



Το Διαχρονικό Σώμα Ελληνικών Κειμένων του 20ού αιώνα

Διονύσης Γούτσος

Διάλεξη στον Τομέα Γλωσσολογίας, 30/3/2016

Δομή παρουσίασης

- Πρόσφατη γλωσσική αλλαγή
- Διαχρονικό Σώμα Ελληνικών Κειμένων 20ού αι.
 - Στόχοι προγράμματος
 - Ιδιαιτερότητες ελληνικής
 - Αρχές σχεδιασμού
- Έρευνα: Γραμματικά είδη στη διαχρονία
- Συμπεράσματα
- Άλλα ερευνητικά ερωτήματα

Πρόσφατη γλωσσική αλλαγή I

- πρόσφατη γλωσσική αλλαγή: αλλαγή στο πρόσφατο παρελθόν που αφορά την τρέχουσα γλώσσα (Aarts et al. 2013: 2)
 - > πρόσφατη χρονικά
 - > «διαχρονία μέσα στη συγχρονία»
- η γλωσσική ποικιλία αποτελεί προϋπόθεση της γλωσσικής αλλαγής

«Διαχρονία μέσα στη συγχρονία» I

κλασ(σ)ικός

- ταξινομική λειτουργία

Συνάψεις: αθλητισμός, μουσική, περίοδος, αρχαιότητα, κείμενα, σπουδές, πολιτισμός, ρεπερτόριο

- αξιολογική λειτουργία

Συνάψεις: παράδειγμα, περίπτωση, ευκαιρία

ο οποίος με τη σειρά του χάνει **κλασική** ευκαιρία σουτάροντας τωρινή ύφεση εκδηλώνεται ως μια **κλασική** κρίση υπερπαραγωγής ησαν, ε: τι – // τι σας είπαν; **κλασική** περίπτωση φοβίας, και περίπτωση φοβίας, και άγχους – **κλασικότητα** (.) δε φοβάται ου άνοιξε και θεωρείται από τα **κλασικότερα** ροκάδικα, στην κα

«Διαχρονία μέσα στη συγχρονία» II

άγριος

- ταξινομική λειτουργία

Συνάψεις: ζώα, ζωή, πανίδα, φύση, πουλιά, άλογα

- περιγραφική λειτουργία

Συνάψεις: διαθέσεις, ομορφιά, φωνή, επιφάνεια, κύματα

- περιγραφική-επιτατική λειτουργία

Συνάψεις: εκμετάλλευση, συγκρούσεις

- επιτατική λειτουργία

Συνάψεις: μεσάνυχτα, χαράματα

Πρόσφατη γλωσσική αλλαγή II

- ιστορική κοινωνιογλωσσολογία (Romaine 1982, Hernandez-Campoy & Conde-Silvestre 2012)
- ιστορική πραγματολογία (Jucker 1995, Jucker & Taavitsainen 2010)
- γραμματικοποίηση (Hopper & Traugott 1993/2003)
- γλωσσολογία σωμάτων κειμένων: χρήση συγκρίσιμων σωμάτων κειμένων, που επιτρέπουν τη συστηματική ανάλυση διαχρονικών δεδομένων από αυθεντικά (γραπτά και προφορικά) κείμενα

Πρόσφατη γλωσσική αλλαγή III

- *Διαχρονικά σώματα κειμένων*: καλύπτουν μία ή περισσότερες ιστορικές περιόδους ή ολόκληρη την ιστορική εξέλιξη της υπό μελέτη γλώσσας
 - Corpus of Historical American English (COHA, 1810-2009)
- *«Οικογένειες» σωμάτων κειμένων*: σώματα που έχουν σχεδιαστεί με τα ίδια κριτήρια, αλλά αφορούν διαφορετικές χρονικές περιόδους
 - BLOB-1931 (1928-1934), LOB (1961) και F-LOB (1991): 1 εκατ. λέξεις, ανά 30 χρόνια μέσα στον 20ό αιώνα
- *Συγχρονικά σώματα κειμένων*
 - BNC (δεκαετία του 1960 έως σήμερα)

Πρόσφατη γλωσσική αλλαγή IV

➤ Αγγλικά

- Γεωγραφική ποικιλία: βρετανικά, αμερικανικά, αυστραλιανά, ινδικά αγγλικά κ.ά. και διάλεκτοι των βρετανικών αγγλικών (π.χ. βόρεια αγγλικά, σκοτσέζικα)
- Όλες οι ιστορικές περίοδοι της αγγλικής: παλαιά, μέση, πρώιμη σύγχρονη, μεταγενέστερη σύγχρονη έως σήμερα
- Ποικιλία κειμενικών ειδών (π.χ. COHA) ή συγκεκριμένα κειμενικά είδη (π.χ. TIME Magazine Corpus, 1923-2006, The Diachronic Corpus of Present-Day Spoken English, DCPSE, 1950-1990)

Πρόσφατη γλωσσική αλλαγή V

- Ουαλικά: Historical Corpus of Welsh Language, 1500-1850
- Ισλανδικά: Icelandic Parsed Historical Corpus, 12ος-20ός αι.
- Γερμανικά: Deutsch Diachron Digital, 750-1900
- Τσεχικά: Diachronic section of the Czech National Corpus, 13ος αι.-1989
- Γαλλικά: Corpus MCVF, 1100-1837
- Ιταλικά: Diachronic Corpus di Italiano Scritto, 1861-1945
- Ισπανικά: Corpus Diacrónico del Español, πριν από το 1290-1974
- Πορτογαλικά: O corpus do Português, 1300-1900
- Κινεζικά: Sheffield Corpus of Chinese, 960-1279, 1366-1644, 1644-1911

Πρόσφατη γλωσσική αλλαγή VI

➤ Πολύ μεγάλα σώματα κειμένων:

- Corpus of Historical American English (1810-2009): 400 εκατ.
- Corpus Diacrónico del Español (1290-1974): 125 εκατ.

➤ Μεγάλα σώματα κειμένων:

- O corpus do Português (1300-1900): 45 εκατ.
- Corpus of English Novels (1881-1922): 26 εκατ.
- The Old Bailey Corpus (1720-1913): 14 εκατ.

➤ Μεσαία σώματα κειμένων:

- Diachronic Czech National Corpus (13ος αι.-1989): 2 εκατ.
- The Helsinki Corpus of English Texts (730-1710): 1,5 εκατ.
- BLOB-1931 (1928-1934): 1 εκατ.

➤ Μικρά σώματα κειμένων:

- Parsed Corpus of Old English (-1300): 106.210
- Sheffield Corpus of Chinese (960-1911): 18.000 (Πρβλ. Kytö 2012: 1516)

Πρόσφατη γλωσσική αλλαγή VII

- Γλωσσικά φαινόμενα που έχουν μελετηθεί:
 - λεξιλόγιο (λ.χ. Baker 2011)
 - γραμματική (Rudanko 2006: συμπληρώματα, Leech κ.ά. 2009)
 - μορφολογία (λ.χ. Baayen & Renouf 1996: παραγωγικότητα γραμματικών μορφημάτων, Duguid 2010)
 - φρασεολογία (λ.χ. Davies 2012)
 - πολιτισμικές αλλαγές (λ.χ. Baker 2010: φύλο, Marchi 2010: ηθική, Partington 2012: αντισημιτισμός)

Πρόσφατη γλωσσική αλλαγή VIII

- Αλλαγές στη γραμματική:
τροπικότητα (βλ. λ.χ. Krug 2000, Millar 2009)
 - Leech (1992): ρήματα που δηλώνουν τροπικότητα (*can, could, should, must* κ.λπ.) μειώνονται σε συχνότητα κατά την περίοδο 1961-1991, ενώ ημι-τροπικά ρήματα (*need to, have to*) αυξάνονται σε συχνότητα

Πρόσφατη γλωσσική αλλαγή IX

- Αλλαγές στη συχνότητα του λεξιλογίου

Baker (2011): Οικογένεια σωμάτων κειμένων Brown και BE06 (1931-1961-1991-2006):

- Μεγαλύτερη μείωση συχνότητας: *upon, shall, Mr, certain, Mrs, great, sir, must, whole, present*
- Μεγαλύτερη αύξηση συχνότητας: *around, health, information, it's, didn't, says, social, family, children, need*
- αλλαγή στη χρήση
- πολιτισμικές αλλαγές

Διαχρονικό Σώμα Ελληνικών Κειμένων 20ού αι.

Πρόγραμμα «Αριστεία»

Στόχοι



1. εξέταση των ζητημάτων δημιουργίας ενός διαχρονικού σώματος κειμένων στα ελληνικά
2. συλλογή δεδομένων με στόχο τη δημιουργία
 - ενός πιλοτικού σώματος κειμένων
 - του Διαχρονικού Σώματος Ελληνικών Κειμένων του 20ού αι.
 - > 20 εκατ. λέξεις (μεγάλο διαχρονικό ΣΚ)
 - > συγκρίσιμο με το ΣΕΚ
3. ανάλυση του σώματος κειμένων για τη διερεύνηση της γλωσσικής αλλαγής στη διάρκεια των δεκαετιών του 20ού αιώνα.

Ιδιαιτερότητες: Η Ελλάδα του 20ού αιώνα



Ιδιαιτερότητες: Τα ελληνικά του 20ού αιώνα I

	<i>Γεγονότα στην Ελλάδα</i>	<i>Γλωσσικός σχεδιασμός και στάσεις</i>
1900		1901/1902: Ευαγγελικά, Ορεστικά 1905: Εταιρεία «Η Εθνική Γλώσσα»
1910	1912-1913: Βαλκανικοί πόλεμοι 1914: Α΄ Παγκόσμιος Πόλεμος, εθνικός διχασμός	1910: Εκπαιδευτικός Όμιλος 1911: Άρθρο 107: καθαρεύουσα επίσημη γλώσσα 1917: Εισαγωγή δημοτικής στο Δημοτικό
1920	1920: Ήττα Βενιζέλου στις εκλογές 1922: Μικρασιατική καταστροφή 1925: Δικτατορία Πάγκαλου 1928: Κυβέρνηση Βενιζέλου	1920: Αλλαγή κυβερνητικής πολιτικής 1923: Επαναφορά δημοτικής 1927: Το ΚΚΕ εγκαταλείπει την καθαρεύουσα
1930	1936: Δικτατορία Μεταξά	1933: Πρώτος τόμος <i>Ιστορικού Λεξικού</i> , Λεξικά Δημητράκου, Πρωίας
1940	1940: Β΄ Παγκόσμιος Πόλεμος 1944-1949: Εμφύλιος	1941: <i>Γραμματική</i> Τριανταφυλλίδη, «Δίκη των τόνων» 1946: <i>Νεοελληνική Σύνταξις</i> Τζάρτζανου
1950	Μεταπολεμική ανασυγκρότηση: δεξιές κυβερνήσεις	1955: Μετάφραση <i>Ιλιάδας</i> Καζαντζάκη-Κακριδή
1960	1964: Κυβέρνηση Παπανδρέου 1967: Δικτατορία	1964: Εκπαιδευτική μεταρρύθμιση
1970	1974: Αποκατάσταση δημοκρατίας	1976: Δημοτική επίσημη γλώσσα του κράτους
1980	1981: Είσοδος στην ΕΟΚ	1982: Εισαγωγή μονοτονικού 1986: Σύνταγμα στη δημοτική

Ιδιαιτερότητες: Τα ελληνικά του 20ού αιώνα II

- η γλωσσική πρακτική απέκλινε από το γλωσσικό σχεδιασμό και τις απόλυτες διχοτομήσεις

Πρώτο μισό 20ού αι.: Mirambel (1937)

- *καθαρεύουσα*
- «μεικτή» (επιστημονικά, πολιτικά κείμενα, ειδήσεις)
- «καθομιλουμένη» (καθημερινή πόλεων, δημοσιογραφία, ενημερωτικά κείμενα)
- *δημοτική* (καθημερινή χωριών, με διαλεκτική ποικιλία)
- «μαλλιαρή»: προτυποποίηση δημοτικιστών

Αρχές σχεδιασμού I

Επιλογή υλικού

1. Χρονική περίοδος: βασική μονάδα η δεκαετία
> 2.200.000 λέξεις για τις 9 πρώτες δεκαετίες του 20ού αι.
2. Δειγματοληψία: αποφυγή συστηματικής προκατάληψης με
τυχαία επιλογή των μεταβλητών
> ανώτατο όριο λέξεων ανά κείμενο, συγγραφέων,
δημοφιλών κειμένων
3. Εκδόσεις: πρώτη έκδοση

Αρχές σχεδιασμού II

Επιλογή υλικού

4. Κοινωνιογλωσσικές μεταβλητές: προσπάθεια για τυχαία δειγματοληψία με μεγάλο εύρος

5. Γλωσσική ποικιλία:

το κειμενικό είδος ως βασική μονάδα σύγκρισης

το ερώτημα της διγλωσσίας ως ερευνητική υπόθεση που θα διερευνηθεί στο στάδιο της ανάλυσης και όχι ως αρχή συλλογής δεδομένων

Αρχές σχεδιασμού III

Διαθεσιμότητα και πηγές δεδομένων

- προβλήματα με προφορικά δεδομένα
- κενά στην αντιπροσώπευση κειμενικών ειδών
- υποχώρηση παλιών/ανάδυση νέων κειμενικών ειδών
(πρβλ. Kytö 2012: 1521-1523)

Στρατηγικές:

- > έμφαση σε κειμενικά είδη που σχετίζονται με τη συζήτηση (speech-like/-based/-purposed)
- > ανάπτυξη «μικρο-σωμάτων κειμένων»
- > αναμόρφωση κειμενικών ειδών ΣΕΚ με στόχο τη συγκρισιμότητα

Αρχές σχεδιασμού IV

Επεξεργασία, επισημείωση και οπτικοποίηση κειμένων

- ανάπτυξη πολυτονικού OCR
- συσχέτιση μονοτονικών και πολυτονικών παραλλαγών
- συσχέτιση μορφολογικών παραλλαγών (π.χ. *έργαστήριον-εργαστήριο-εργαστήρι*)
- επισημείωση για μέρη του λόγου, λημματοποίηση
- διαθεσιμότητα σε χρηστική ιστοσελίδα

Έρευνα: Δεδομένα και μεθοδολογία I

Επιλογή δεδομένων από το ΣΕΚ20:

Είδη	Κειμενικοί τύποι	Αριθμός λέξεων
Προφορικά	Κινηματογραφικά επίκαιρα	78,441
Δημόσιες ομιλίες	Βουλή, Ακαδημαϊκά, Άλλα	339.194
Συνομιλία	Διάλογοι ταινιών	208.206
Λογοτεχνία	Μυθιστορήματα, Διηγήματα, Ποίηση, Θεατρικά	1.086.687
Ακαδημαϊκά	Ανθρωπιστικά, Κοινωνικά/ Οικονομικά, Θετικές επιστήμες	1.044.200
Εκλαϊκευτικά		285.252
Ιδιωτικά	Επιστολές	178.720
ΣΥΝΟΛΟ		3.220.700

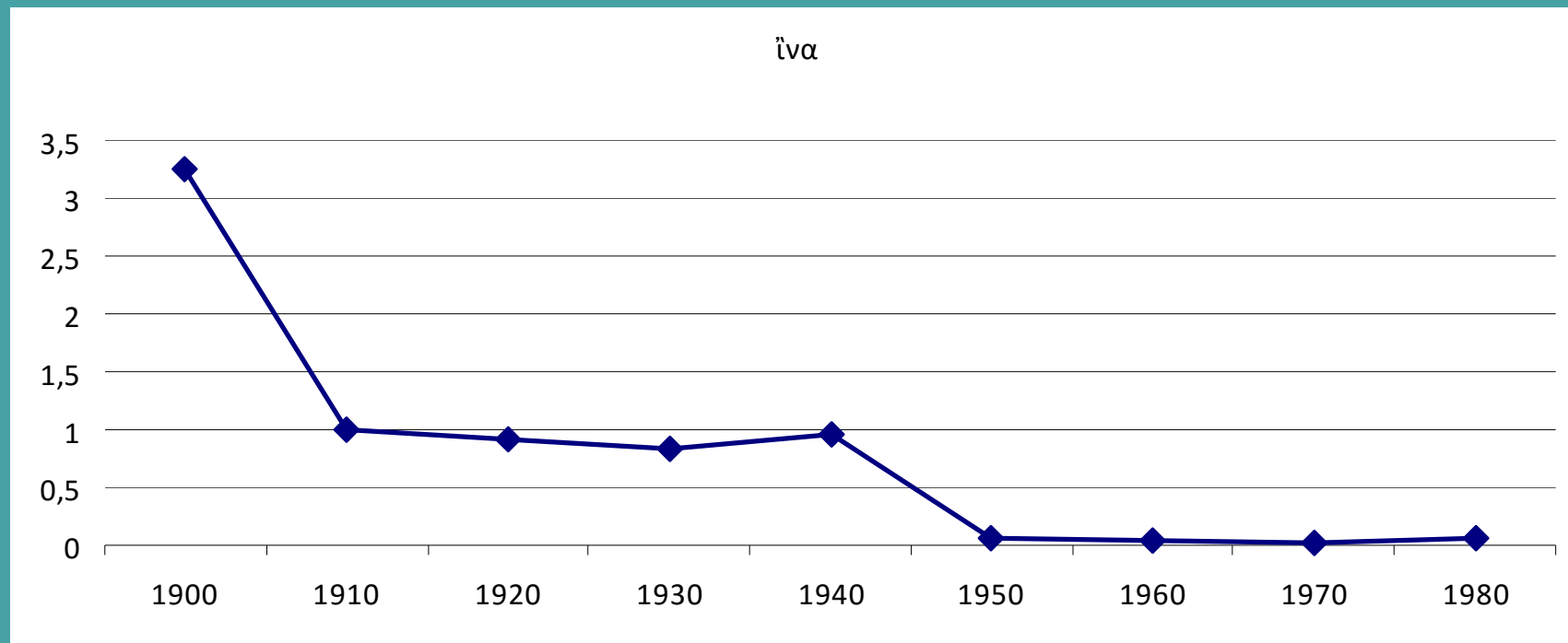
Έρευνα: Δεδομένα και μεθοδολογία II

Συχνότητα απλή και κανονικοποιημένη (σε 10.000 λέξεις)
βασικών γραμματικών στοιχείων: ἵνα και ζεύγη: ποῖος -
ποιός/ποιος, διά - γιά/για, εἰς - σέ/σε, (και: δι', γι', σ', στ*)

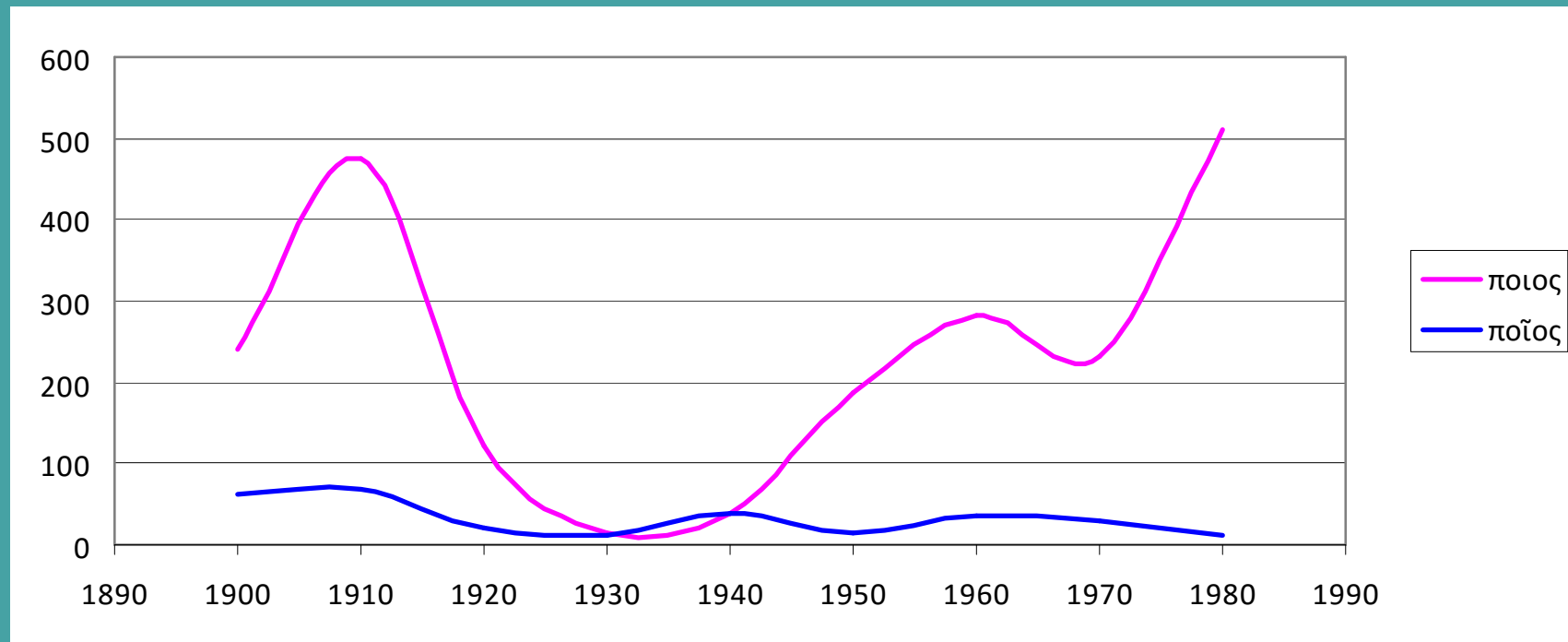
Σε 7 κειμενικά είδη, με έμφαση σε είδη που προσφέρουν
στοιχεία για την προφορική γλώσσα (Culpeper & Kytö 2010:
speech-like π.χ. ιδιωτικές επιστολές, *speech-based* π.χ.
δημόσιες ομιλίες και *speech-purposed* π.χ. διάλογοι ταινιών,
θεατρικά)

Δεδομένα από 9 δεκαετίες (1900-1989)

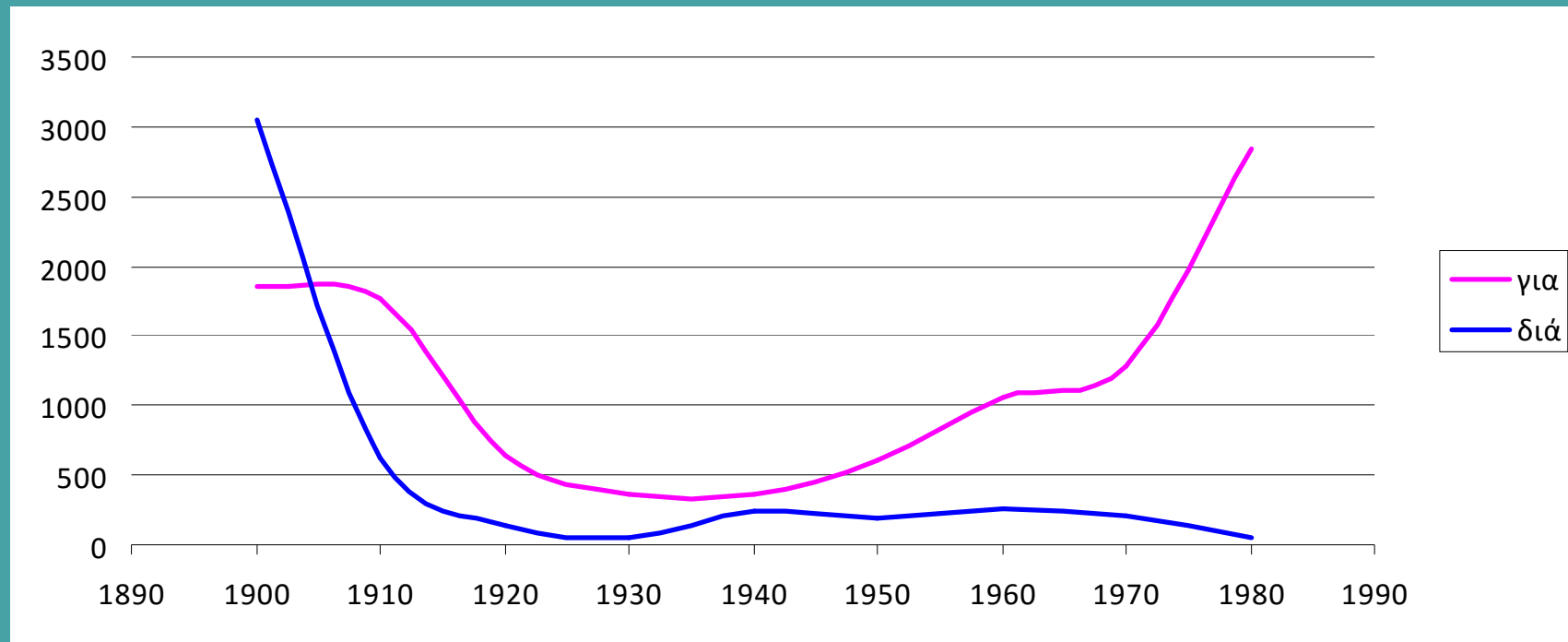
Γραμματικοί τύποι στη διαχρονία I



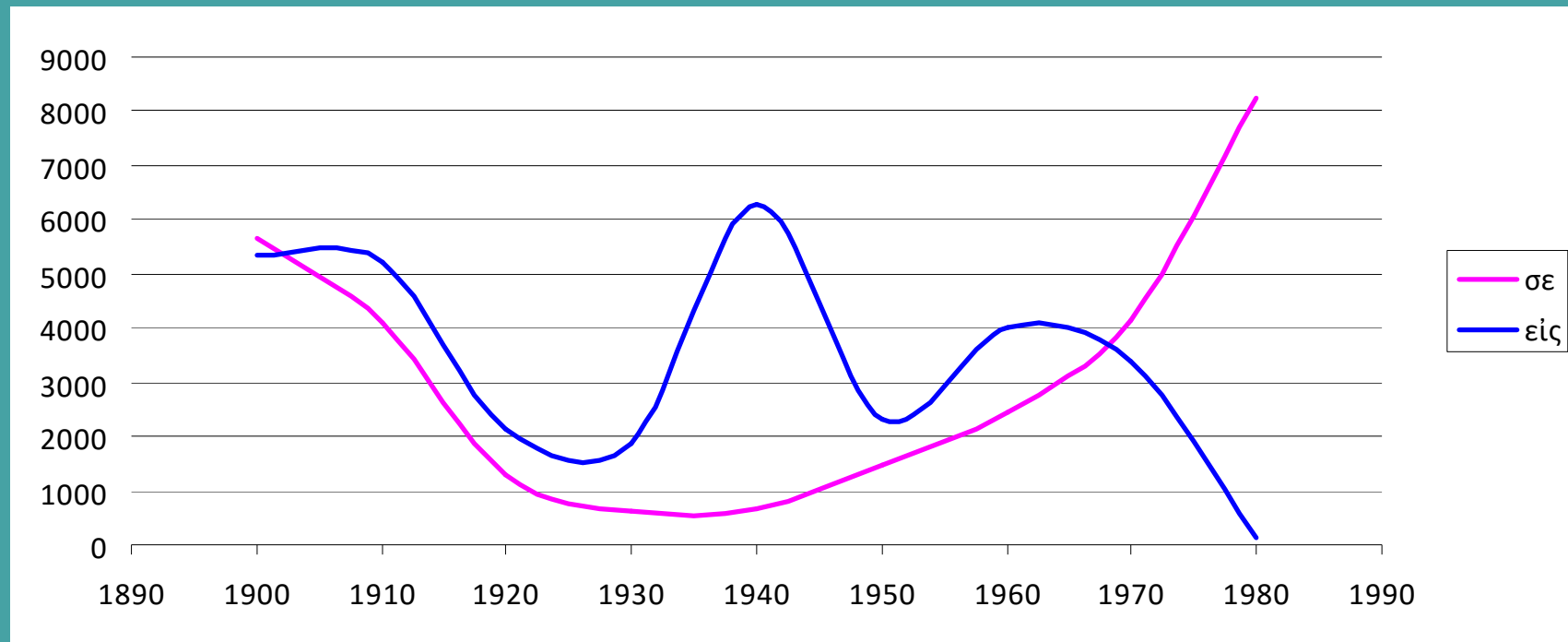
Γραμματικοί τύποι στη διαχρονία II



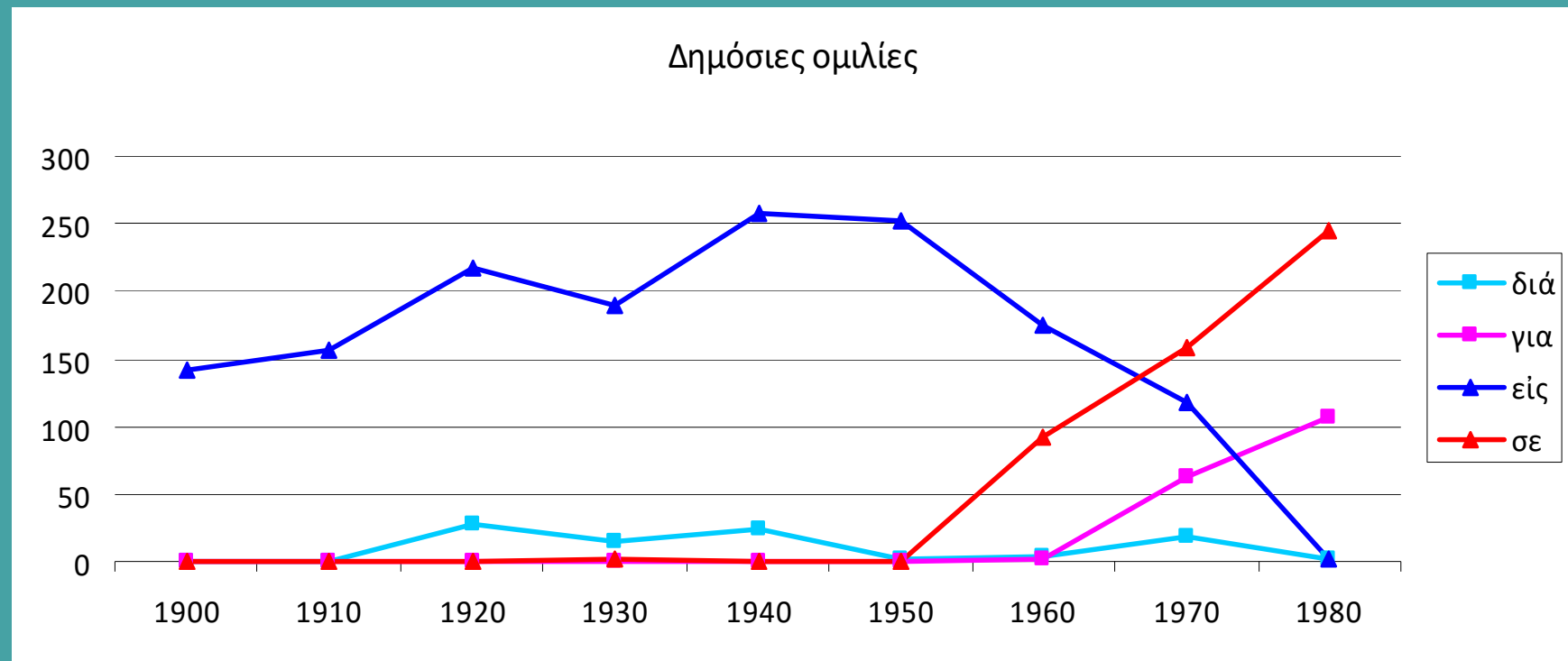
Γραμματικοί τύποι στη διαχρονία III



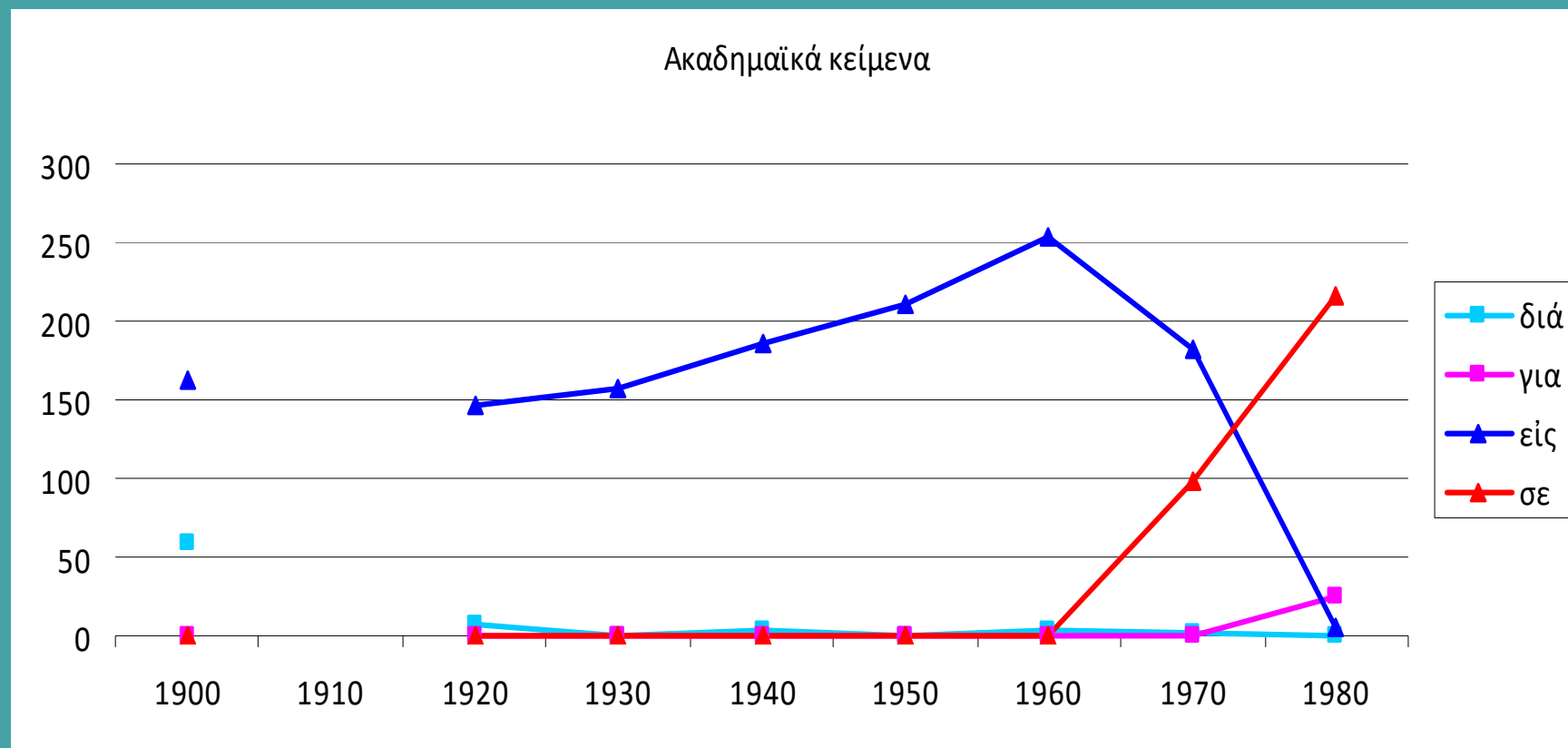
Γραμματικοί τύποι στη διαχρονία IV



Κειμενικά είδη I

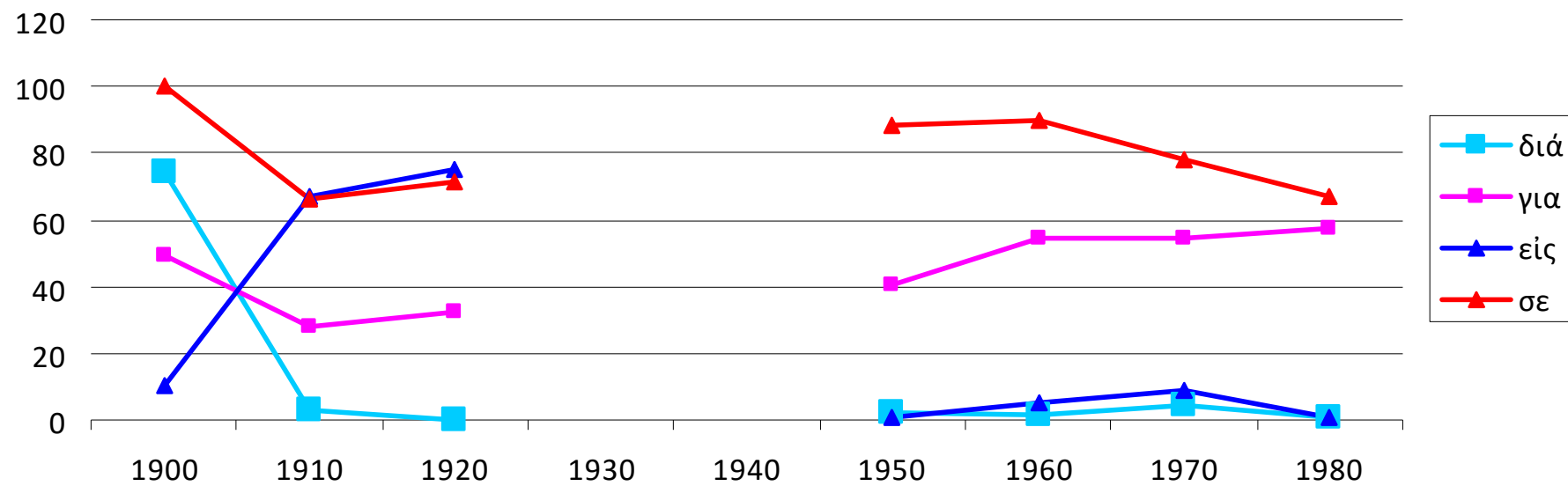


Κειμενικά είδη II

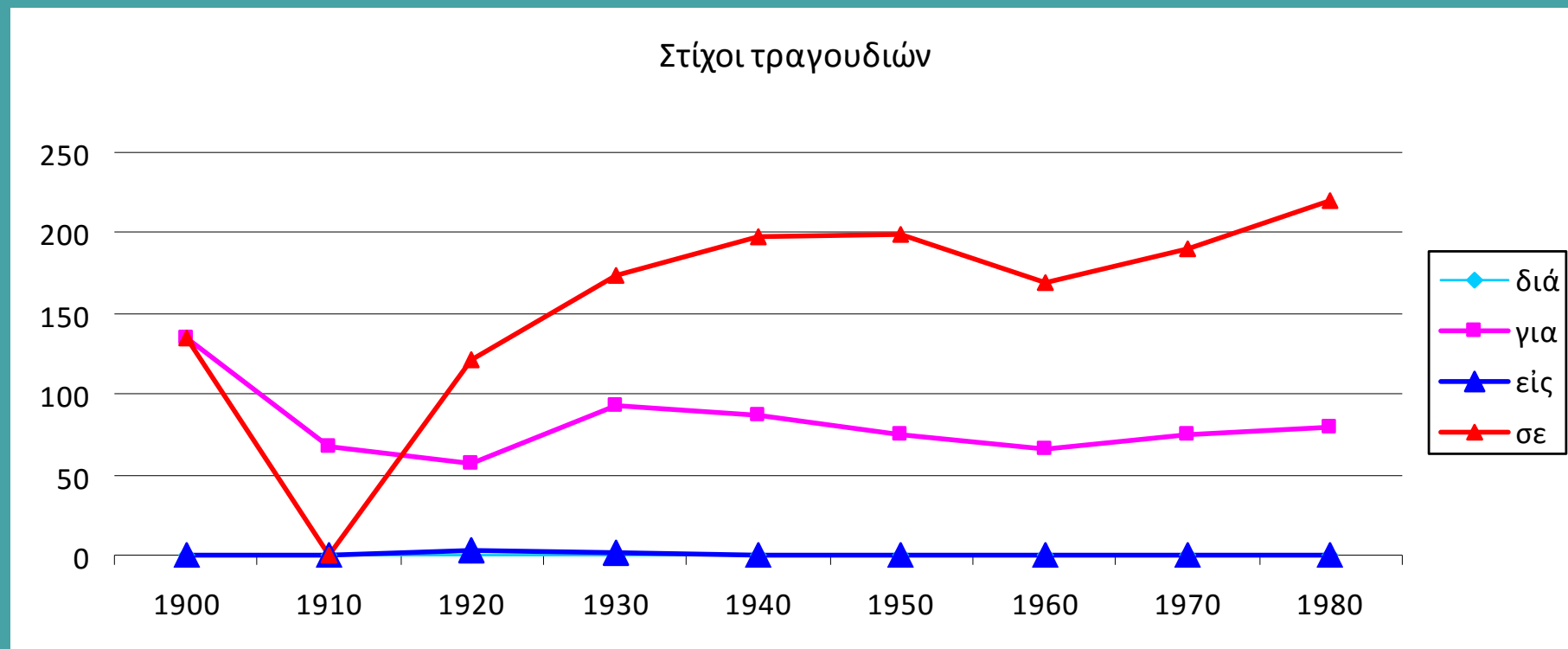


Κειμενικά είδη III

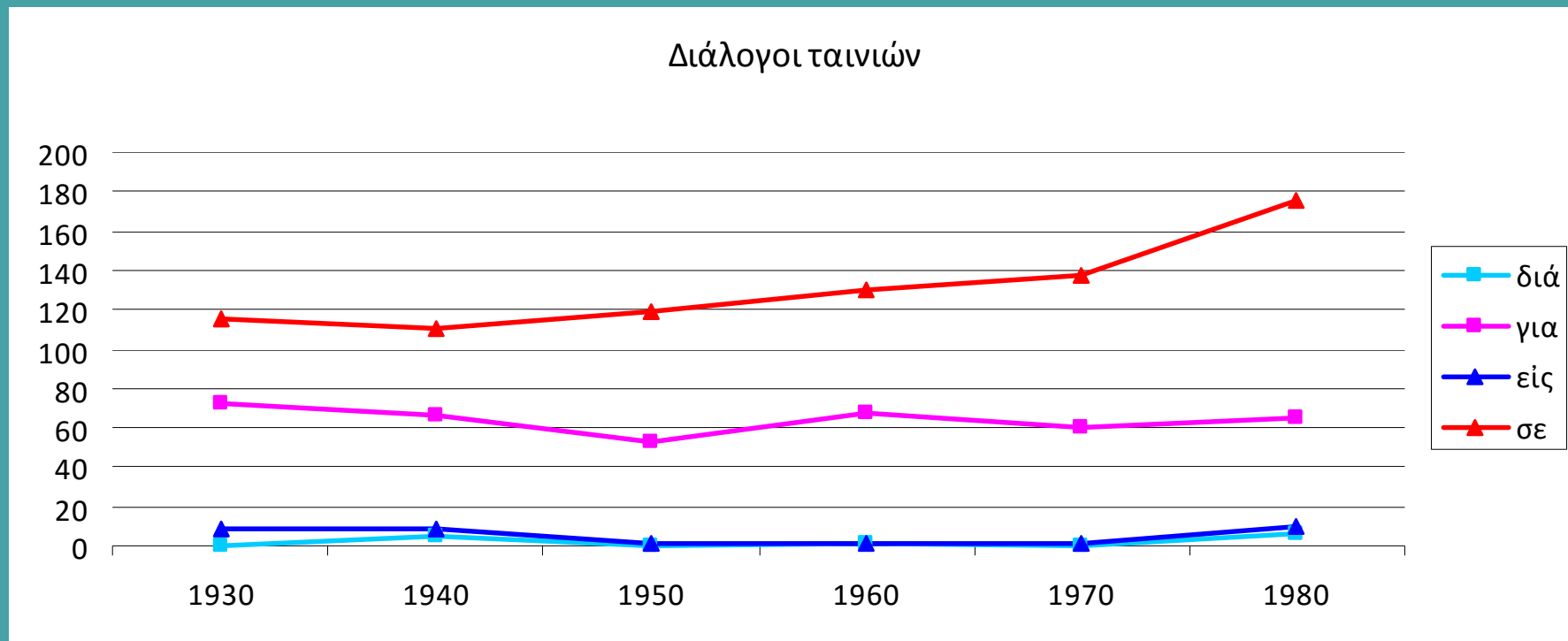
Λογοτεχνία



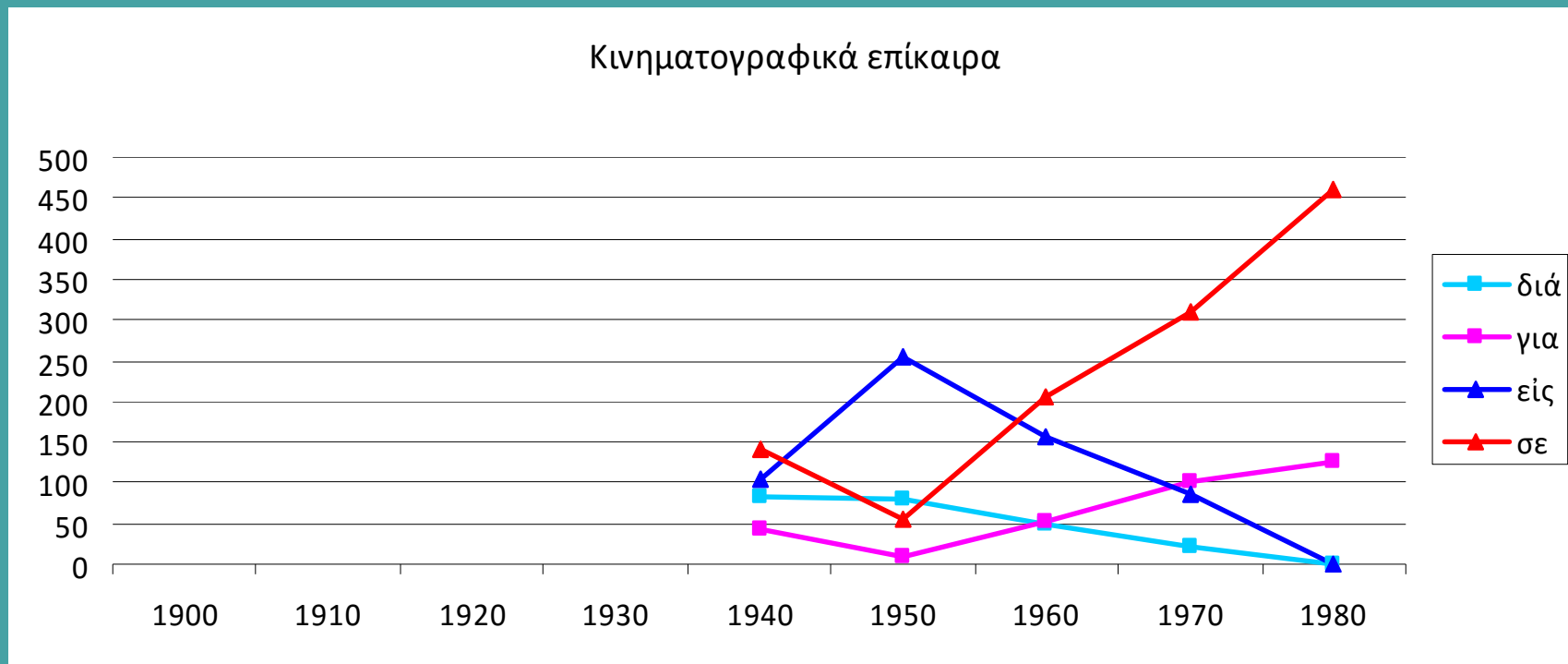
Κειμενικά είδη IV



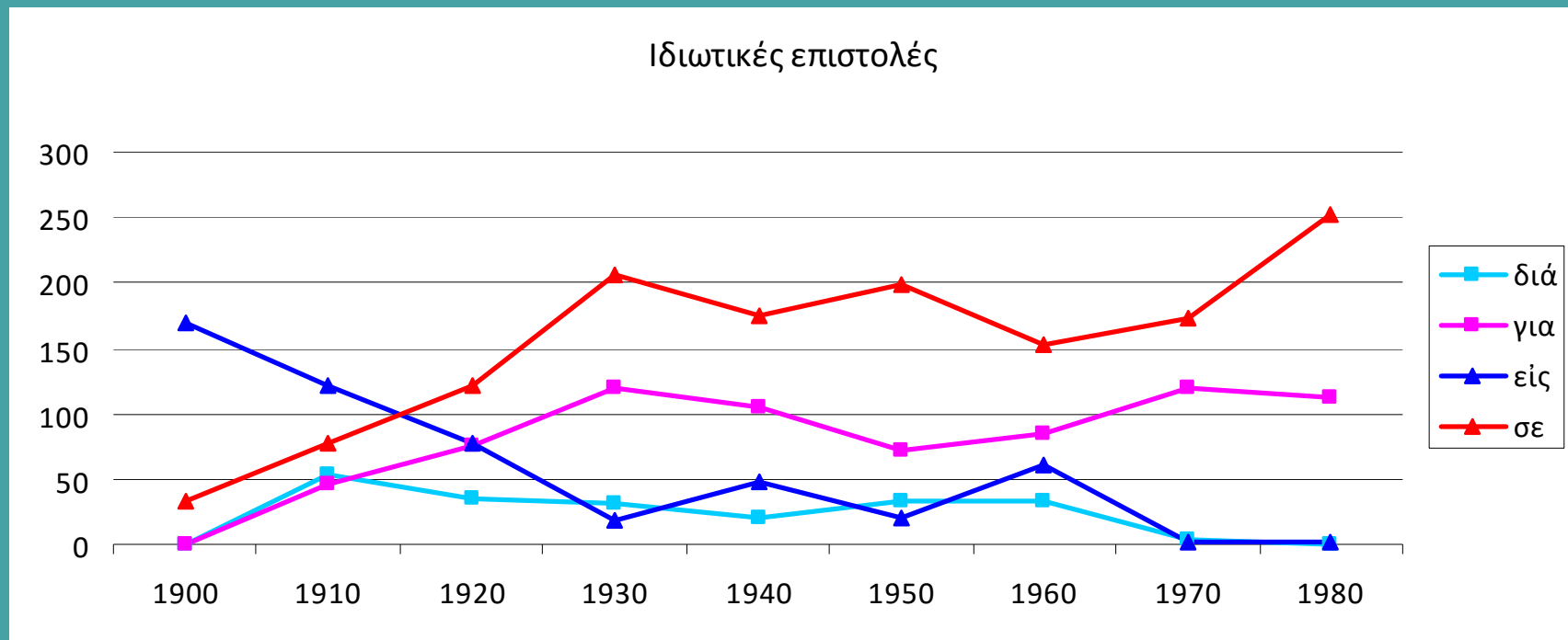
Κειμενικά είδη V



Κειμενικά είδη VI



Κειμενικά είδη VII



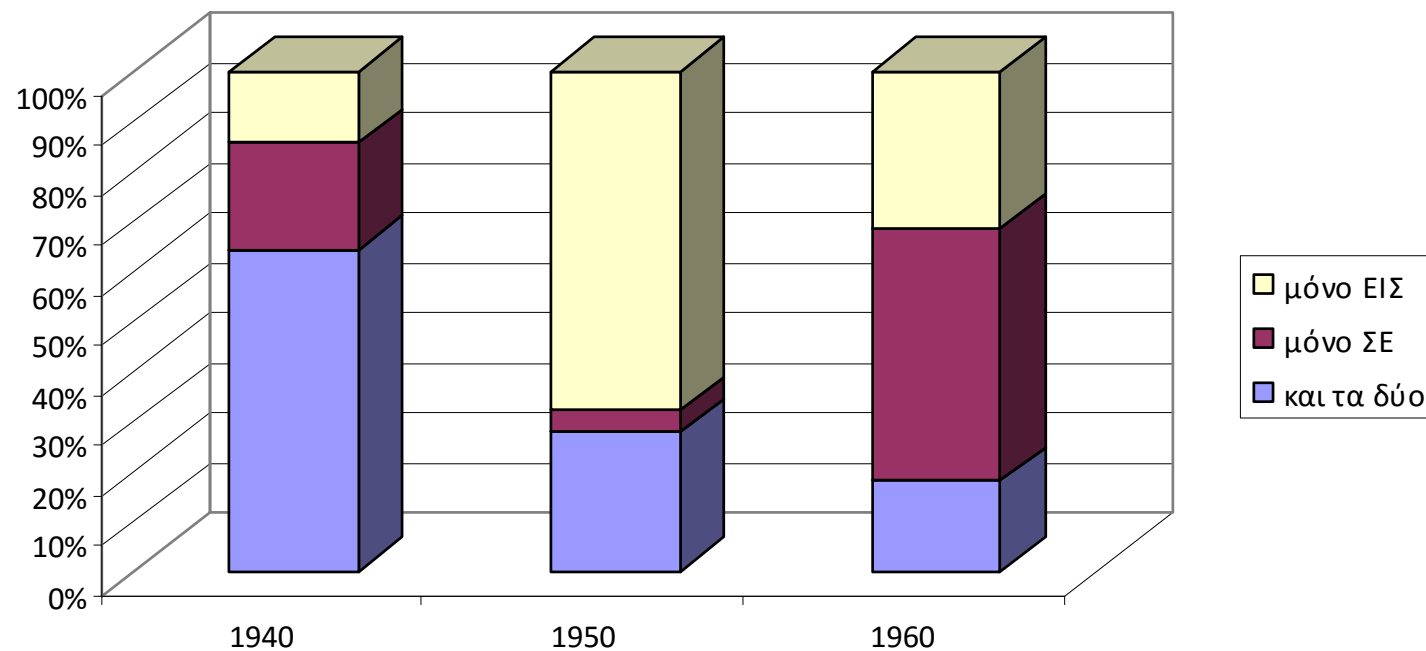
Συμπεράσματα I

- > Σχήμα U των χαμηλών εναλλακτικών τύπων στην ελληνική διγλωσσία
- > Ακανόνιστο σχήμα των υψηλών εναλλακτικών τύπων > στερεοτυπική (με την έννοια του Labov), εμβληματική χρήση
- > Κεντρικός ρόλος του κειμενικού είδους για την κατανόηση της διγλωσσίας και της γλωσσικής αλλαγής, γενικότερα.
- > Διάκριση μεταξύ επίσημων και ανεπίσημων ειδών: συντηρητικά με ξαφνική αλλαγή σε αντίθεση με τη σταδιακή, «φυσιολογική» γλωσσική αλλαγή

Συμπεράσματα II

Παραλλαγές που συνυπάρχουν σε κειμενικά είδη ή μεμονωμένα κείμενα

- εις-σε σε κινηματογραφικά επίκαιρα



Συμπεράσματα III

- εις - σε σε κινηματογραφικά επίκαιρα

Αι μηχανοκίνηται δυνάμεις του ελληνικού στρατού παρελαύνουν εις επιβλητικούς σχηματισμούς και εμπνέουν εις τας μυριάδας του λαού που τις παρακολουθεί το αίσθημα της ασφαλείας και της υπερηφανείας. [...] Χιλιάδες των θεατών χειροκροτούν και επευφημούν τους άνδρας αυτούς, οι οποίοι, με τους υπερόχους αγώνας των, έδωσαν στην Ελλάδα μας την μεγάλην και ωραίαν της νίκην (SRF01-1940-0002)

Μετά την άφιξιν εις την πόλιν, τα παιδιά δέχονται την ευλογίαν της Εκκλησίας. Προσφέρονται εις τα παιδιά διάφορα δώρα. Τα παιδάκια φθάνουν εις παιδόπολιν της Θεσσαλονίκης. Με τη χαρά ζωγραφισμένα στα πρόσωπά των παίζουν με τα δώρα που τους προσεφέρθησαν [...]. (SRF01-1950-0019)

Στον αερολιμένα του Ελληνικού συνεχίζεται με εντατικότατο ρυθμό στα πλαίσια του Τεχνικού Προγράμματος του 1969 η ανέγερσις νέων κτηριακών εγκαταστάσεων [...] Στο Τεχνικό Πρόγραμμα του 1969 προβλέπεται [...] δια της χρησιμοποιήσεως ραντάρ εις τον έλεγχον της εναερίου κυκλοφορίας. [...] να ανταποκρίνεται άνετως εις τας υποχρεώσεις του ως διεθνούς αεροπορικού κόμβου και εις τας ανάγκας διακινήσεως εμπορευμάτων και επιβατών (SRF01-1960-0050)

Συμπεράσματα IV

- εις - σε σε ιδιωτικές επιστολές 1900-1920:

μόνο 4% και με τους δύο τύπους > προσωπική προτίμηση

- διά - για στη λογοτεχνία 1900-1910:

τα περισσότερα κείμενα έχουν και τους δύο τύπους>

υφολογικός μηχανισμός:

Μία εξ αυτών κατελήφθη υπό σπασμωδικού ακαταβλήτου γέλωτος και ενώ δια της μιας χειρός εκράτει τα στήθη της, δια της άλλης εκτύπα τον ώμον της Λενιάς [...] συζητούντας και λύοντας τας διαφοράς των ενώπιον των προεστών, "για να παίρνη πράξι" [...] Ήτο τότε πολύ μικρός· επειδή δε κάθε φορά εζήτει να τον σηκώνουν για να παίξη την καμπάνα, αυτός κατ' ουδένα λόγον να πλησιάση άλλους ανθρώπους. Ούτε για να μεταλάβη δεν έστεργε να καταιβή στο χωριό. [...] Ο Μανώλης εγέλασεν, αλλ' εγέλασε μάλλον δια την μωρίαν εκείνου του νέου (WFB08-1910-0004)

Συμπεράσματα V

Βασικά στοιχεία για τα αναπάντητα ερωτήματα της
διγλωσσίας:

- αντίθεση προφορικού-γραφτού λόγου;
 - πότε συνέβησαν οι αλλαγές;
 - πώς επηρέασαν οι γλωσσικές στάσεις τη γλωσσική χρήση;
-
- > ρόλος των κειμενικών ειδών και γλωσσικών ποικιλιών
 - > ολοκληρωμένη αντίληψη για την τυποποίηση της Ελληνικής

Άλλα ερευνητικά ερωτήματα

- αλλαγές στην έκφραση της τροπικότητας
- νόρμες για την τοποθέτηση των συνδετικών στοιχείων
- αναφορικές αντωνυμίες: *ο οποίος-που*
- πτώσεις-προθετικές φράσεις
- ποικιλία στους τύπους της γενικής (-ης ή -εως)
- μορφολογική παραγωγικότητα
- σημασιολογική αλλαγή: αξιολόγηση
- τύποι που σχετίζονται με τη δήλωση του φύλου
- νόρμες των κειμενικών ειδών
- κειμενικό ύφος/συνομιλιοποίηση/επικράτηση καθομιλουμένης



www.greekcorpus20.phil.uoa.gr