,97

# Στοιχεία

- Είναι φανερό ότι στις νεανικές συνομιλίες κυριαρχούν φαινομενικά στοιχεία και δείκτες λόγου όπως **ρε, ντάξει** και **εντάξει, γώ** (από τη φράση ξέρω 'γώ), **βασικά**, οι τύποι **μου**, όπως και παραπάνω, και **συ** (από το σύμπλεγμα ρε συ) και θεματικά στοιχεία όπως **σχολή**. Τα στοιχεία αυτά φαίνεται να χαρακτηρίζουν πρωτοτυπικά τον νεανικό λόγο στις συγκεκριμένες συνομιλίες.
- Αντίθετα, στις συνομιλίες στις οποίες συμμετέχουν και **μεγαλύτεροι** αντικατοπτρίζονται στις λέξεις-κλειδιά οι **συγγενικές** σχέσεις που συνδέουν τους τελευταίους με τους νεότερους (π.χ. *μαμά, θεία, θείος, γιαγιά*), αλλά και τα **ενδιαφέροντα** σε αυτού του είδους συνομιλίες (π.χ. *παιδάκια, άνθρωποι*). Ιδιαίτερο ενδιαφέρον για περαιτέρω έρευνα έχει η εμφάνιση του **επισημότερου προς** και του **λοιπά** (από το σύμπλεγμα *και τα λοιπά*). Συμπερασματικά, η εικόνα που προκύπτει από μια απλή αναζήτηση των λέξεων-κλειδιών σε προφορικά κείμενα του **ΣΕΚ συνάδει με τις δαισθητικές αντιλήψεις που έχουμε ως ομιλητές της ελληνικής για τις τυπικές λεξιλογικές επιλογές σε σχέση με το φύλο και την ηλικία των συμμετεχόντων.**

## Σημασίες των λέξεων άνδρας/άντρας και γυναίκα σε κείμενα του ΣΕΚ.

Φύλο	(1) στον <b>άντρα</b> το κόμπλεξ μου τη σπάει	(2) το Κοράνι σέβεται τη <b>γυναίκα</b>
Άτομο	(3) ένας ρακένδυτος <b>άνδρας</b> μαζεύει τα απομεινάρια	(4) συνάντησα μια <b>γυναίκα</b> σε αξιοθρήνητη κατάσταση
Σύζυγος	(5) ο <b>άνδρας</b> της αγόρασε τη δεκατριάχρονη	(6) μπήκαμε σε μεγάλο δίλημμα με τη <b>γυναίκα</b> μου
+ Επαγγελματικό	(7) να δουλέψεις με πέντε <b>άντρες</b> ηλεκτρονικούς	(8) το Λουξεμβούργο δεν έχει καμία <b>γυναίκα</b> ευρωβουλευτή
Ενήλικος/Ενήλικη	(9) τα «παιδιά» σας γίνανε <b>άνδρες</b>	(10) είναι <b>γυναίκα</b> , δεν είναι κοριτσάκι πλέον η Δήμητρα
Προσωπικό	(11) οι <b>άνδρες</b> της προσωπικής του φρουράς	
Πρόσωπο με κύρος	(12) οι δύο <b>άνδρες</b> συζήτησαν θέματα Μεσογειακής συνεργασίας	
Στερεοτυπικός ρόλος	(13) έχει να κάνει με <b>άντρα</b> και όχι με... ρόμπα	
Μέλος οικογένειας	(14) οι <b>άντρες</b> της οικογένειας το έχτισαν με τα χέρια τους	
Παράνομη		(15) η παράνομη διακίνηση <b>γυναικών</b>
Οικιακή βοηθός		(16) ... τις δουλειές του σπιτιού σε κάποια <b>άλλη γυναίκα</b>
Ειδική δομή		(17) για να έχει την ησυχία της η <b>γυναίκα</b>

# Στοιχεία

- οι σημασίες που αναφέρονται αποκλειστικά στους άντρες είναι κατά βάση ευνοϊκές, ενώ όσες αναφέρονται μόνο σε γυναίκες είναι υποτιμητικές.
- οι κοινές συνάψεις (των λέξεων άνδρας – γυναίκα) αναφέρονται στην εμφάνιση, τη σεξουαλικότητα, την ισότητα, τις σχέσεις και την εξουσία, οι συνάψεις του **άνδρας/άντρας** σχετίζονται με τον **δημόσιο βίο** και τη **σύγκρουση**, ενώ οι συνάψεις του **γυναίκα** με τα **προβλήματα** που αντιμετωπίζουν, την **κακοποίηση**, την **πορνεία**, τη **δουλειά**, τη **συμμετοχή** στην **κοινωνία** ή με ομάδες μη **προνομιούχων** ατόμων.
- Εξίσου ενδιαφέρον έχουν οι συνάψεις που χρησιμοποιούνται στα **γυναικεία** περιοδικά σε αντίθεση με τα **ανδρικά** περιοδικά, τα **ποικίλης ύλης** και τις **εφημερίδες**. Για παράδειγμα, τα ανδρικά περιοδικά δεν αποφεύγουν το **χυδαίο λεξιλόγιο** όταν **αναφέρονται στις γυναίκες**, ενώ τα γυναικεία περιοδικά δίνουν έμφαση στις συνάψεις που έχουν σχέση με την επαγγελματική καταξίωση των γυναικών.

# Στοιχεία

- Ένα ακόμη εύρημα **αφορά την ηλικία στην οποία αναφέρονται οι λέξεις** *αγόρι* και *κορίτσι*.  
 Παραδείγματα όπως τα παρακάτω είναι τυπικά για το ότι η λέξη **κορίτσι** αποδίδεται σε **γυναίκες**, ενώ κάτι αντίστοιχο δεν συμβαίνει για τη λέξη **αγόρι**.  
*... τι μήνα του μέλιτος να κάνει το **κορίτσι** με τον Πάγκαλο;*  
*... και νυν κυρία Καραμανλή. Και όμορφο **κορίτσι** είναι ...*
- Τέλος, μια από τις πιο ενδιαφέρουσες παρατηρήσεις είναι ότι η λέξη γυναίκα χρησιμοποιείται πολύ περισσότερο με τη σημασία του **μεμονωμένου ατόμου** απ' ό,τι η λέξη άντρας, κάτι που ίσως οφείλεται στο ότι είναι πιο πιθανό να χρησιμοποιηθεί στον προφορικό λόγο το ψευδο-γενικευτικό άνθρωπος για την αναφορά σε πρόσωπο αρσενικού φύλου (Makri-Tsilirakou 1989).

## Ομιλητές – Ηλικία - Φύλο(στοιχεία)

- Αυτό που βρέθηκε είναι ότι κάθε ομάδα ομιλητών (π.χ. νέες γυναίκες) τείνει να **χρησιμοποιεί περισσότερο** τις **λέξεις** που την **χαρακτηρίζουν** (π.χ. κοπέλα, κορίτσι στην περίπτωση αυτή). Επίσης, οι γυναίκες κυρίως μιλούν για συγκεκριμένα άτομα, άντρες ή γυναίκες, στις συνομιλίες τους, ενώ οι άντρες τείνουν να **μιλούν** περισσότερο για το φύλο, π.χ. σε συγκρίσεις του τι κάνουν γενικά οι άντρες και οι γυναίκες.
- Έρευνα της Hatzidaki (2011), που είναι η μόνη άλλη ανάλυση για το φύλο σε ελληνικά σώματα κειμένων, η οποία βρίσκει ότι τόσο τα **αντρικά όσο** και τα **γυναικεία περιοδικά** παρουσιάζουν **ομοιότητες** στα **λεξικά** συμπλέγματα **τροπικότητας** (προφορικός - γραπτός λόγος) που χρησιμοποιούν (π.χ. *θα πρέπει να, για να μην, καλό είναι να κ.ά.*).



# Κατάλογος συχνότερων λέξεων σε κοινοβουλευτικές ομιλίες των ετών 1911 και 1990-2010

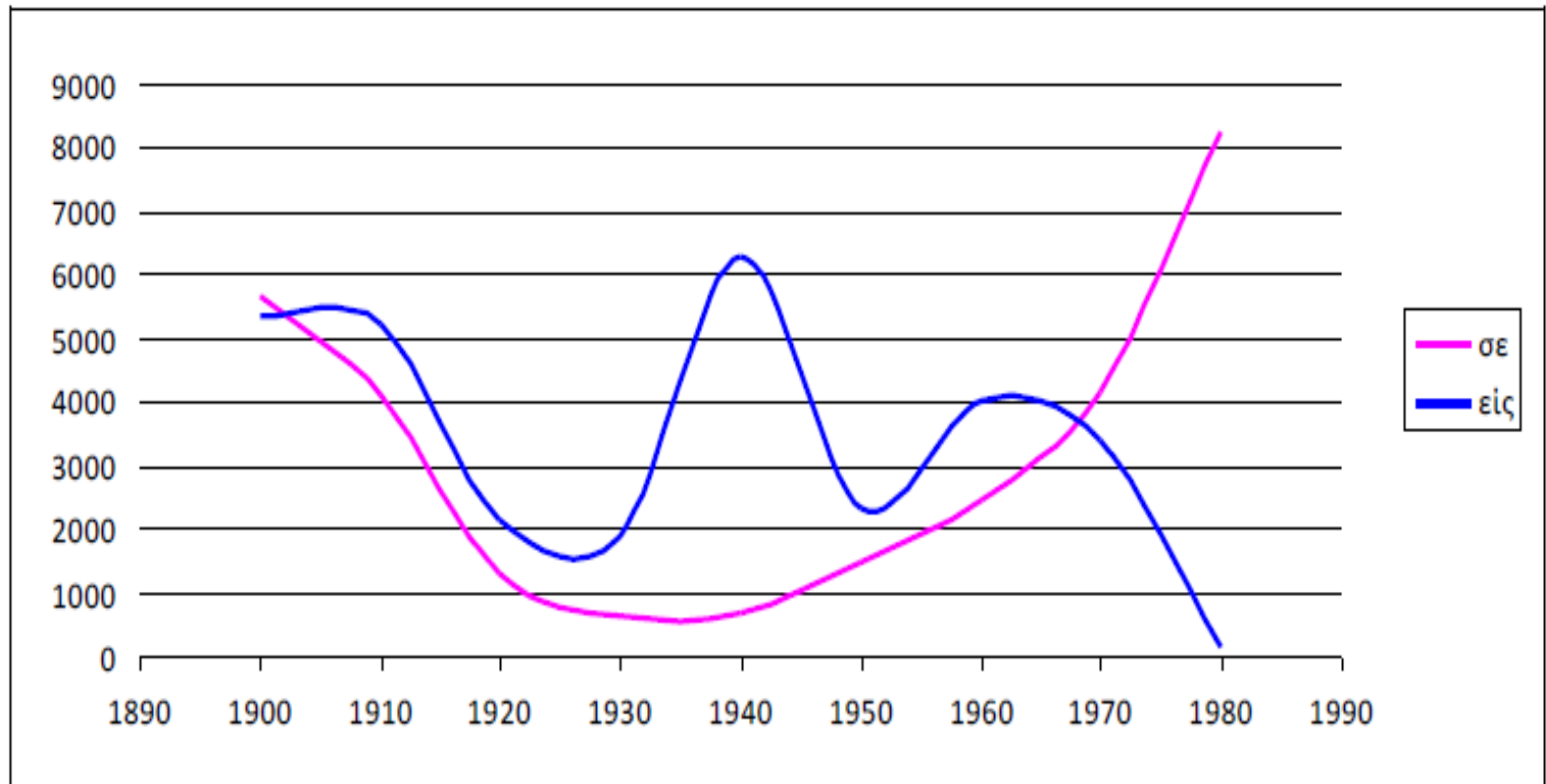
Δεδομένα 1911	Δεδομένα 1990-2010
καί	και
νά	το
τήν	να
τῆς	του
ή	της
τοῦ	την
τό	η
εἰς	με
ὅτι	που
εἶναι	για
τῶν	των
δέν	εἶναι
ό	δεν
πρός	τα
έν	από

τόν	θα
θά	ὅτι
τά	ο
διά	σε
τάς	η
κύριοι	τη
ὅπως	στην
τούς	στο
ἀπό	τις
κατά	αυτό

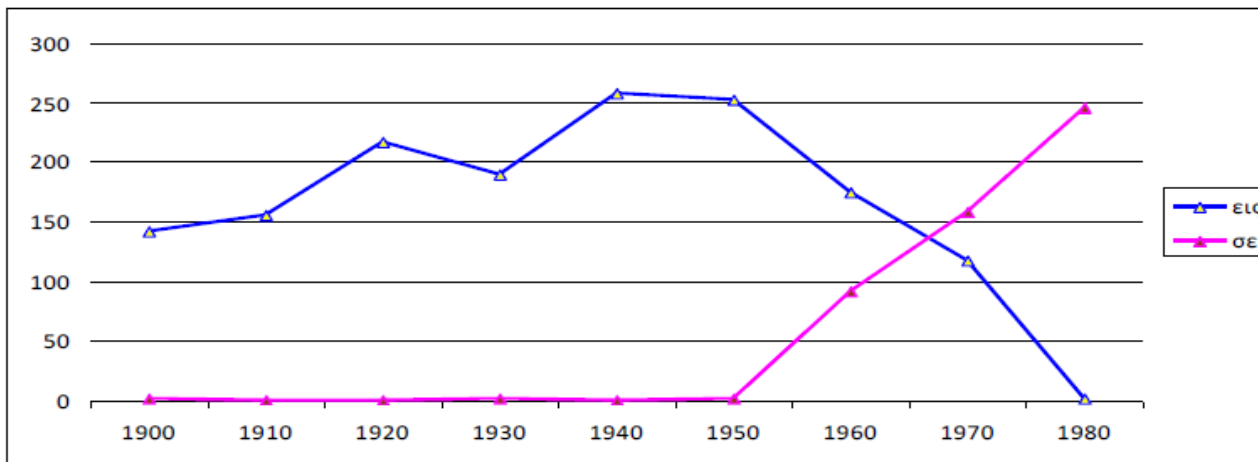
## Κατάλογος συχνότερων λέξεων σε κοινοβουλευτικές ομιλίες των ετών 1911 και 1990-2010

- Μεταξύ άλλων, η σύγκριση των δεδομένων του 1911 με εκείνα της περιόδου 1990-2010 εμφανίζει ιδιαίτερο **ενδιαφέρον ως προς τις προθέσεις**: στα δεδομένα από το 1911 πολύ υψηλή συχνότητα εμφάνισης παρουσιάζουν προθέσεις όπως οι είς, πρὸς, ἐν, διὰ και κατά, ενώ στα δεδομένα του ΣΕΚ ιδιαίτερα συχνές προθέσεις είναι οι με, για, σε/στην/στο. Η μόνη πρόθεση που είναι κοινή στις δύο λίστες είναι η από. Τα στοιχεία αυτά υποδεικνύουν τις **εκτεταμένες αλλαγές στο σύστημα των προθέσεων** από το 1911 έως σήμερα.

## Συχνότητα των τύπων σε και είς στα δεδομένα του Διαχρονικού ΣΕΚ του 20ού αιώνα

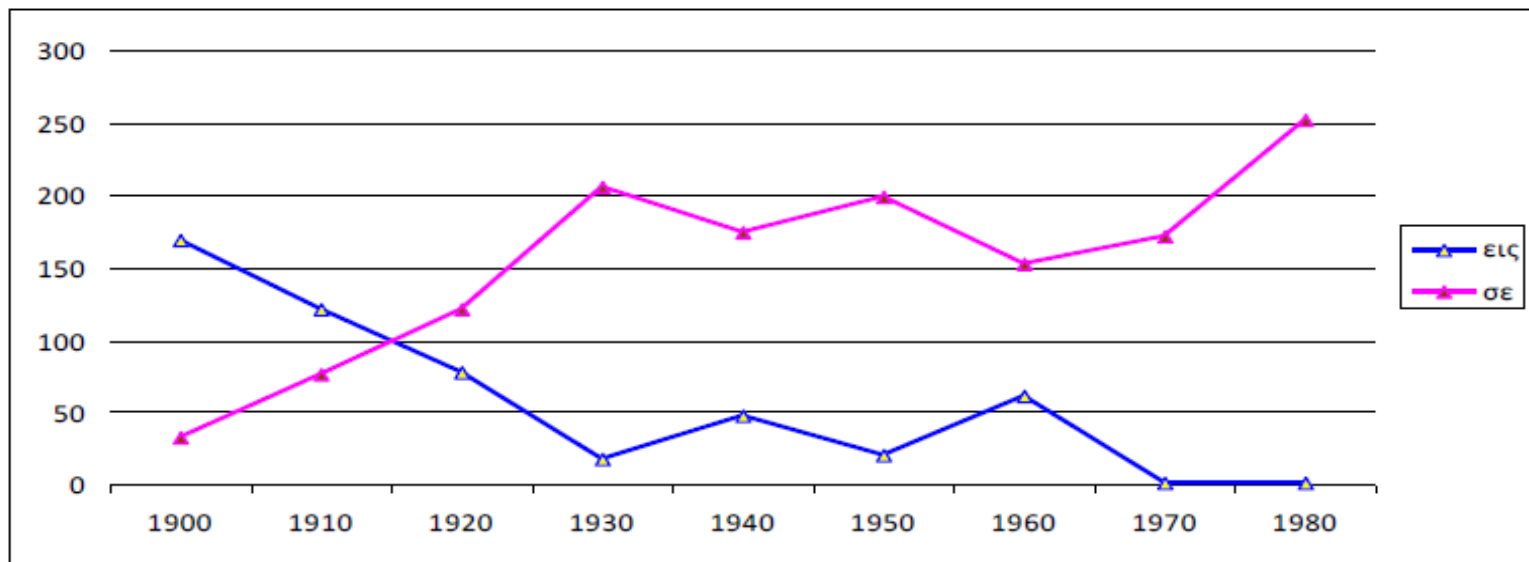


- Όπως μπορούμε να δούμε στα δύο αυτά σχήματα, στο κειμενικό είδος των δημόσιων ομιλιών ο τύπος *είς* είναι ο μόνος που εμφανίζεται έως τη δεκαετία του 1950, οπότε αρχίζει η κατακόρυφη πτώση του και η απότομη άνοδος του σε.



*Συχνότητα των τύπων σε και είς στο κειμενικό είδος των δημόσιων ομιλιών*

- Στο κειμενικό είδος **των ιδιωτικών επιστολών**, αντίθετα, η αλλαγή αυτή είναι πιο **σταδιακή** και η επικράτηση του δημοτικού τύπου σε αρχίζει από τη δεκαετία του 1920



**Συχνότητα των τύπων σε και εις στο κειμενικό είδος των ιδιωτικών επιστολών.**

# Λέξεις-κλειδιά στα λογοτεχνικά έναντι των νομοθετικών/διοικητικών κειμένων του ΣΕΚ.

	Λέξη-κλειδί	Συχνότητα (λογοτεχνικά κείμενα)	Συχνότητα (νομοθετικά/ διοικητικά κείμενα)	Σημαντικότητα
1	μου	24330	8	23438,54
2	κι	24270	102	22383,23
3	ήταν	13423	180	11409,84
4	σου	10886	2	10487,86
5	να	74658	22368	10463,21
6	μα	12578	452	9099,89
7	τι	10072	146	8483,68
8	θα	23063	3455	8416,08
9	είχε	9795	375	6973,58
10	το	69488	24920	6074,45
11	απ	8073	282	5874,88
12	μας	9537	621	5732,33
13	είπε	5587	1	5378,55
14	πως	6482	125	5261,35
15	γιατί	5771	36	5205,03
16	σαν	6906	219	5134,64
17	τον	26980	6864	5123,67
18	εγώ	5242	0	5062,41

## Λέξεις-κλειδιά στα λογοτεχνικά έναντι των νομοθετικών/διοικητικών κειμένων του ΣΕΚ.

29	μάτια	3071	5	2900,16
30	έτσι	4363	241	2784,35
31	σε	14248	15508	-2356,48
32	της	36342	33090	-2570,56
33	από	23659	23524	-2636,22
34	οποία	573	3021	-3204,83
35	κάθε	2306	5564	-3399,18
36	του	55080	52090	-4847,51
37	απόφαση	272	3593	-5200,34
38	κατά	1210	7277	-8209,38
39	ή	3902	21064	-22718,96
40	των	7221	27570	-24601,46

# Στοιχεία

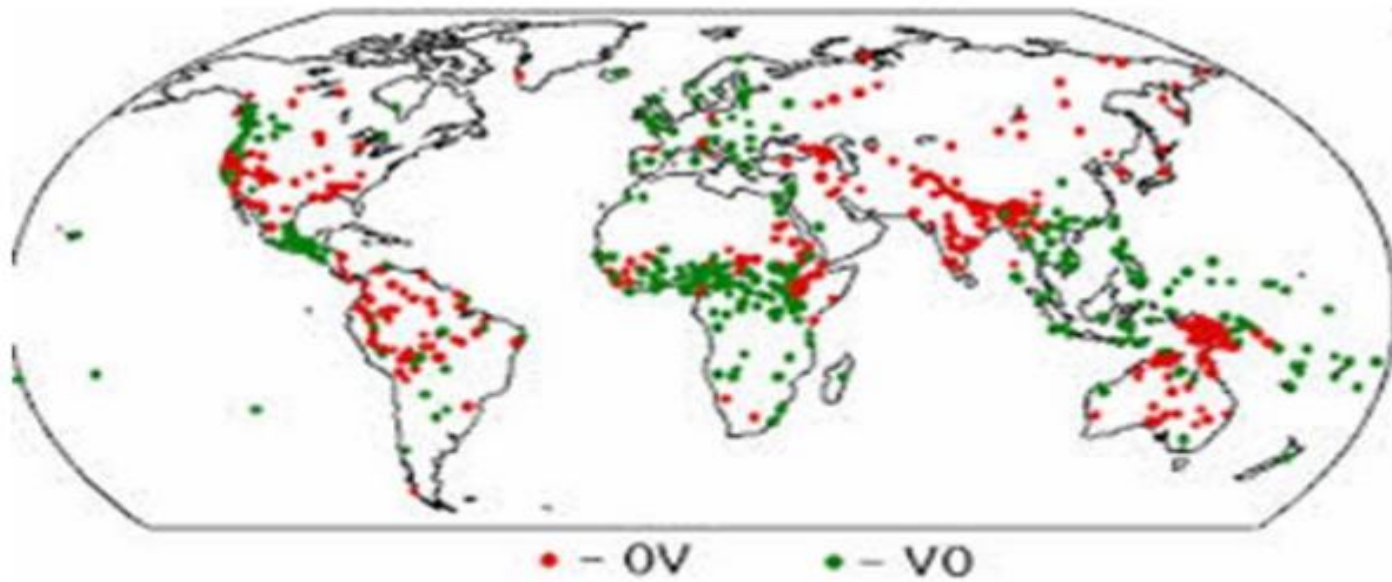
- Στις συχνότερες λέξεις-κλειδιά στα **λογοτεχνικά κείμενα** περιλαμβάνονται προσωπικές αντωνυμίες (μου, σου, μας, εγώ), σύντομοι ή συγκεκριμένοι τύποι (κι, απ), δείκτες τροπικότητας (να, θα), ρήματα στον αόριστο (ήταν, είχε, είπε), αρνήσεις (δεν, δε), λεξιλόγιο μελών του σώματος (μάτια), **γλωσσικά στοιχεία** δηλαδή που ο Biber θα θεωρούσε ότι **χαρακτηρίζουν τη δημιουργία με εμπλοκή, τον αφηγηματικό λόγο και το μη απρόσωπο ύφος**
- οι συχνότερες λέξεις-κλειδιά στον **νομοθετικό/διοικητικό λόγο** (που δηλώνονται με αρνητικό πρόσημο στον πίνακα) είναι γενικές άρθρων (των, του, της), τύποι της αντωνυμίας (οποίος), προθέσεις σε πλήρεις τύπους (κατά, από, σε), το διαζευκτικό ή κ.ά., στοιχεία δηλαδή που χαρακτηρίζουν τον επίσημο λόγο στα ελληνικά
- Με παρόμοιο τρόπο μπορούμε να χαρτογραφήσουμε το λεξιλόγιο, αλλά και τις κυρίαρχες γραμματικές επιλογές στα κειμενικά είδη του ΣΕΚ, μέσω δηλαδή της **αμοιβαίας σύγκρισης των πιο σημαντικών συχνών λέξεων** που απαντούν σε καθένα από αυτά.
- Σε κάθε περίπτωση τα σώματα κειμένων όχι μόνο επισημαίνουν τον ρόλο των κειμενικών ειδών στη λειτουργία της γλώσσας, αλλά και προσφέρουν τα εργαλεία για να εντοπίσουμε και να αξιολογήσουμε τις διαφορές μεταξύ τους



# ΤΥΠΟΛΟΓΙΑ ΓΛΩΣΣΩΝ

# Τυπολογία

## Εισαγωγή



# Τι είναι η Τυπολογία;

- Σημαντικός – ίσως ο πιο σημαντικός (;) – κλάδος της Γλωσσολογίας
- Προσπαθεί έμμεσα να απαντήσει στο βασικό ερώτημα «Τι είναι γλώσσα;»

## Τι ερευνά η Τυπολογία;

- Η τυπολογία είναι η αναζήτηση δομικών ομοιοτήτων μέσα στο χάος της διαγλωσσικής ποικιλίας
- Ερευνά τις δομικές ομοιότητες που υπάρχουν ανάμεσα στις γλώσσες

# Τυπολογία

- Η Τυπολογία αποτελεί έναν από τους πιο σημαντικούς κλάδους της Γλωσσολογίας, στα πλαίσια του οποίου εκπονείται ένας **σημαντικός αριθμός ερευνητικών εργασιών**.
- Το βασικότερο αντικείμενο της Τυπολογίας είναι η απάντηση στο ερώτημα «**Τι είναι γλώσσα;**» μέσα από έναν συγκεκριμένο τρόπο: Αυτόν της περιγραφής, **ανάλυσης** και **ποσοτικής** διερεύνησης των **χαρακτηριστικών** όλων των γλωσσών του κόσμου.
- Κυρίως στόχο της Τυπολογίας αποτελεί η αναζήτηση της **σύγκλισης, των ομοιοτήτων** μεταξύ των γλωσσών, δεδομένης της τεράστιας ποικιλίας (και επομένως αμέτρητων διαφορών) που υπάρχει ανάμεσα στις γλώσσες του κόσμου

# Τυπολογία

- Το αντικείμενο έρευνας λοιπόν της Τυπολογίας είναι η αναζήτηση, η επισήμανση και η **ανάλυση των δομικών ομοιοτήτων** που υπάρχουν ανάμεσα σε κάποιες (ή στις περισσότερες από τις) γλώσσες του κόσμου
- Εμφάνιση και εξέλιξη του κλάδου: Ο όρος «Τυπολογία» προέρχεται από την **Βιολογία**, όπως και πολλοί άλλοι όροι που εισήλθαν στην Γλωσσολογία κατά τον 19ο αι. Στη Βιολογία σημαίνει την **ταξινόμηση** σε είδη με βάση **εξωτερικά** («μορφολογικά») **χαρακτηριστικά** και **όχι γενετικά**, και το ίδιο ακριβώς σημαίνει και για τις γλώσσες: την **ταξινόμησή** τους με **βάση εξωτερικά** τους **χαρακτηριστικά** (π.χ. σειρά των όρων, παρουσία ή απουσία πτώσεων κλπ.) και **όχι με βάση πιθανή συγγένεια** (αν ανήκουν δηλ. στην ίδια γλωσσική οικογένεια) –
- Στην Γλωσσολογία ο όρος χρησιμοποιήθηκε πρώτη φορά από τους Νεογραμματικούς ήδη από τα τέλη του 19ου αι. (von der Gabelentz, 1891). Στη συνέχεια η Τυπολογική έρευνα αναπτύχθηκε κυρίως από τον Αμερικανικό δομισμό (με προεξάρχουσα μορφή τον Sapir), χάρη στην περιγραφή πολλών Ινδιάνικων γλωσσών της Β. Αμερικής. Δεν υπήρχε όμως η έννοια της Τυπολογίας ως ξεχωριστού κλάδου της Γλωσσολογίας, αυτό έγινε με τις έρευνες του διάσημου ερευνητή J.Greenberg ('60). Τις τελευταίες δεκαετίες ο κλάδος γνωρίζει μεγάλη άνθηση

# Γλωσσικές ομοιότητες και γλωσσικοί τύποι

- Τι μπορούν να μας πουν οι ομοιότητες; Δεδομένης της διαγλωσσικής ποικιλίας σε παγκόσμιο επίπεδο, τα σημεία σύγκλισης ανάμεσα στις γλώσσες είναι αυτά που ενδιαφέρουν περισσότερο την γλωσσολογική κοινότητα, μια και αυτά θα μπορούσαν να μας δείξουν τι είναι κοινό σε όλες ή στην μεγάλη πλειοψηφία των γλωσσών και, κατ' επέκταση, τι είναι κοινό στην γλωσσική ικανότητα όλων των ανθρώπων.
- Θα πρέπει να επισημανθεί ότι η ισοδυναμία των δομών που συγκρίνονται **βασίζεται στην σημασιολογία περισσότερο, παρά στην μορφή τους.** Με άλλα λόγια, κοιτάζουμε λ.χ. να δούμε αν υπάρχει Αιτιατική πτώση, δηλ. κάποιο μαρκάρισμα στο Αντικείμενο του ρήματος, ανεξάρτητα αν αυτό το μαρκάρισμα σε κάποιες γλώσσες γίνεται με πρόθημα, επίθημα, με ξεχωριστό κλιτικό ή αν δηλώνεται πάνω στο Ρήμα.

## 4 είδη εξήγησης

- Σε κάθε δομική ομοιότητα που παρατηρούμε σε ένα μεγάλο αριθμό γλωσσών, μπορούμε να δώσουμε μία από τις παρακάτω εξηγήσεις:
  - α) **Γλωσσική συγγένεια**: η ομοιότητα οφείλεται στην κοινή καταγωγή των γλωσσών και την διατήρηση χαρακτηριστικών από την «μητέρα-γλώσσα»
  - β) **Γλωσσική επαφή**: η ομοιότητα οφείλεται στην σύγκλιση που παρατηρείται ανάμεσα σε γλώσσες που βρίσκονται σε μακροχρόνια επαφή
  - γ) Η ομοιότητα μπορεί επίσης να οφείλεται σε κοινές **κοινωνιο-γλωσσικές ή περιβαλλοντικές συνθήκες** (π.χ. κοινωνική διαστρωμάτωση επιβάλλει τύπους ευγενείας ή παρόμοιο φυσικό περιβάλλον ωθεί προς ομοιότητες στο λεξιλόγιο). Συνήθως στο λεξιλογικό επίπεδο
  - δ) Όταν δεν μπορούμε να αιτιολογήσουμε την ομοιότητα με βάση **τίποτα από τα παραπάνω**, τότε θεωρούμε ότι απλά οι γλώσσες όπου εμφανίζεται το ίδιο χαρακτηριστικό ανήκουν σε έναν κοινό «τύπο»



# Γλωσσικές ομοιότητες και γλωσσικοί τύποι

- Με αυτόν τον τρόπο, δημιουργούνται όροι όπως «γλώσσες ΥΡΑ [με αυτή την σειρά των όρων]», «γλώσσες εργασιακές [με αυτό το σύστημα πτώσεων]» κ.ά.
- Οι γλωσσικοί τύποι μπορούν να αναφέρονται σε **οποιοδήποτε επίπεδο γλωσσικής ανάλυσης** (φωνολογία, μορφολογία, σύνταξη, λεξιλόγιο / σημασιολογία)
- Αυτό αποτελεί και βασική διαφορά ανάμεσα στην **Τυπολογική** και την **Ιστορικο-συγκριτική έρευνα**: η πρώτη **αφορά όλα τα επίπεδα**, ενώ η δεύτερη κυρίως την **φωνολογία και το λεξιλόγιο**

# Καθολικά

- Ο βασικός ερευνητικός στόχος της Τυπολογίας είναι η διατύπωση καθολικών
- Τι είναι τα **Καθολικά**: «**Εμπειρικά** διαπιστωμένες γενικεύσεις που περιγράφουν συστηματικές κατανομές για συγκεκριμένα γραμματικά φαινόμενα στις γλώσσες του κόσμου» (Cristofaro, 2010). Με άλλα λόγια, απόλυτες ή ισχυρές τάσεις που παρατηρούνται στις γλώσσες του κόσμου.
- Έμφαση στο **εμπειρικά**: Είναι **διαπιστώσεις ισχυρών τάσεων** στις υπαρκτές γλώσσες, **όχι υποθέσεις** με βάση τα δεδομένα κάποιων γλωσσών.

# Καθολικά

- Μπορούμε να ξεχωρίσουμε τα ακόλουθα είδη καθολικών:
- **Απόλυτα καθολικά:** Η διαπίστωση ισχύει για όλες τις γλώσσες του κόσμου, λ.χ. «Όλες οι γλώσσες έχουν φωνήεντα», «όλες οι γλώσσες έχουν τρόπους να παράγουν ερωτήσεις» κλπ.
- **Στατιστικά καθολικά:** Η διαπίστωση ισχύει για έναν μεγάλο αριθμό γλωσσών, όχι όμως για το σύνολό τους, λ.χ. «Οι περισσότερες γλώσσες έχουν τον φθόγγο /η/ (όχι όμως η Central Rotokas, από την Παπούα Νέα Γουινέα)». –
- **Επαγωγικά καθολικά:** Αυτά είναι της μορφής ‘Αν μια γλώσσα Χ, τότε Υ...’ Π.χ. «Αν σε μια γλώσσα το ρήμα έπεται τόσο του υποκειμένου όσο και του αντικειμένου, τότε αυτή η γλώσσα έχει πτωτικό σύστημα»

# Καθολικά

- Πώς δημιουργούνται τα καθολικά;
- Σύμφωνα με τους Τυπολόγους, αυτά **δεν είναι αποτέλεσμα κάποιας Καθολικής γραμματικής** (όπως θα υποστήριζαν οι εκπρόσωποι της Γενετικής Γραμματικής), αλλά **ισχυρών διαγλωσσικών πανανθρώπινων τάσεων** που σχετίζονται με την επεξεργασία των γλωσσικών δεδομένων από τον ανθρώπινο εγκέφαλο, αρχές οικονομίας της ανθρώπινης σκέψης κ.ά.
- Τα Τυπολογικά καθολικά είναι εξαιρετικά χρήσιμα και για την **Ιστορική γλωσσολογία**, καθώς μπορούν να **κατευθύνουν** τις **εικασίες** μας για γλωσσικές δομές του παρελθόντος με βάση τα όσα παρατηρούμε σήμερα στις γλώσσες του κόσμου

# Καθολικά

- Τέλος, μία παρατήρηση για τα δεδομένα στα οποία βασίζεται η Τυπολογία:
- Είναι προφανές ότι η **Τυπολογία χρειάζεται μια πληθώρα ποιοτικών δεδομένων** από πολλές γλώσσες. Αυτό μόνο τα τελευταία χρόνια έχει καταστεί δυνατό, ειδικά λόγω των μεγαλύτερων δυνατοτήτων δημιουργίας βάσεων δεδομένων και επεξεργασίας των δεδομένων με ηλεκτρονικά μέσα
- Υπάρχει όμως πολύ μεγάλη ανάγκη για μελέτες-πεδίου σε πάρα πολλές περιοχές του πλανήτη, ώστε να περιγραφούν όσο το δυνατόν περισσότερες γλώσσες. Οι περισσότερες είναι ακόμα τελείως άγνωστες.
- Το **ζήτημα του «γλωσσικού θανάτου» είναι απειλή** για την Τυπολογία αλλά και για την κατανόηση του φαινομένου της γλώσσας

# Γλωσσική ποικιλομορφία



# Γλωσσική ποικιλομορφία



# Με τον ίδιο τρόπο σε όλες τις γλώσσες του κόσμου..

- Τα παιδιά κατακτούν τα γλωσσικά ορόσημα με **τον ίδιο τρόπο ανεξάρτητα από τη γλώσσα** στην οποία εκτίθενται
- **6 – 8 μήνες:** αρχίζουν το «βάβισμα» (επαναλαμβανόμενες συλλαβές όπως *bababa*)
- **20 – 24 μήνες:** αρχίζουν να βάζουν τις λέξεις τη μια μετά την άλλη
- **2 – 3 χρόνια:** χρησιμοποιούν απαρέμφατα, παραλείπουν τα υποκείμενα, ομαλοποιούν ανώμαλους τύπους
- Κωφά παιδιά αρχίζουν το βάβισμα (με χειρονομίες) την ίδια χρονική στιγμή με τα ακούοντα παιδιά



# Το λογικό πρόβλημα στην Κατάκτηση Γλώσσας

- Τα παιδιά έχουν **πλούσια γλωσσική γνώση** που τους δίνει τη δυνατότητα να παράγουν άπειρο αριθμό προτάσεων, τη στιγμή που έχουν ακούσει μόνο έναν περιορισμένο αριθμό προτάσεων
- Μαθαίνουν βασιζόμενα σε «**θετικά**» στοιχεία
- **Μαθαίνουν** τη γλώσσα **χωρίς** να τους **πει** κάποιος ξεκάθαρα **ποιες δομές** είναι **λανθασμένες**
- Παρόλο που κάνουν λάθη, σε πολλές περιπτώσεις **αποφεύγουν να κάνουν λανθασμένες γενικεύσεις**

# Η έννοια της Γραμματικής

- Γνωρίζω μια γλώσσα σημαίνει γνωρίζω τη **γραμματική** της
- **Γραμματική:** ένα σύστημα πεπερασμένων κανόνων το οποίο βρίσκεται στο μυαλό των ομιλητών
- **Νοητική Γενετική Διαδικασία** (mental generative process, Chomsky): χρησιμοποιώντας ένα μικρό αριθμό κανόνων κατασκευάζουμε άπειρο αριθμό προτάσεων
- Γνώση της γραμματικής μας επιτρέπει να παράγουμε προτάσεις που δεν τις έχουμε ακούσει πριν και να κρίνουμε εάν μια πρόταση είναι γραμματική ή όχι

# Περιορισμοί

- Γλωσσικές αρχές οι οποίες απαγορεύουν κάποιες αλληλουχίες λέξεων, κάποιες γραμματικές λειτουργίες, κάποιους συσχετισμούς ήχων και σημασίας
- Υπάρχουν σε κάθε γλώσσα και είναι απαραίτατοι
- Περιορισμοί στη **μορφή της πρότασης**
- Περιορισμοί **στην ερμηνεία**

# Από που προέρχεται η γνώση μας για τη γλώσσα;

- ❖ Πώς γνωρίζουμε ότι μια πρόταση είναι αντιγραμματική, δεν μπορεί να έχει μια συγκεκριμένη σημασία ή είναι αμφίσημη;

## **4εκδοχές:**

- Μίμηση
- Ενίσχυση
- Συσχετισμοί
- Καθολική Γραμματική

# Κατάκτηση Γλώσσας μέσω Μίμησης

- ❖ **Τα παιδιά μαθαίνουν τη γλώσσα μιμούμενα τους ενήλικες**
  - Δεν υπάρχει απόλυτη αντιστοιχία μεταξύ των εισερχόμενων και εξερχόμενων δεδομένων
  - **Ενήλικες:** ερωτήσεις και προστακτικές
  - **Παιδιά:** καταφατικές προτάσεις
  - Τα παιδιά παράγουν «νεολογιστικές» προτάσεις και προτάσεις που δεν θα μπορούσαν να έχουν ακούσει ποτέ

# Κατάκτηση Γλώσσας μέσω Ενίσχυσης

- ❖ Τα παιδιά μαθαίνουν τη γλώσσα επειδή «**ενισχύονται θετικά**» όταν παράγουν κάτι σωστά και «**ενισχύονται αρνητικά**» όταν κάνουν λάθη (c.f. Skinner, 1957)
- Αυτό είναι κάτι που συμβαίνει, **αλλά δεν** μπορεί να εξηγήσει σε βάθος τη γλωσσική ικανότητα
- **Δεν** εξηγεί τον άπειρο αριθμό προτάσεων που παράγουμε
- Οι γονείς προσέχουν **ΤΙ** λένε τα παιδιά τους και όχι **ΠΩΣ** το λένε

# Κατάκτηση Γλώσσας

- **Έμφυτη γνώση** (Chomsky, 1959)
- Έμφυτοι περιορισμοί: κοινοί για όλες τις γλώσσες
- ❖ **Η έμφυτη γνώση εξηγεί:**
  - ✓ Γιατί είναι δυνατή η κατάκτηση γλώσσας, δεδομένης της διαφοροποίησης στις συνθήκες εκμάθησης
  - ✓ Ομοιότητες σε όλες τις γλώσσες ως προς τον τρόπο και χρόνο μάθησής τους

Όμως! **Δεν είναι ΟΛΗ** η γνώση έμφυτη!

Οι γλώσσες διαφέρουν (π.χ. κανόνας υποκειμένου για Ελληνικά και Αγγλικά)

**ΑΛΛΑ:** η ποικιλία αυτή δεν είναι χωρίς περιορισμούς

# Κατάκτηση Γλώσσας ως Καθολική Γραμματική

- **ΚΑΘΟΛΙΚΗ ΓΡΑΜΜΑΤΙΚΗ (ΚΓ):** το σύνολο των περιορισμών με τους οποίους είναι προικισμένοι ΟΛΟΙ οι άνθρωποι εκ γενετής και που ευθύνονται για τη διαδικασία κατάκτησης της γλώσσας. Η ΚΓ καθορίζει το εύρος της ποικιλίας και συνεπώς χαρακτηρίζει την έννοια της πιθανής ανθρώπινης γλώσσας.
- **ΚΑΤΑΚΤΗΣΗ ΓΛΩΣΣΑΣ:** αποτέλεσμα της διεπίδρασης των έμφυτων παραγόντων και του περιβάλλοντος. Η γλώσσα δεν «μαθαίνεται» αλλά «προκύπτει», «αναδύεται» την κατάλληλη στιγμή δεδομένου ότι το παιδί εκτίθεται σε κάποια μορφή της (ομιλούμενη ή νοηματική).





"Καλά Χριστούγεννα"

Σε όλους!!!!!!