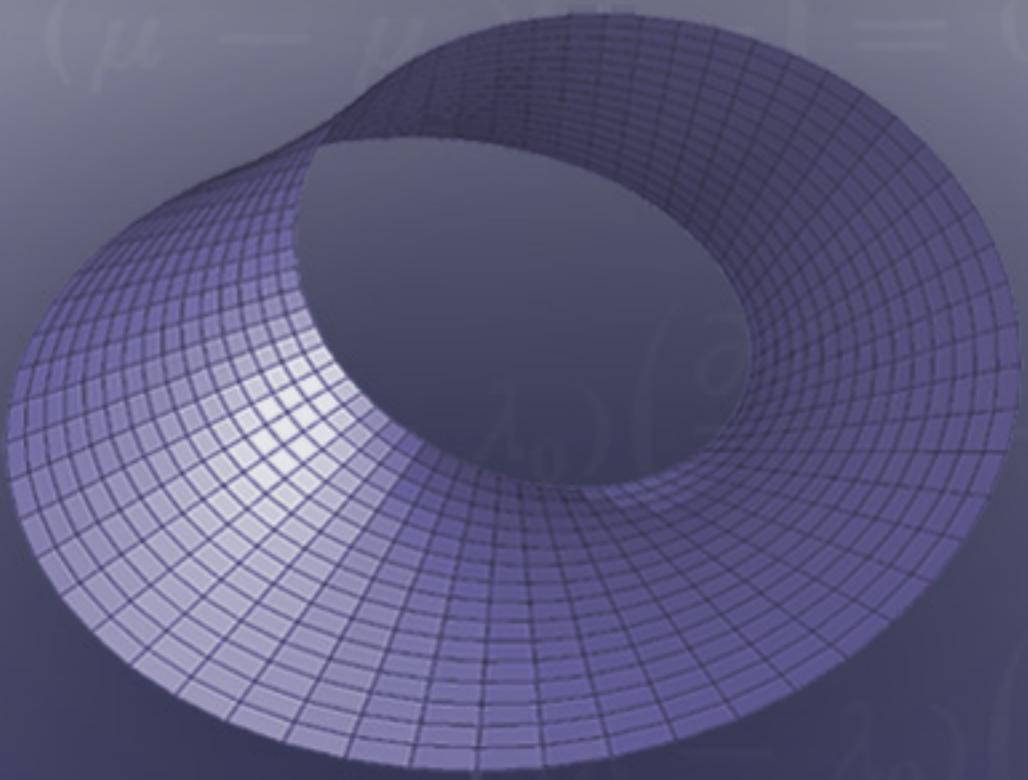


# Οι νέες τεχνολογίες στην κοινωνική έρευνα



## ΕΠΙΜΕΤΡΟ

# ΟΙ ΝΕΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΣΤΗΝ ΚΟΙΝΩΝΙΚΗ ΕΡΕΥΝΑ

### 1. Εισαγωγή

Εύκολα μπορεί κανείς να παρατηρήσει ότι τα τελευταία είκοσι χρόνια —που μας χωρίζουν από την πρώτη έκδοση του παρόντος έργου— η έρευνα στις Κοινωνικές Επιστήμες άλλαξε αρκετά μορφή λόγω της αυξημένης χρήσης των Νέων Τεχνολογιών (NT).

Πράγματι, ο Ηλεκτρονικός Υπολογιστής (H/Y), που ήδη εκείνα τα χρόνια χρησιμοποιείτο σαν εργαλείο επεξεργασίας και ανάλυσης ποσοτικών κυρίως δεδομένων, έχει σήμερα μετατραπεί σ' ένα «ψηφιακό μέσο» του οποίου η χρήση επεκτείνεται, για να συμπεριλάβει, μεταξύ άλλων, τη συλλογή, διακίνηση και αποθήκευση των δεδομένων, τη διενέργεια πολύπλοκων αναλύσεων και την παρουσίαση αποτελεσμάτων, την εξασφάλιση πρόσβασης του ερευνητή σε διεθνή δίκτυα πληροφοριών και τη δυνατότητα επικοινωνίας του με την επιστημονική κοινότητα.

Είναι φανερό ότι μια εισαγωγική κατάρτιση στις επιμέρους εφαρμογές των NT στην κοινωνική έρευνα απαιτεί ένα συνδυασμό θεωρητικής προσέγγισης αλλά και πρακτικών εφαρμογών με χρήση προϊόντων των NT. Καθώς όμως αυτό είναι αδύνατο στο πλαίσιο της συγκεκριμένης έκδοσης, θα περιοριστούμε σε μια όσο το δυνατόν περισσότερο προσιτή και επιγραμματική εισαγωγική παρουσίαση των ευρύτερων και ήδη δοκιμασμένων τομέων όπου χρησιμοποιούνται οι NT στην κοινωνική έρευνα.

Η χρήση των NT παρέχει σημαντικές τεχνικές διευκολύνσεις στον ερευνητή της Κοινωνιολογίας, όπως:

i. άμεσο έλεγχο στη γραφή, την επανάληψη και την παραγωγή κάθε

γραπτού του (χειρόγραφα ερευνητικής κατεύθυνσης, πίνακες αποτελεσμάτων, γραφικές παραστάσεις, διδακτικό υλικό, κλπ.).

ii. ηλεκτρονική πρόσβαση σε ευρύτατα σύνολα πληροφορίας είτε μέσω δικτύων (on-line data bases) είτε μέσω οπτικών δίσκων (off-line CD-ROM).

iii. ταχεία, αποτελεσματική και εύληπτη συλλογή, διακίνηση, αποθήκευση και ανάκληση στοιχείων-δεδομένων ερευνών και σχετικής πληροφορίας.

Επίσης οι NT δίνουν νέες δυνατότητες για:

iv. αμφιδρομες ποσοτικές και ποιοτικές αναλύσεις δεδομένων με επιλογή χρήσης μεγάλης ποικιλίας μεθόδων

v. ταχύτατη και απρόσκοπτη επικοινωνία με μεμονωμένους ερευνητές ή ερευνητικές ομάδες ανά τον κόσμο, μέσω διεθνών δικτύων τύπου INTERNET, JANET, κ.ά.·

vi. αποτελεσματικότερη, από πλευράς κατανόησης και ερμηνείας, παρουσίαση του υλικού των δεδομένων και των αποτελεσμάτων της κοινωνιολογικής έρευνας, αφού με τη χρήση του υπερκειμένου (hypertext), των πολυμέσων (multimedia), των γραφικών με υπολογιστή (computer graphics) και των συστημάτων γεωγραφικής πληροφορίας (Geographical Information Systems - GIS) μπορούν πλέον να χρησιμοποιηθούν μη γραμμικά κείμενα, ήχοι, εικόνες, κινούμενο σχέδιο, βίντεο, γραφικά, καθώς και χάρτες, φωτογραφήσεις από δορυφόρους κλπ.

Ακόμη,

vii. οι αρχές, η μεθοδολογία και οι εφαρμογές επιστημονικών περιοχών, συνδεδεμένων άμεσα με τις NT, όπως αυτές της τεχνητής νοημοσύνης (artificial intelligence), των έμπειρων συστημάτων (expert systems) και της προσομοίωσης (simulation), χρησιμοποιούνται για την ανάπτυξη κοινωνιολογικής θεωρίας, για την ερμηνεία κοινωνικών φαινομένων, για την κατασκευή πολύπλοκων μοντέλων, αλλά και, συμβουλευτικά, σε διάφορα στάδια της έρευνας.

Στο πλαίσιο της παρούσας εισαγωγής θα περιοριστούμε στην παρουσίαση τριών τομέων της κοινωνικής έρευνας που έχουν επηρεαστεί σημαντικά από την παρουσία των NT:

(α') τη συλλογή των δεδομένων της έρευνας:

(β') τις επικοινωνιακές δυνατότητες, τόσο στο επίπεδο πρόσβασης του ερευνητή σε πηγές γνώσης και πληροφορίας (βάσεις δεδομένων, οπτικοί δίσκοι), όσο και στο επίπεδο της συνεργασίας των ερευνητών (μέσω διεθνών δικτύων).

(γ') την επεξεργασία και ανάλυση ποσοτικών και ποιοτικών δεδομένων.

## 2. Συλλογή των Δεδομένων

Ενώ οι H/Y χρησιμοποιούνται χρόνια τώρα στην κοινωνική έρευνα, κυρίως για την ανάλυση ποσοτικών δεδομένων, στον τομέα της συλλογής δεδομένων άρχισαν να χρησιμοποιούνται σε ευρεία κλίμακα μόνο πρόσφατα.

Η συλλογή των στοιχείων μιας έρευνας, το γνωστό field research (έρευνα πεδίου), γίνεται πια, ολοένα συχνότερα, με τη βοήθεια υπολογιστή. Ο ερευνητής που αναλαμβάνει να πάρει τις συνεντεύξεις, εφοδιασμένος μ' έναν φορητό H/Y, βλέπει να ξετυλίγονται στην οθόνη οι διάφορες ερωτήσεις του ερωτηματολογίου της έρευνας συνοδευόμενες από σχετικές παρατηρήσεις-οδηγίες, τις οποίες οφείλει να λάβει υπόψη του. Οι κατά περίπτωση επιμέρους διαδρομές στο ερωτηματολόγιο υιοθετούνται αυτόματα, χωρίς να αφήνουν περιθώρια παρεκκλίσεων, ενώ οι έλεγχοι στην είσοδο των στοιχείων περιορίζονται στο ελάχιστο τα πιθανά λάθη κωδικοποίησης.

Η συνέντευξη με τη βοήθεια H/Y (Computer Assisted Interviewing - CAI) έχει ρίζες στην από χρόνια χρήση των H/Y (στις ΗΠΑ), στον όλο έλεγχο και συντονισμό τής πολύ διαδεδομένης εκεί μεθόδου της τηλεφωνικής συνέντευξης (Computer Assisted Telephone Interviewing - CATI). Στην Ευρώπη, όπου για πολλούς λόγους η έρευνα μέσω τηλεφώνου ήταν μέχρι πρότινος πολύ περιορισμένη, η συνέντευξη με τη βοήθεια H/Y έχει διαφορετική ιστορία. Οργανισμοί όπως το Κεντρικό Γραφείο Στατιστικής της Ολλανδίας και το Γραφείο Απογραφών Πληθυσμού και Ερευνών της Μεγάλης Βρετανίας επιχείρησαν το 1987, για πρώτη φορά, μια έρευνα υπολογισμού του εργατικού δυναμικού με τη βοήθεια H/Y μέσω προσωπικών συνεντεύξεων (Computer Assisted Personal Interviewing - CAPI).

Στις μέρες μας, διεθνώς, οι μεγαλύτερες δημόσιες και ιδιωτικές έρευνες μέσω συνεντεύξεων γίνονται με τη βοήθεια υπολογιστή. Στον ελληνικό χώρο μόνον τα τελευταία χρόνια βλέπουμε μια πρώτη εμφάνιση των μεθόδων αυτών.

Η λεγόμενη «αυτο-συνέντευξη» (Computer Assisted Self-Interviewing) αφορά περιπτώσεις όπου το ερωτηματολόγιο υπό μορφή δισκέτας H/Y αποστέλλεται προς συμπλήρωση στον ερωτώμενο (όπως συμβαίνει συνήθως με τα ερωτηματολόγια οργανισμών, π.χ., της Γενικής Γραμματείας Έρευνας και Τεχνολογίας και της Ευρωπαϊκής Ένωσης, που αφορούν το επιστημονικό δυναμικό, τον εξοπλισμό εταιρειών κλπ.) ή όπου ο ερευνητής που παίρνει τη συνέντευξη στρέφει τον υπολογιστή προς τον ερωτώμενο όταν οι ερωτήσεις θίγουν ιδιαιτέρως ευαίσθητα θέματα (όπως χρήση ναρκωτικών ουσιών, παρενόχληση σε ξουαλικού χαρακτήρα κλπ.). Επίσης αφορά την περίπτωση όπου παραχωρούνται υπολογιστές σε νοικοκυριά για την περιοδική εισαγωγή και αποστολή στοιχείων.

## 2.1. Τα προτερήματα της συνέντευξης με τη βοήθεια H/Y είναι κυρίως τα παρακάτω:

(α') Αποφυγή χαρακτηριστικών λαθών κωδικοποίησης, με αποτέλεσμα τη βελτιωμένη ποιότητα των δεδομένων.

Λόγω του ότι η όλη ροή του ερωτηματολογίου και οι διαδρομές μέσα σ' αυτό είναι εκ των προτέρων πολύ προσεκτικά προγραμματισμένες, η επόμενη κατάλληλη ερώτηση εμφανίζεται αυτόμata στην οθόνη του φορητού H/Y, ελαχιστοποιώντας έτσι το ενδεχόμενο ο ερευνητής να υποβάλει λάθος ερώτηση, να παραλείψει ερωτήσεις, να εισαγάγει λάθος απαντήσεις ή απαντήσεις πέραν των προβλεπομένων, που είναι μερικά από τα πιο χαρακτηριστικά λάθη στις έρευνες πεδίου. Επίσης, επειδή κατά τη διάρκεια της εισαγωγής των στοιχείων γίνεται και ο έλεγχος τους, είναι δυνατόν να διαλευκανθούν επιτόπου πιθανές παρανόησεις, πράγμα που συνήθως είναι αδύνατο να επιτευχθεί εκ των υστέρων από την κεντρική υπηρεσία.

(β') Ελάττωση του χρόνου ελέγχου και συγκέντρωσης των στοιχείων, με αποτέλεσμα την αμεσότερη διαθεσιμότητα των στοιχείων για περαιτέρω επεξεργασία.

Είναι προφανές ότι ο χρόνος ελέγχου των στοιχείων περιορίζεται στο ελάχιστο, μιας και ο έλεγχος γίνεται πλέον, στο μεγαλύτερο μέρος του, κατά τη διάρκεια της εισαγωγής. Επίσης περιορίζεται και ο χρόνος συγκέντρωσης των στοιχείων στην κεντρική υπηρεσία, αφού, καθώς βρίσκονται σε ψηφιακή μορφή, διακινούνται ευκολότερα και ταχύτερα.

(γ') Αποτελεσματικότερη διαχείριση της έρευνας πεδίου.

Η ψηφιακή μορφή των δεδομένων, και ως εκ τούτου η δυνατότητα άμεσης διακίνησής τους, επιτρέπει τον μεγαλύτερο έλεγχο του έργου των ερευνητών.

(δ') Μείωση του κόστους της έρευνας.

Η αποφυγή του σταδίου της εισόδου (πληκτρολόγησης) των δεδομένων από το χαρτί όπου ήταν άλλοτε γραμμένα, η μείωση των ελέγχων στην κεντρική υπηρεσία, καθώς και η μη βασισμένη στο χαρτί πια έρευνα πεδίου (με τα σχετικά έξοδα εκτύπωσης, ταχυδρομείου κλπ.), συμβάλλουν στη μείωση του κόστους της έρευνας.

(ε') Ασφάλεια και απόρρητο των δεδομένων του ερωτηματολογίου.

Υπό την προϋπόθεση ότι ακολουθούνται συστηματικά ορισμένοι κανόνες, τα σε ψηφιακή μορφή δεδομένα της έρευνας μπορούν να διαφυλαχθούν και, εάν είναι αναγκαίο, να κρυπτογραφηθούν αμέσως με την είσοδό τους στον H/Y, κάτι που δεν ήταν δυνατόν όταν τα στοιχεία βρίσκονταν στο χαρτί. Το ίδιο ισχύει και για τη διακίνηση των στοιχείων που τώρα γίνεται ηλεκτρονικά, συνήθως μέσω modem. Φόβοι ότι τα δεδομένα στην ψηφιακή τους μορφή μπορεί πιο εύκολα να γίνουν κτήμα ακατάλληλων προσώπων ή υπηρεσιών, να αντιγραφούν και να χρησιμοποιηθούν για σκοπούς ανεξάρτητους από την έρευνα —κάτι που αφορά ενγένει την πληροφορία σε ψηφιακή μορφή—, εκ πρώτης όψεως δεν θα έπρεπε να υπάρχουν, τουλάχιστον σε ευρεία κλίμακα. Ωστόσο στην εποχή μας, την εποχή των hackers και των crackers, συχνά επαληθεύονται.

Η τεχνική της συλλογής στοιχείων με τη βοήθεια υπολογιστή έγινε γρήγορα αποδεκτή τόσο από τους ερευνητές όσο και από τους ερωτώμενους. Πιο συγκεκριμένα, οι ερωτώμενοι παρουσιάζονται γενικώς αδιάφοροι ή και θετικοί στην αποτύπωση της προσωπικής τους πραγματικότητας μέσω ενός ισχυρού και «έξυπνου» μηχανήματος, οι δε ερευνητές επικεντρώνουν τον προβληματισμό τους σε θέματα όπως είναι, για πα-

ράδειγμα, ο πλέον εύχρηστος τύπος της μηχανής ή του λογισμικού (hardware, software) που χρησιμοποιούν, καθώς και το ζήτημα της προσωπικής τους ασφάλειας, αφού ο εξοπλισμός που διαθέτουν είναι πολύ ακριβός.

## 2.2. Τα μειονεκτήματα της χρήσης Η/Υ στη συλλογή των στοιχείων μιας έρευνας είναι κυρίως τα παρακάτω:

(α') Απαιτείται περισσότερος χρόνος για την προετοιμασία της έρευνας, διότι πρέπει να αναπτυχθούν τα σχετικά προγράμματα του υπολογιστή (λογισμικό) ή/και να γίνει προσαρμογή κάποιων χαρακτηριστικών προγραμμάτων πακέτων.

(β') Το κόστος για τη μετάβαση από το χειρωνακτικό, ούτως ειπείν, σύστημα στο ηλεκτρονικό είναι σημαντικό. Ο εξοπλισμός (φορητοί Η/Υ, modems κλπ.) που απαιτείται, το εξειδικευμένο λογισμικό και η απαραίτητη κατάρτιση των ερευνητών είναι σημαντικές δαπάνες που διαμορφώνουν το αρχικό κόστος του εγχειρήματος. Θεωρητικά, τα αρχικά αυτά έξοδα αποσβένονται μέσα σε λίγα χρόνια, λόγω του χαμηλότερου κόστους διενέργειας των ερευνών.

(γ') Η μετάβαση από το ένα σύστημα στο άλλο απαιτεί μεταβολές στην οργανωτική δομή του οργανισμού που διενεργεί τις έρευνες, δηλαδή κατάργηση ορισμένων θέσεων εργασίας, προσαρμογή σε νέους τρόπους σκέψης, και αλλαγές που τελικά επηρεάζουν τους πάντες στην επιχείρηση.

(δ') Οι απαντήσεις που περιγράφονται με κείμενο μεγάλου εύρους είναι δύσκολο ή και αδύνατο να εισαχθούν στον υπολογιστή, μιας και απαιτείται η πληκτρολόγησή τους. Πράγματι, οι ερευνητές, που κατά κανόνα εισάγουν αριθμούς, σπάνια έχουν ευχέρεια στην πληκτρολόγηση μακροσκελών κειμένων (π.χ. απαντήσεις σε ανοιχτές ερωτήσεις) που, υπό τις προηγούμενες συνθήκες, θα συμπληρώνονταν με το χέρι από τους ίδιους ή τους ερωτώμενους.

(ε') Η χρήση της NT (δηλαδή ενός εργαλείου όπως ο Η/Υ και του σχετικού λογισμικού) στη συλλογή των στοιχείων μιας έρευνας επιβάλλει, όπως είναι φυσικό, τα όριά της στην όλη διαδικασία. Έτσι, η ευελιξία του επαγγελματία ερευνητή είναι δυνατόν να περιορίζεται, λόγω του

ότι οφείλει να αντιμετωπίσει κάποια προβλήματα που δεν έχουν προβλεφθεί από το λογισμικό. Με το ερωτηματολόγιο στο χαρτί μπορούσε να σημειώνει, π.χ., στο περιθώριο τις παρατηρήσεις του, να αλλάζει τη σειρά ορισμένων ερωτήσεων, και γενικά να το χειρίζεται έτσι ώστε να κατορθώνει τελικά να συλλέγει, ανάλογα με την περίπτωση, την επιθυμητή πληροφορία. Αυτό σήμερα, όσο προσεκτικά κι αν είναι σχεδιασμένο το λογισμικό για την πρόβλεψη τέτοιων περιπτώσεων, είναι αδύνατον, πρακτικά, να επιτευχθεί.

(στ') Είναι απαραίτητο να οργανωθεί η τεχνική υποστήριξη του εξοπλισμού που χρησιμοποιείται και η παροχή βοήθειας στον ερευνητή όταν συναντά προβλήματα σχετικά με την τεχνολογία. Ακόμη πρέπει να είναι διαθέσιμος επιπλέον εξοπλισμός για αντικαταστάσεις λόγω πιθανής βλάβης.

## 2.3. Μεταξύ των σπουδαιότερων εξελίξεων που προβλέπονται για τα συστήματα συνεντεύξεων με υπολογιστή είναι:

(α') Οι νέες δυνατότητες — πολύ πέρα από αυτές που γνωρίζουμε για τη συλλογή πληροφοριών, που θα εξασφαλίσουν οι συνεχείς βελτιώσεις στον τομέα των υπολογιστών: για παράδειγμα, η χρησιμοποίηση χρωμάτων, ώστε να παρουσιάζονται καλύτερα οι οδηγίες στον ερευνητή και στον ερωτώμενο, η εμφάνιση ρεαλιστικών εικόνων και σκηνών video στην οθόνη του Η/Υ που καλείται να σχολιάσει ο ερωτώμενος, κλπ.

(β') Η μετάβαση από ένα σύστημα πρόσωπο-με-πρόσωπο (face-to-face), όπως ήταν το σύστημα γραπτής συμπλήρωσης ερωτηματολογίων, σ' ένα σύστημα πλάι-πλάι (side-by-side), αφού ερευνητής και ερωτώμενος στην ουσία συζητούν πλέον μπροστά σε μια οθόνη.

(γ') Η χρήση ειδικών ηλεκτρονικών στιλό, μέσω των οποίων θα γίνεται πια η εισαγωγή των στοιχείων γράφοντας απευθείας στην οθόνη, θα καταστήσει τους φορητούς Η/Υ ελαφρύτερους — κάτι πολύ σημαντικό για τον ερευνητή — και θα επιτρέψει την εισαγωγή εκτενών κειμένων-απαντήσεων.

(δ') Οι βελτιώσεις στο σχετικό λογισμικό αναμένεται να αντιμετωπίσουν αποτελεσματικότερα πολύπλοκα ερωτηματολόγια.

(ε') Η δυνατότητα σύνδεσης μέσω της ηλεκτρονικής οδού καθιστά

τον ερευνητή μέλος ενός διαλογικού δικτύου, εφόσον βέβαια δεν στερείται με αυτό τον τρόπο ζωτικές κοινωνικές δραστηριότητες και ελευθερίες απαραίτητες για τη δουλειά του.

### 3. Πρόσβαση στις Πηγές και Συνεργασία Ερευνητών

Ένα άλλο ενδιαφέρον σημείο συνάντησης των NT και της σύγχρονης κοινωνιολογικής έρευνας αφορά στις νέες δυνατότητες που παρέχουν αφενός μεν οι οπτικοί ψηφιακοί δίσκοι με την τεράστια περιεκτικότητα στοιχείων που έχουν, αφετέρου δε ο H/Y, που σαν κόμβος ενός ευρύτατου δικτύου είναι πλέον ένα επικοινωνιακό μέσο.

**3.1. Τη δεκαετία του 1970 έχουμε τις πρώτες βιβλιογραφικές βάσεις δεδομένων, προσπελάσιμες μέσω του τηλεφωνικού δικτύου. Μέχρι τότε η έρευνα σε καταλόγους, βιβλία και περιοδικά ήταν δυνατή μόνο στους χώρους των βιβλιοθηκών. Τη δεκαετία του '80 το εύρος των υπηρεσιών σε δίκτυο (on-line), που ενδιαφέρουν τον κοινωνικό ερευνητή, αυξάνεται, με τη δημιουργία μεγάλου αριθμού βάσεων δεδομένων πλήρους κειμένου (full text) καθώς και βιβλιογραφικού υλικού. Επίσης, νέου τύπου πηγές πληροφοριών είναι πλέον προσπελάσιμες μέσω του δικτύου, όπως κατάλογοι βιβλιοθηκών, ευρετήρια και δεδομένα από κοινωνικές έρευνες. Προς τα τέλη της ίδιας αυτής δεκαετίας του '80 έχουμε και τις πρώτες βάσεις δεδομένων που αφορούν τον κοινωνικό ερευνητή σε μορφή οπτικών ψηφιακών δίσκων CD-ROM (Compact Disc - Read Only Memory). Το ενδιαφέρον με αυτούς τους —μόνον αναγνώσματος— δίσκους, που λειτουργούν βάσει της τεχνολογίας laser, έγκειται αφενός στην περιεκτικότητά τους (ένα CD-ROM κειμένου μπορεί να περιέχει το ισοδύναμο περίπου πεντακοσίων μέσου μεγέθους βιβλίων) και αφετέρου στη διαθεσιμότητα των πληροφοριών, που είναι πια απαλλαγμένη από τα μειονεκτήματα του δικτύου (κόστος πρόσβασης και συνδρομής, τεχνικά προβλήματα δικτύου), καθώς και στις ευκολίες διαχείρισης που παρέχει ο υπολογιστής, δηλαδή δημιουργία ευρετηρίου, δυνατότητα αναζήτησης και αναδιάταξης σε μια επιθυμητή σειρά κλπ.**

Τη δεκαετία μας χαρακτηρίζει μια ακόμη μεγαλύτερη στροφή προς

την παραγωγή εξειδικευμένων παραγωγών-τίτλων σε CD-ROM. Καθώς το κόστος της τεχνολογικής φάσης της δημιουργίας ενός CD-ROM συνεχώς μειώνεται, γίνεται πλέον δυνατή η παραγωγή και η κυκλοφορία μη εμπορικών, πιο εξειδικευμένων παραγωγών σε περιορισμένο αριθμό. Έτσι βλέπουμε συχνά να χρησιμοποιούνται οι οπτικοί δίσκοι για τη διακίνηση μεγάλων ποσοτήτων δεδομένων, μεγάλου εύρους κειμένων κλπ. Παράλληλα, οι σχετικές υπηρεσίες βάσεων δεδομένων σε δίκτυο πληθύνονται και εξειδικεύονται.

Η πληροφορία που είναι δυνατόν να βρει κανείς είτε στη μορφή των off-line οπτικών δίσκων είτε στη μορφή των on-line βάσεων δεδομένων είναι, όπως ήδη αναφέρθηκε, κυρίως συλλογές κειμένων και βιβλιογραφικές αναφορές. Έτσι, στη μια ή/και στην άλλη μορφή περιλαμβάνονται αυτή τη στιγμή περιλήψεις (abstracts) των άρθρων των σπουδαιοτέρων επιστημονικών περιοδικών (από το 1970 περίπου), διατριβές, και σε πλήρες κείμενο το σώμα πολλών γνωστών εφημερίδων (όπως ο *Economist*, οι *Times* και άλλες), καθώς και εξειδικευμένα για τον κοινωνικό ερευνητή περιοδικά.

Ο τρόπος με τον οποίο ο ερευνητής διαχειρίζεται την πληροφορία στα συστήματα αυτά στηρίζεται στην επιλογή λέξεων-κλειδιών ή φράσεων που εκφράζουν τη γνωστική περιοχή που τον ενδιαφέρει και στον κατάλληλο συνδυασμό τους μέσω λογικών παραστάσεων (boolean operators) του τύπου OR, AND και NOT. Κατασκευάζει δηλαδή ο ερευνητής δομές όπως («gender» OR «sex») AND («pay» OR «earnings» OR «salaries») AND («teaching»), που σημαίνει ότι ενδιαφέρεται π.χ. για άρθρα και κείμενα σχετικά με τις διαφορές στα εισοδήματα μεταξύ ανδρών και γυναικών στην εκπαίδευση. Ο υπολογιστής προσδιορίζει τις περιλήψεις ή τα κείμενα που ανταποκρίνονται στους συνδυασμούς των λέξεων-κλειδιών που προτάθηκαν, και ο ερευνητής διαβάζει ή εκτυπώνει τις σχετικές πληροφορίες.

Το κύριο πρόβλημα στη χρήση των on-line βάσεων δεδομένων είναι το κόστος της σύνδεσης, που, όταν δεν καλύπτεται από τον οργανισμό στον οποίο ανήκει ο ερευνητής, αποτελεί αληθινή τροχοπέδη στην έρευνά του. Μια άλλη, μικρότερη δυσκολία είναι η ανάγκη εκμάθησης της ειδικής γλώσσας αναζήτησης (query language), αν και αυτή συνήθως γίνεται με τη συνδρομή ενός ειδικού βιβλιοθηκονόμου. Μια λύση στο πρό-

βλημα του κόστους καθώς και στη χρονική πίεση (που δρα με τη μορφή «ταξιμέτρου») στην περίπτωση των on-line υπηρεσιών δίνεται, όπως είπαμε, με τους οπτικούς δίσκους, εφόσον βέβαια υπάρχουν και περιέχουν αυτά που μας ενδιαφέρουν.

Σήμερα διεθνώς υφίστανται περίπου 5.000 βάσεις δεδομένων, στις οποίες κανένας έχει πρόσβαση είτε κατευθείαν μέσω της τηλεφωνικής γραμμής είτε μέσω ειδικής υπηρεσίας πρόσβασης, όπως είναι η Dialog Information Service, το European Space Agency, και άλλες 500 περίπου, διεθνώς, ανάλογες υπηρεσίες (παράδειγμα τέτοιας υπηρεσίας στον ελληνικό χώρο είναι ο Ερμής του Εθνικού Κέντρου Τεκμηρίωσης - EKT).

Σχετικά με τις διαθέσιμες βάσεις δεδομένων σε μορφή CD-ROM, αξίζει να αναφέρουμε μεταξύ άλλων τις εξής:

- Sociofile, που περιέχει 1.800 περιοδικά από το 1974, διατριβές από το 1986, και ενημερώνεται τρεις φορές ετησίως,
- ASSIA PLUS (Applied Social Science Index and Abstracts), που περιέχει 550 περιοδικά από το 1987 και ενημερώνεται τέσσερις φορές το χρόνο,
- Cross-cultural, με πλήρες κείμενο για θέματα όπως το σεξ, ο γάμος, η οικογένεια, η εκπαίδευση κλπ. σε διάφορες κοινωνικές και εθνικές ομάδες,
- PsychLIT της Αμερικανικής Ψυχολογικής Εταιρείας, που περιέχει 1.300 περιοδικά και βιβλία από το 1974 και ενημερώνεται τέσσερις φορές το χρόνο,
- ERIC, του Educational Resources Center του αμερικανικού Υπουργείου Παιδείας, με ερευνητικά αποτελέσματα και 750 περιοδικά με ενημέρωση επίσης τέσσερις φορές το χρόνο,
- Econlit, που αντιστοιχεί στο *Journal of Economic Literature* και αφορά περιοδικά, διατριβές και βιβλία από το 1969, και
- η βάση δεδομένων του International Statistical Yearbook, που περιέχει περίπου 500.000 χρονολογικές σειρές από υπηρεσίες όπως η Ευρωπαϊκή Ένωση, ο OECD και η γερμανική Bundesbank.

Η χρήση των NT δίνει, όπως βλέπουμε, στον σύγχρονο κοινωνικό ερευνητή τη δυνατότητα να εργάζεται ουσιαστικά με μια παγκόσμια βι-

βλιοθήκη, που δεν τον υποχρεώνει να βρίσκεται κάπου σαν φυσική παρουσία, ενώ του δίνει τη δυνατότητα να «κατεβάσει» στον υπολογιστή του πληροφορίες και κείμενα που τον αφορούν και, όπως είπαμε και προηγουμένως, να τα διαχειρίστει σαν ψηφιακή πληροφορία, βοηθούμενος από τα χαρακτηριστικά εργαλεία που διαθέτει κάθε H/Y. Βέβαια η διαδικασία ανεύρεσης της πληροφορίας σε μια τέτοια εικονική (virtual) βιβλιοθήκη είναι διαφορετική απ' ό, τι στις γνωστές μας βιβλιοθήκες. Πέρα από τα τεχνολογικά προβλήματα, που ακόμη είναι συχνά στο δίκτυο, και τις μεγάλες καθυστερήσεις λόγω υπερφόρτωσης των γραμμών παλιάς τεχνολογίας, ο ερευνητής πρέπει να εξοικειωθεί με την ειδική γλώσσα αναζήτησης.

Ένα άλλο είδος στοιχείων —που ενδιαφέρουν άμεσα τον κοινωνικό ερευνητή και είναι πλέον, με τη βοήθεια της νέας τεχνολογίας, διαθέσιμα σε μορφή ψηφιακής πληροφορίας και των δύο τύπων που αναφέραμε— είναι τα στατιστικά δεδομένα που προέρχονται από διάφορες έρευνες πεδίου. Πράγματι, τα στοιχεία κάθε έρευνας, είτε έχουν συλλεχθεί με τη βοήθεια H/Y είτε έχουν εκ των υστέρων εισαχθεί σε υπολογιστές, όντας σε ψηφιακή μορφή και με την προϋπόθεση της ανωνυμίας των υποκειμένων, είναι διαθέσιμα στον οποιονδήποτε, οπουδήποτε και αν αυτός βρίσκεται, οποτεδήποτε κι αν θέλει να τα επεξεργαστεί και με όποιον τρόπο επιθυμεί. Έτσι σήμερα υπάρχουν διεθνώς, σε ψηφιακή μορφή, πολλές συλλογές δεδομένων (microdata, όπως συνηθίζεται να λέγονται), που προέρχονται από διάφορες έρευνες, δημοσκοπήσεις και απογραφές. Οι συλλογές αυτές αφορούν όλες τις κύριες ανθρώπινες δραστηριότητες, από απογραφές πληθυσμού και κτηρίων μέχρι έρευνες για το εργατικό δυναμικό, το έγκλημα, τη σεξουαλική συμπεριφορά, το στιλ ζωής κλπ. Έτσι, για παράδειγμα, το General Household Survey στη Μεγάλη Βρετανία διαθέτει σε ψηφιακή μορφή δεδομένα σχετικά με την εκπαίδευση, την ανεργία, την υγεία, την κατοικία, τα έσοδα κλπ. που αφορούν 20.000 νοικοκυριά, με ετήσια ενημέρωση και έτος έναρξης το 1971, ενώ στα δεδομένα του Family Expenditure Survey, που βασίζονται σε ετήσιες έρευνες σε 12.000 νοικοκυριά της Μεγάλης Βρετανίας από το 1953, μπορεί να βρει ο ερευνητής στοιχεία για την όλη οικονομική διαχείριση στο επίπεδο του νοικοκυριού.

**3.2.** Οι τεχνολογίες και υπηρεσίες στις οποίες αναφερθήκαμε στο προηγούμενο Κεφάλαιο, που με τη χρήση του Η/Υ δίνουν νέες δυνατότητες στον κοινωνικό ερευνητή, αφορούν πληροφορίες προερχόμενες από άλλους (τρίτους). Όμως η ψηφιακή επικοινωνία (Computer Mediated Communication - CMC) μέσω των διεθνών δικτύων του τύπου INTERNET και JANET (Joint Academic Network) παρέχει επιπλέον τη δυνατότητα στους ερευνητές να επικοινωνούν μεταξύ τους, να ανταλλάσσουν πληροφορίες, να μοιράζονται εμπειρίες, να επεξεργάζονται θέματα από κοινού — να παράγουν δηλαδή νέα πληροφορία.

Πράγματι, ένα σύνολο υπηρεσιών, από το απλό ηλεκτρονικό ταχυδρομείο (electronic-mail, E-mail) και τους καταλόγους συνδρομητών-χρηστών (Listserv) που παρέχουν δυνατότητα ενός με πολλούς (one-to-many) επικοινωνίας, μέχρι τα bulletin boards και τις computer conferences με την πολλοί-με-πολλούς (many-to-many) επικοινωνία:

- φέρνουν σε επαφή τους ερευνητές μεταξύ τους ταχύτατα, όσο μακριά κι αν βρίσκονται — χωρίς μάλιστα να απαιτούν την ταυτόχρονη φυσική παρουσία τους στο δίκτυο, αφού τα μηνύματα μπορούν να αποθηκευτούν στον Η/Υ.
- επιτρέπουν την εύκολη αποστολή μηνυμάτων σε ευρύτατα και διασκορπισμένα στο χώρο σύνολα αποδεκτών, καθώς και τη συνεργασία και επεξεργασία ενός θέματος από κοινού.

#### 4. Επεξεργασία και Ανάλυση Δεδομένων

**4.1.** Πολύ συχνά ο κοινωνικός ερευνητής επιθυμεί τα δεδομένα του — που συγκεντρώθηκαν συνήθως από μια έρευνα μέσω ερωτηματολογίου, κωδικοποιήθηκαν και αποθηκεύτηκαν σε κάποιον Η/Υ, και που θ' αποτελέσουν την αφετηρία αναλύσεων, για τον προσδιορισμό της «κοινωνιολογικής ιστορίας» του φαινομένου που αυτά περιγράφουν — να οργανωθούν συστηματικά και αποτελεσματικά. Αυτό, που αποτελεί συχνά και το πρώτο στάδιο της επεξεργασίας των δεδομένων, επιτυγχάνεται μέσω κατάλληλου λογισμικού, που ονομάζεται σύστημα διαχείρισης βάσης δεδομένων (Data Base Management System - DBMS).

Οι κυριότεροι λόγοι για τους οποίους ο κοινωνικός ερευνητής θα θελήσει να χρησιμοποιήσει ένα πρόγραμμα διαχείρισης βάσης δεδομένων — με άλλα λόγια, δηλαδή, θα θελήσει να αποθηκεύσει στον Η/Υ τα δεδομένα του με μια συγκεκριμένη λογική δομή —, είναι γιατί μ' αυτό τον τρόπο έχει τη δυνατότητα πολύ εύκολα και γρήγορα:

(α') να συνδύασει δεδομένα διαφόρων τύπων, που εμπειρικλείουν όμως κοινή πληροφορία, δημιουργώντας έτσι νέα πληροφορία.

(β') να αναδιατάξει τα δεδομένα σε σχέση με διάφορα χαρακτηριστικά, επιτυγχάνοντας έτσι την από διαφορετικές οπτικές γωνίες θεώρησή τους.

(γ') να δημιουργήσει νέα σύνολα στοιχείων, υποσύνολα του αρχικού συνόλου, που χαρακτηρίζονται από συγκεκριμένες προδιαγραφές για τα στοιχεία τους (τιμές, ιδιότητες κλπ.).

(δ') να ανακτήσει το σύνολο ή μέρος των στοιχείων της βάσης δεδομένων σε μια επιθυμητή μορφή ανάλογα με διάφορα κριτήρια.

(ε') να μεταβάλει (ενημερώσει, διαγράψει) τις τιμές των δεδομένων καθ' όλη τη διάρκεια ζωής της βάσης.

Ο τύπος των δεδομένων για τη διαχείριση των οποίων θα μπορούσε να πει κανείς ότι είναι ιδιαίτερα αποτελεσματική η δημιουργία μιας βάσης δεδομένων περιγράφεται σαν ένα ευρύ σύνολο περιπτώσεων που μπορούν αρκετά εύκολα να ταξινομηθούν με βάση διάφορα κριτήρια τα οποία περιγράφουν συγκεκριμένα αντικείμενα ενδιαφέροντος. Κατ' αυτό τον τρόπο, δεν φαίνεται να είναι ενδεδειγμένη η δημιουργία μιας βάσης δεδομένων όταν διαθέτει κανείς περιορισμένα σύνολα στοιχείων (της τάξης των μερικών εκατοντάδων), γιατί η όλη προσπάθεια για την κατασκευή της βάσης και την εκμάθηση της διαχείρισής της είναι αναντίστοιχη με το αποτέλεσμα, ούτε όταν τα δεδομένα έχουν τη μορφή πλήρους κειμένου, οπότε είναι προτιμότερο να χρησιμοποιηθεί ένα εξειδικευμένο πρόγραμμα διαχείρισης κειμένων.

Υπάρχουν διαφόρων ειδών προγράμματα βάσεων δεδομένων, ανάλογα με το μοντέλο που χρησιμοποιούν για να ορίσουν τη δόμηση των δεδομένων και τους κανόνες που διέπουν την όλη διαχείρισή τους. Σήμερα τα περισσότερα συστήματα βάσεων δεδομένων βασίζονται στο σχεσιακό μοντέλο, σύμφωνα με το οποίο η πραγματικότητα ενός φαινομένου

νου περιγράφεται από μια συλλογή πινάκων που ονομάζονται σχέσεις (relations). Το προτέρημα της σχεσιακής λογικής είναι ότι δεδομένα από διαφορετικούς πίνακες μπορούν να συσχετιστούν και να θεωρηθούν σαν ένα για τις ανάγκες της αναζήτησης, της αποθήκευσης και της παροχής πληροφοριών. Η διαχείριση μιας βάσης δεδομένων γίνεται με τη βοήθεια μιας γλώσσας αναζήτησης (query language) που βασίζεται σε απλά μαθηματικά της θεωρίας συνόλων, όπως είδαμε στη σχετική με τις σε δίκτυο βάσεις δεδομένων παράγραφο (§3.1.).

Τρεις διαφορετικές προσεγγίσεις του σχεσιακού μοντέλου βάσης δεδομένων δημιουργούν και ισάριθμες κατηγορίες προϊόντων. Έτσι, προγράμματα - διαδοχικές εκδόσεις του τύπου dBase, καθώς και το Clipper, το FoxBase και το FoxPro, που χαρακτηρίζονται από μια περιορισμένη δυνατότητα σύνδεσης (linkage) των στοιχείων των διαφόρων πινάκων σχέσεων, αποτελούν την πρώτη κατηγορία τύπου dBase. Το «ιστορικό» προτέρημα των βάσεων αυτών είναι το ότι απαιτούν περιορισμένο χώρο στη μνήμη του υπολογιστή — κάτι που με την εξέλιξη των υπολογιστών δεν έχει πια ιδιαίτερη σημασία. Στη δεύτερη κατηγορία ανήκουν συστήματα διαχείρισης βάσεων δεδομένων όπως τα Ingres, Oracle και Informix, που χρησιμοποιούνται για ευρύτατες σχεσιακές βάσεις και χαρακτηρίζονται από μια δομημένη γλώσσα αναζήτησης (Structured Query Language - SQL). Σαν τρίτη κατηγορία σχεσιακών βάσεων δεδομένων μπορεί να χαρακτηρισθούν τα προϊόντα που βασίζονται σε μια «φιλική» προσέγγιση του χρήστη, που επιτυγχάνουν δηλαδή τη δημιουργία της βάσης με τη χρήση παραθύρων, εικονίδιων, του ποντικιού και διαφόρων δεικτών, ανταποκρινόμενα πλήρως στην οπτική άποψη που έχει ο ερευνητής για τις διασυνδέσεις στα δεδομένα του. Είναι προφανές ότι ο χωρίς ιδιαίτερες απαιτήσεις χρήστης, που δεν θέλει να μπει στη διαδικασία έστω και περιορισμένης εκμάθησης στοιχείων προγραμματισμού για τη δημιουργία και τη διαχείριση της βάσης των δεδομένων του, θα βρει στην κατηγορία αυτή ενδιαφέροντα εργαλεία. Προϊόντα αυτής της κατηγορίας είναι το FileMaker Pro, το Paradox και το 4th Dimension.

Προτού προχωρήσουμε στο δεύτερο στάδιο επεξεργασίας των δεδομένων μιας έρευνας, στο στάδιο δηλαδή της ανάλυσης, του προσδιορισμού και της ερμηνείας των σχέσεων μεταξύ των δεδομένων, θα συμπλη-

ρώσουμε τα σχετικά με τις βάσεις δεδομένων με δύο τύπους προϊόντων που ενδιαφέρουν τον κοινωνικό ερευνητή:

Το σύστημα SIR (Scientific Information Retrieval) ανήκει στις σχεσιακές βάσεις δεδομένων και προσφέρει πολύ ικανοποιητικές δυνατότητες οργάνωσης και διαχείρισης ιεραρχικής μορφής πληροφορίας, όπως είναι η σχετική με τη μονάδα του νοικοκυριού. Σ' αυτή την τόσο χαρακτηριστική περίπτωση δεδομένων στις έρευνες πεδίου η μονάδα είναι το νοικοκυριό που περιέχει έναν μεταβλητό αριθμό ατόμων. Για κάθε νοικοκυριό το πρόγραμμα δημιουργεί μια εγγραφή και διαφορετικές εγγραφές για κάθε άτομο. Το SIR, διαθέτοντας ένα λεξικό που καταχωρεί την περιγραφή και τον ορισμό των δεδομένων, καθώς επίσης και τη δυνατότητα συνεχούς ενημέρωσης με νέα στοιχεία, δημιουργίας συσχετίσεων μεταξύ των εγγραφών και υπολογισμού νέων μεταβλητών, είναι ένα χρήσιμο βοήθημα για τον κοινωνικό ερευνητή, κυρίως για την αντιμετώπιση μεγάλων και σύνθετων συλλογών ιεραρχικών δεδομένων.

Τα προγράμματα QUANVERT και QuickTAB είναι ιδιαίτερα χρήσιμα για την ανάλυση των ευρύτατων συνόλων στοιχείων των δημοσκοπήσεων από τον μη ειδικό. Η λογική της δόμησης της πληροφορίας με βάση τη μεταβλητή και όχι την περίπτωση, καθώς και ο τρόπος διαχείρισης που διαθέτουν, ελαττώνει σημαντικά τον απαιτούμενο χώρο στο σκληρό δίσκο και περιορίζει τον υπολογιστικό χρόνο στον H/Y, με αποτέλεσμα να γίνεται δυνατή η επεξεργασία μεγάλων συνόλων δεδομένων σε πλατφόρμα μικρούπολογιστή.

4.2. Σχετικά με τη στατιστική ανάλυση των δεδομένων που προέρχονται από τις διάφορες έρευνες, με τη βοήθεια H/Y, θα πρέπει να υπενθυμίσουμε ότι η πρώτη έκδοση του παρόντος βιβλίου συνέπεσε με την ευρεία διάδοση της χρήσης στατιστικών προγραμμάτων σε μεγάλα συστήματα H/Y (mainframe). Προγράμματα-πακέτα όπως το SPSS (Statistical Package for the Social Sciences) και το BMDP (BioMeDical Program) παρέχουν τη δυνατότητα στον κοινωνικό ερευνητή να επιχειρήσει στατιστικές επεξεργασίες, τόσο κλασικές (χαρακτηριστικές τιμές, κατανομές, έλεγχοι υποθέσεων) όσο και πιο πολύπλοκες (παραγοντικές αναλύσεις, διακριτική ανάλυση). Την ίδια δεκαετία (μέσα '60 με μέσα '70) έχουμε

την ανάπτυξη και ουσιαστική συμβολή στην ανάλυση δεδομένων (Analyse des Données) της γαλλικής σχολής του καθηγητή J.-P. Benzécri. Εξελίσσονται κυρίως η Παραγοντική Ανάλυση Αντιστοιχιών (Analyse Factuelle des Correspondances) και η Αυτόματη Ταξινόμηση (Classification Automatique). Η χωρίς a priori μεθοδολογία της πολυδιάστατης Περιγραφικής Στατιστικής με τη βοήθεια του H/Y θα αποτελέσει στη συνέχεια χρήσιμο εργαλείο για επιστήμονες απ' όλους τους χώρους των δυνατών εφαρμογών, καταγραφόμενη σαν εμπειρία στο περιοδικό *Les Cahiers de l'Analyse Des Données*.

Τη δεκαετία του '80 η διάδοση των μικροϋπολογιστών προσαρμόζει τα ήδη υπάρχοντα για μεγάλους υπολογιστές στατιστικά πακέτα, και δημιουργεί πλήθος άλλων για τους συμβατούς IBM και τους Macintosh. Τα πιο συχνά χρησιμοποιούμενα προγράμματα-πακέτα για τη στατιστική επεξεργασία δεδομένων είναι αυτή τη στιγμή, τόσο σε προσωπικούς υπολογιστές όσο και σε μεγαλύτερα συστήματα (workstations, mainframes), το SPSS (Statistical Package for the Social Sciences) και το SAS (Statistical Analysis System). Και τα δύο προγράμματα στο βασικό τους τμήμα προσφέρουν, εκτός από συγκεκριμένες στατιστικές τεχνικές, ισχυρή και ευέλικτη διαχείριση δεδομένων, δυνατότητες πινακοποίησης δεδομένων και αποτελεσμάτων, δυνατότητες δημιουργίας και επεξεργασίας στατιστικών γραφικών, καθώς και βοήθημα για τη δημιουργία αναφορών. Στις στατιστικές τεχνικές περιλαμβάνονται οι κατανομές συχνοτήτων, τα κλασικά περιγραφικά στατιστικά μέτρα, η κατασκευή πινάκων διπλής εισόδου και οι σχετικές δοκιμασίες, η ανάλυση διακύμανσης (ANOVA), η γραμμική παλινδρόμηση κ.ά. Τα πρόσθετα τμήματα αμφοτέρων των προγραμμάτων είναι αυτά μέσω των οποίων είναι δυνατό να επιχειρήσει κανείς μια πιο προχωρημένη, πιο εξειδικευμένη στατιστική ανάλυση, όπως π.χ. παραγοντική ανάλυση (factor analysis), ανάλυση συστάδων (cluster analysis), διακριτική ανάλυση (discriminant analysis), ανάλυση αξιοπιστίας (reliability analysis), μη γραμμική παλινδρόμηση, πολυπαραγοντική ανάλυση διακύμανσης (MANOVA), χρονολογικές σειρές, ποιοτικό έλεγχο κ.ά. Επίσης και τα δύο προγράμματα επεξεργάζονται, σε περιορισμένο βαθμό, ποιοτικές μεταβλητές.

Σχετικά με τη σύγκριση των δύο πακέτων, που ουσιαστικά παρέχουν

τις ίδιες δυνατότητες ανάλυσης, αναφέρουμε πως έχει παρατηρηθεί ότι, ενώ το ευρύ κοινό προτιμά τον όλο τρόπο λειτουργίας του πακέτου SPSS (συνοδευτικά εγχειρίδια, παρουσίαση κλπ.), οι στατιστικοί και οι ειδικοί αναλυτές τείνουν περισσότερο προς τον τρόπο αντιμετώπισης των διαφόρων προβλημάτων που παρέχει το SAS. Και για τα δύο προγράμματα πάντως, στις για περιβάλλον windows σε μικροϋπολογιστή πρόσφατες εκδόσεις τους, απαιτείται πλατφόρμα με επεξεργαστή 486 τουλάχιστον, μνήμη RAM 16Mb, αποθηκευτικός χώρος στο σκληρό δίσκο τουλάχιστον 45Mb για το βασικό μόνο τμήμα τους κι επιπλέον χώρος για κάθε πρόσθετο τμήμα, SVGA οθόνη, ενώ συνιστάται η ύπαρξη μαθηματικού συνεπεξεργαστή.

Εκτός από τα δύο προγράμματα στα οποία, λόγω της ευρείας διάδοσής τους, αναφερθήκαμε, υπάρχει στην αγορά μια πληθώρα προγραμμάτων-πακέτων, κατάλληλων για τη στατιστική ανάλυση δεδομένων, που «τρέχουν» τόσο σε μεγάλα συστήματα όσο και σε προσωπικούς υπολογιστές, και μάλιστα κάτω από windows. Έναν αριθμό από αυτά, συνοδευόμενα από μια περιορισμένη περιγραφή τους, δίνουμε αμέσως παρακάτω. Πάντως, εκείνος που τελικά θα αποφασίσει για το ποιο πρόγραμμα θα επιλέξει είναι ο κοινωνικός ερευνητής, σε συνάρτηση με τη χρήση που επιθυμεί και με τις σχετικές πληροφορίες που έχει από τον ευρύτερο χώρο δράσης του:

- Systat, που συνδυάζει πολύ καλές δυνατότητες διαχείρισης δεδομένων, διενέργειας στατιστικών αναλύσεων και κατασκευής γραφικών από κοινού με μεγάλη ευκολία χρήσης.
- BMDP (BioMeDical Program), που αποτελείται από μια συλλογή στατιστικών προγραμμάτων και συνήθως χρησιμοποιείται για ιατρικές και επιδημιολογικές έρευνες. Έχει πολλές δυνατότητες στατιστικής επεξεργασίας, αλλά θεωρείται λίγο δύσκαμπτο σχετικά με τη διαχείριση δεδομένων.
- SPAD.N (Système Portable pour l'Analyse des Données), που είναι ένα πλήρες πακέτο στατιστικής επεξεργασίας και ανάλυσης δεδομένων της γαλλικής σχολής. Καλύπτοντας όλο το εύρος από τις απλές περιγραφές μέχρι τις πλέον εξειδικευμένες μεθόδους, δίνει ιδιαίτερη ση-

- μασία στις μεθόδους της παραγοντικής ανάλυσης αντιστοιχιών και τις ταξινομήσεις.
- STATA, που είναι ένα καλό πλήρες πακέτο για στατιστικές αναλύσεις, κατασκευή μοντέλων και γραφικών παραστάσεων, καθώς και διαχείριση δεδομένων.
  - Statgraphics, που αν και είναι πιο περιορισμένο από τα προηγούμενα ως προς τις δυνατότητές του, ο συνδυασμός των στατιστικής επεξεργασίας με τα υψηλής ποιότητας γραφικά που παρέχει το καθιστούν ιδιαίτερα ενδιαφέρον για τον κοινό χρήστη.
  - GLIM (General Linear Interactive Modelling), που κυρίως χρησιμοποιείται όταν απαιτείται μια διαλογική σχέση με τα δεδομένα, με σκοπό τη δυνατότητα μεταβολής μοντέλου αναλόγως των αποτελεσμάτων. Η όλη διαχείριση πάντως των δεδομένων καλύτερα είναι να επαφίεται σε άλλα πακέτα και μετά να χρησιμοποιείται το GLIM, που διαθέτει εντούτοις εξαιρετικές δυνατότητες γραφικών και απεικόνισης.
  - MiniTab, που χαρακτηρίζεται από την ιδιαίτερη ευκολία εκμάθησης του καθώς και τη σχετικά φτηνή τιμή του. Η λογική του «φύλλου εργασίας» που διαθέτει το καθιστά εύχρηστο για απλές στατιστικές αναλύσεις, για επιτυχημένη γραφική παρουσίαση αποτελεσμάτων καθώς και για διδακτικούς στόχους, αλλά δεν το ευνοεί για την επεξεργασία μεγάλων συνόλων δεδομένων.
  - STATVIEW, που κάνει έναν περιορισμένο αριθμό αναλύσεων και γραφικών, με ιδιαίτερα φιλικό όμως τρόπο.

Τέλος, τα πολύ διαδεδομένα προγράμματα - φύλλα εργασίας (spreadsheets) του τύπου Lotus 1-2-3, Excel κλπ., μπορούν να χρησιμοποιηθούν για τη διενέργεια περιορισμένου αριθμού και πολυπλοκότητας αναλύσεων ιδιαίτερα από τον μη ειδικό ή όταν δεν είναι δυνατή η πρόσβαση σ' ένα εξειδικευμένο στατιστικό πακέτο. Στα θετικά των πρωταρχικών αυτών εργαλείων πρέπει κανείς να συμπεριλάβει τα ικανοποιητικά γραφικά που τα προγράμματα αυτά μπορούν εύκολα να ενσωματώσουν στις παρουσιάσεις, καθώς και τη δυνατότητα «επικοινωνίας» τους με τα προγράμματα επεξεργασίας κειμένου και τα στατιστικά πακέτα.

4.3. Επί σειρά ετών οι κοινωνικοί ερευνητές που ασχολούνταν με ποιοτικά δεδομένα (σημειώσεις από έρευνες πεδίου, συλλογές κειμένων, συνεντεύξεις και αφηγήσεις, απαντήσεις σε ανοιχτές ερωτήσεις ερωτηματολογίων) ήταν διστακτικοί σχετικά με τη χρήση των H/Y στις έρευνές τους ενγένει, και ιδιαίτερα για την ανάλυση των δεδομένων τους. Πράγματι, επειδή αφενός η ανάλυση των ποιοτικών δεδομένων θεωρείται ως μια άκρως προσωπική διαδικασία όπου κάθε προσπάθεια τυποποίησης πρέπει να αποφεύγεται, και αφετέρου επειδή τα διατιθέμενα υπό μορφή software προϊόντα δεν ήταν για πολύ χρόνο τα κατάλληλα, θεωρήθηκε ότι η χρήση H/Y για την ποιοτική έρευνα δεν ενδείκνυται. Τα τελευταία όμως χρόνια, κι αφού πρώτα οι ερευνητές εξοικειώθηκαν με τις βάσεις δεδομένων και τους επεξεργαστές κειμένου, σημειώνεται ένα αυξημένο ενδιαφέρον σχετικά με τη χρήση λογισμικού στη διαχείριση κι επεξεργασία και ποιοτικών πλέον δεδομένων. Μεταξύ των προγραμμάτων που φαίνεται να καλύπτουν σε μεγάλο βαθμό τις προσδοκίες των ερευνητών είναι τα THE ETHNOGRAPH, TAP (Text Analysis Package), QUALPRO και QUALOG, HyperRESEARCH, ETHNO, NUDIST(Non-numerical Unstructured Data Indexing, Searching and Theory building), TextBase Alpha, HyperQual, ATLAS, Sonar Professional, SPAD.T (Système Portable pour l'Analyse des Données. Textes), Lexico1, ALCESTE, HyperBASE κ.ά. Είναι ενδιαφέρον να αναφερθεί ότι τα περισσότερα έχουν δημιουργηθεί από μεμονωμένους ερευνητές και πανεπιστημιακές εργαστηριακές πρωτοβουλίες.

Για να γίνει κατανοητή η χρησιμοποίηση λογισμικού στην ποιοτική έρευνα, θα παραθέσουμε στοιχεία από ένα συγκεκριμένο παράδειγμα ανάλυσης ποιοτικών δεδομένων καθώς και τη μεθοδολογία που χρησιμοποιήθηκε. Πρόκειται για τα κείμενα - ελεύθερες απαντήσεις που έδωσαν σε δύο ανοιχτές ερωτήσεις 650 επιμορφωμένοι εκπαιδευτικοί όλων των ειδικοτήτων, στο πλαίσιο μιας ευρύτερης έρευνας σχετικά με τις γνώσεις και στάσεις των εκπαιδευτικών ως προς τις νέες τεχνολογίες και τις προσδοκίες τους από τη χρήση αυτών στην εκπαίδευση. Η πρώτη ερώτηση ζητούσε να αναφέρουν τους λόγους για τους οποίους κρίνουν θετική ή αρνητική την εισαγωγή της τεχνολογίας στην εκπαίδευση. Η δεύτερη ερώτηση ζητούσε την περιγραφή ενός τρόπου πιθανής χρήσης (σενάριο) των NT στη σχολική τάξη, με την υπόθεση ότι θα είχαν στη διάθε-

σή τους όποιο μέσο επιθυμούσαν (τηλεόραση, βίντεο, ηλεκτρονικός υπολογιστής κ.τ.λ.). Όλα τα άτομα δεν επωφελήθηκαν από την ευκαιρία που τους δόθηκε να αναπτύξουν τις απόψεις τους, και περιοριστήκαμε στις απαντήσεις 180 ατόμων που είχαν γράψει τουλάχιστον μισή σελίδα κειμένου και για τις δύο ερωτήσεις, τις οποίες κωδικοποιούμε με το (1) και το (2) αντίστοιχα. Για την ανάλυση τόσο του περιεχομένου των απαντήσεων όσο και του ύφους χρησιμοποιήθηκαν προγράμματα παραγοντικής ανάλυσης δεδομένων και αυτόματης ταξινόμησης.

#### Α'. Σημασιολογική Ανάλυση των Απαντήσεων στις Ερωτήσεις (1) και (2)

Επειδή για ορισμένα από τα 180 άτομα η απάντηση σε μία από τις δύο ερωτήσεις απουσιάζει ολοκληρωτικά ή περιορίζεται σε μία μόνο φράση, ο αριθμός των απαντήσεων δεν είναι ακριβώς 360 (δηλαδή  $2 \times 180$ ) αλλά μόνον 354. Η χρήση του λογισμικού μάς έδωσε έναν κατάλογο συμπτώσεων των τύπων λέξεων, ταξινομημένων ανάλογα με τη συχνότητα εμφάνισής τους. Από τον γενικό αυτό κατάλογο δημιουργήσαμε ένα λεξικό 122 τύπων. Επειδή ο στόχος ήταν η ανάλυση περιεχομένου, και άρα το παραπάνω λεξικό θα έπρεπε να περιοριστεί σε λέξεις που αφορούν στη χρήση των νέων τεχνολογιών στην εκπαίδευση ή στη στάση απέναντι σ' αυτήν, κατατάξαμε τελικά τις λέξεις σε 27 θέματα. Χρησιμοποιώντας στη συνέχεια την παραγοντική ανάλυση αντιστοιχιών και την αυτόματη ιεραρχική ταξινόμηση (classification automatique hiérarchique), αναλύσαμε δύο πίνακες που αντιστοιχούν ως προς τις γραμμές αμφότεροι στις 354 απαντήσεις στις ερωτήσεις (1) και (2), ενώ ως προς τις στήλες ο πρώτος στις 27 θεματικές ενότητες και ο δεύτερος στις 122 λέξεις.

Στο σημείο αυτό είναι χρήσιμο να αναφέρουμε ορισμένα στοιχεία των αρχών και της λειτουργίας των μεθόδων της ανάλυσης δεδομένων που χρησιμοποιήσαμε:

(α') Η παραγοντική ανάλυση αντιστοιχιών ανήκει στις λεγόμενες γραμμικές (επειδή χρησιμοποιούν στοιχεία της Γραμμικής Άλγεβρας) μεθόδους της πολυδιάστατης Περιγραφικής Στατιστικής. Βασική αρχή της μεθόδου αυτής είναι ότι κάθε μία από τις διαστάσεις ενός πίνακα δε-

δομένων (γραμμές ή στήλες) μας επιτρέπει να ορίσουμε την έννοια της απόστασης μεταξύ των στοιχείων της άλλης διάστασης. Κατ' αυτό τον τρόπο το σύνολο των στηλών (λεξικό 122 λέξεων ή 27 κατηγορίες αυτών, στο παράδειγμά μας) μας επιτρέπει να ορίσουμε, με τη βοήθεια κατάλληλων μαθηματικών τύπων, αποστάσεις ανάμεσα στις γραμμές — 354 απαντήσεις στις ερωτήσεις (1) και (2). Στο παράδειγμά μας, για δύο γραμμές (απαντήσεις)  $i$  και  $i'$  θα είναι:

$$d^2(i, i') = \sum_{j=1}^p \left( \frac{1}{f_{i,j}} \right) \left( \frac{f_{ij}}{f_{i,j}} - \frac{f_{i',j}}{f_{i',j}} \right)^2,$$

όπου  $f_{ij}$  και  $f_{i,j}$  είναι αντίστοιχα η σχετική συχνότητα της εμφάνισης μιας λέξης  $j$  στις απαντήσεις  $i$  και  $i'$ , και  $f_{i,j}$  και  $f_{i',j}$  οι περιθώριες συχνότητες για τις οποίες ισχύει  $f_{i,j} = \sum_{j=1}^p f_{ij}$  και  $f_{i'} = \sum_{j=1}^n f_{i',j}$ . Η απόσταση αυτή,\* που ονομάζεται απόσταση του  $x^2$ , μοιάζει πολύ με τη γνωστή ευκλίδεια απόσταση, με μόνη διαφορά τον σταθμικό όρο  $\left(\frac{1}{f_{i,j}}\right)$ .

Σύμφωνα μ' αυτόν τον τύπο, αν δύο απαντήσεις  $i$  και  $i'$  μοιάζουν από σημασιολογική πλευρά, δηλαδή περιέχουν με ανάλογες συχνότητες τις ίδιες λέξεις ή κατηγορίες λέξεων ( $j$ ), θα έχουν και μικρότερη απόσταση μεταξύ τους. (Το αντίστροφο θα ισχύει όταν οι απαντήσεις διαφέρουν αρκετά, οπότε και χαρακτηρίζονται από μεγάλη απόσταση μεταξύ τους).

Με βάση αυτή την αρχή, και με τη χρήση των παραπάνω προγραμμάτων, μπορούμε να αντιστοιχίσουμε τους αρχικούς πίνακες σε «νέφη σημείων» ενός «πολυδιάστατου γεωμετρικού χώρου». Αντιστοιχούμε δηλαδή κάθε μία από τις 354 απαντήσεις ( $n=354$ ) σ' ένα σημείο ενός χώρου που έχει σαν διαστάσεις τις 27 κατηγορίες ή τις 122 λέξεις ( $p=27$  ή  $p=122$ ). Με τον τρόπο αυτό, κάθε απάντηση (που συμβολίζεται με το  $i$ ) ορίζεται (δηλαδή τοποθετείται σαν σημείο στο χώρο) από τις σχετικές

\* Με τον ίδιο τρόπο, μέσω του συνόλου των γραμμών είναι δυνατό να υπολογισθούν αποστάσεις στο σύνολο των στηλών:

$$d^2(j, j') = \sum_{i=1}^n \left( \frac{1}{f_{i,j}} \right) \left( \frac{f_{ij}}{f_{i,j}} - \frac{f_{i,j'}}{f_{i,j'}} \right)^2.$$

συχνότητες με τις οποίες περιέχει τις 27 κατηγορίες ή τις 122 λέξεις (που είναι οι συντεταγμένες  $j$  του  $i$ ). Επειδή, δε, έχουμε ορίσει, με τον τρόπο που αναφέραμε, αποστάσεις μεταξύ των σημείων, η μορφή των συγκεντρωμένων ή όχι σημείων (νέφη) αποτελεί, μεταξύ άλλων, εργαλείο ερμηνείας του πίνακα των δεδομένων καθώς περιέχει στατιστικές πληροφορίες.\* Η δυσκολία τώρα έγκειται στο να μπορέσουμε να «πλησιάσουμε», να «δούμε» (σε μικρότερη προφανώς διάσταση) και να ερμηνεύσουμε, με τη μικρότερη δυνατή απώλεια πληροφορίας, την πολυδιάστατη αυτή εικόνα της δομής των αρχικών πινάκων που συνιστούν οι απεικονίσεις αυτές.

Το λογισμικό της παραγοντικής ανάλυσης αντιστοιχιών που χρησιμοποιήσαμε επιτυγχάνει καταρχάς τον προσδιορισμό των διευθύνσεων-αξόνων που ορίζουν τον καλύτερο, από πλευράς διατήρησης της πληροφορίας, γεωμετρικό υπο-χώρο (δηλαδή μικρότερης διάστασης από τον αρχικό χώρο) περιγραφής του νέφους. Στη συνέχεια μας δίνει τις προβολές του νέφους των σημείων, που δεν πρέπει να ξεχνάμε ότι αντιστοιχεί στον πίνακα δεδομένων μας, στα διάφορα επίπεδα που ορίζονται από τους άξονες που θ' αποτελέσουν, ερμηνεύόμενοι, τις νέες μεταβλητές στις οποίες οφείλεται η δομή των δεδομένων μας.

(β') Η αυτόματη ιεραρχική ταξινόμηση που χρησιμοποιήσαμε στο συγκεκριμένο παράδειγμα ανήκει στην οικογένεια των μεθόδων της αυτόματης ταξινόμησης, που έχουν σαν στόχο τη συγκρότηση κατηγοριών (κλάσεις) που κρίνονται ομογενείς υπό το πρίσμα ορισμένων κριτηρίων. Με τις μεθόδους αυτές ορίζονται αποστάσεις\*\* μεταξύ των στοιχείων και των ομαδοποιήσεών τους που μπορεί να δίνονται, π.χ., από τον τύπο:

\* Μη δυνάμενοι στα όρια του παρόντος να επεκταθούμε περισσότερο σε μαθηματικές και τεχνικές έννοιες, περιορίζόμαστε να πούμε ότι, όπως αποδεικνύεται από εξισώσεις μεταφοράς, είναι δυνατή η κοινή παράσταση των στοιχείων των γραμμών —απαντήσεις— και των στοιχείων των στηλών —λέξεις— συγχρόνως στα διάφορα παραγοντικά επίπεδα. Αυτό επιτρέπει την ερμηνεία των γειτνιάσεων στα επίπεδα όχι μόνο μεταξύ δύο σημείων-γραμμών και δύο σημείων-στηλών αλλά και μεταξύ ενός σημείου-γραμμής και του συνόλου των σημείων-στηλών, καθώς και μεταξύ ενός σημείου-στήλης και του συνόλου των σημείων-γραμμών.

\*\* Οι αποστάσεις αυτές δεν είναι ευκλίδειες γιατί δεν ισχύει η ιδιότητα:

$$d(x, z) + d(z, y) > d(x, y).$$

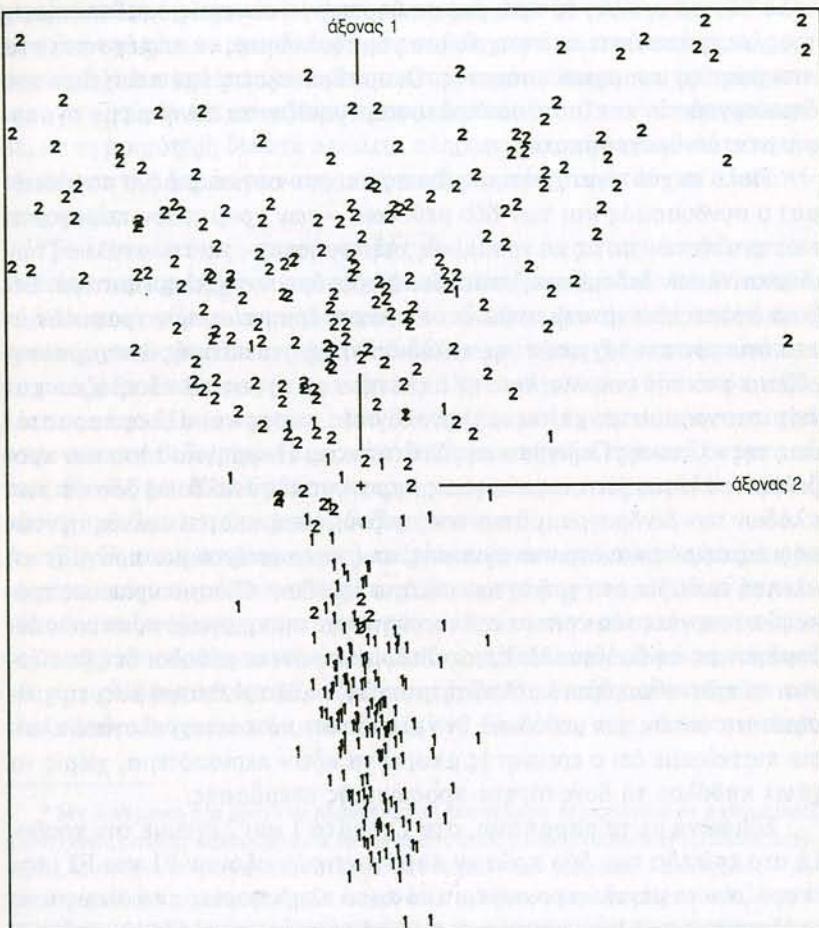
$d(h, z) = \text{Min}\{d(x, z), d(y, z)\}$ , που ονομάζεται ελάχιστο άλμα, ή από τον τύπο:  $d(h, z) = [d(x, z) + d(y, z)]/2$ , που ονομάζεται μέση απόσταση.

Στην περίπτωση της ιεραρχικής ταξινόμησης, ξεκινώντας από το σύνολο των στοιχείων, το πρόγραμμα δημιουργεί συνεχείς ομαδοποιήσεις τους, με τρόπο ώστε τα στοιχεία μιας ομαδοποίησης να περιέχονται στα στοιχεία τής ιεραρχικά επόμενης. Οι ομαδοποιήσεις των στοιχείων που δημιουργούνται κατ' αυτό τον τρόπο παρουσιάζονται συνήθως με τη μορφή ενός δενδρογράμματος.

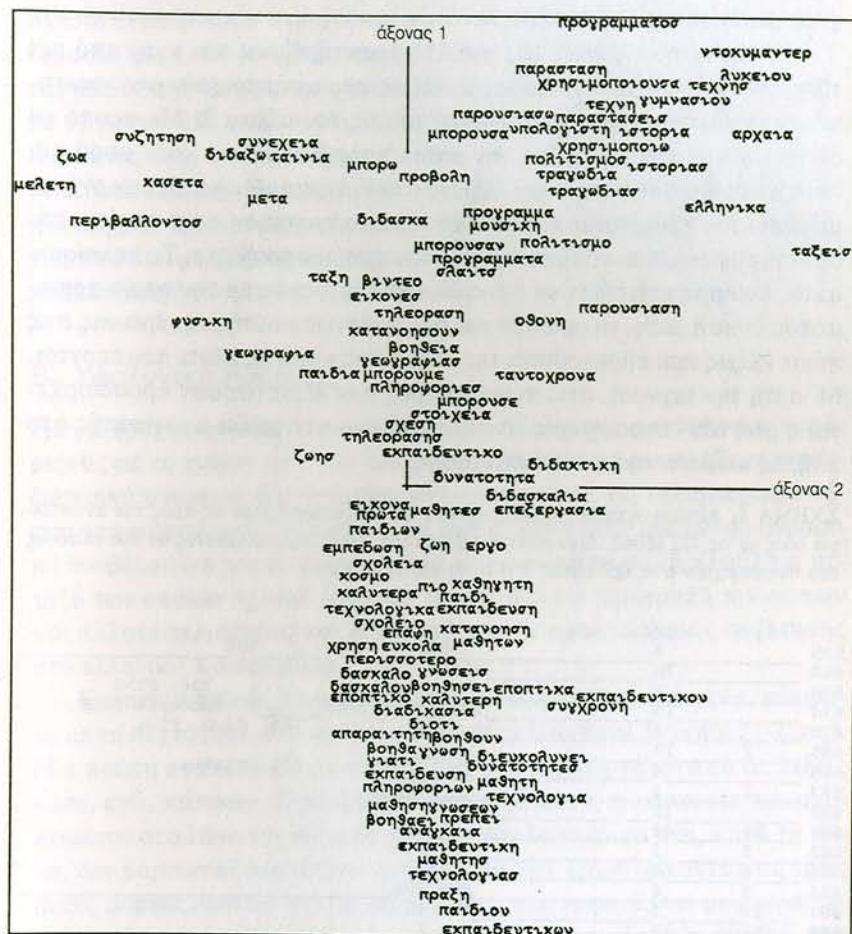
Πολύ συχνά είναι χρήσιμος (όπως και στο συγκεκριμένο παράδειγμα) ο συνδυασμός και των δύο μεθόδων —των γραμμικών παραγοντικών αναλύσεων με τις μη γραμμικές ταξινομήσεις— για την ανάλυση των ιδίων πινάκων δεδομένων, οπότε οι μέθοδοι δρουν συμπληρωματικά. Βέβαια πρέπει εδώ να σημειωθεί ότι ο τρόπος ερμηνείας των γραφικών απεικονίσεων των τεχνικών της πολυδιάστατης στατιστικής δεν χαρακτηρίζεται από την ευκολία και την απλότητα με την οποία «διαβάζει» κανείς ιστογράμματα, «πίτες», χρονολογικές σειρές και άλλες παραστάσεις της κλασικής Περιγραφικής Στατιστικής. Η ερμηνεία τόσο των προβολών του νέφους των σημείων στις παραγοντικές αναλύσεις όσο και των κλάδων των δενδρογραμμάτων στις ταξινομήσεις απαιτεί αφενός τη γνώση μιας σειράς κανόνων ανάγνωσής τους και αφετέρου μια πραγματική κλινική εμπειρία στη χρήση των ενλόγω μεθόδων. Ο συγκεκριμένος τρόπος λειτουργίας του κοινωνικού ερευνητή για την ερμηνεία ποιοτικών δεδομένων με τη βοήθεια H/Y, και δεδομένου ότι οι μέθοδοι δεν βασίζονται σε πιθανοθεωρητικό μοντέλο, ενισχύει, κατά την άποψή μας, τη χρησιμότητα αυτών των μεθόδων. Στο μεθοδολογικό και τεχνολογικό πλαίσιο πιστεύουμε ότι ο ερευνητής μπορεί να «δει» περισσότερα, χωρίς να χάνει καθόλου τη δυνατότητα προσωπικής επέμβασης.

Σύμφωνα με τα παραπάνω, στα Σχήματα 1 και 2 έχουμε την προβολή στο επίπεδο των δύο πρώτων παραγοντικών αξόνων F1 και F2 (που εκφράζουν το μεγαλύτερο συγκριτικά ποσό πληροφορίας από όλους τους υπόλοιπους) των δύο νεφών, των απαντήσεων και των λέξεων αντίστοιχα. Παρατηρώντας το νέφος των απαντήσεων που συμβολίζονται με (1) και (2), βλέπουμε ότι ο διαχωρισμός μεταξύ τους είναι περίπου τέλειος: στο ημι-επίπεδο που χαρακτηρίζεται από θετικές τιμές για τον άξονα F1

ΣΧΗΜΑ 1. Παραγοντική ανάλυση αντιστοιχιών του πίνακα 354 απαντήσεων στις ερωτήσεις (1) και (2) και 122 λέξεων. Προβολή των σημείων-απαντήσεων στο πρώτο παραγοντικό επίπεδο (άξονες 1 και 2). Διακρίνεται ο σαφής διαχωρισμός των ερωτήσεων (1) και (2).



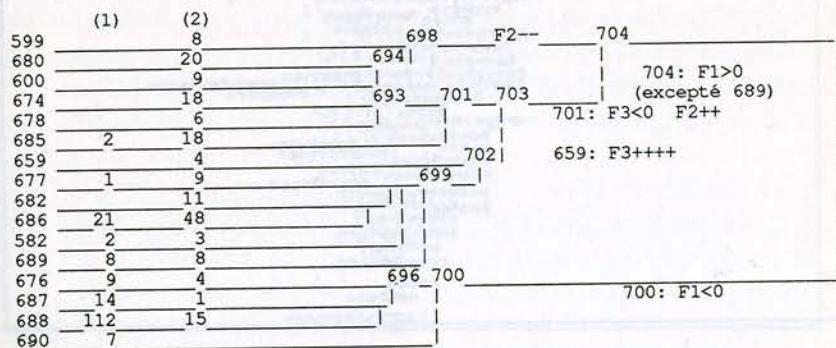
ΣΧΗΜΑ 2. Παραγοντική ανάλυση αντιστοιχιών του πίνακα των 354 απαντήσεων στις ερωτήσεις (1) και (2) και 122 λέξεων. Προβολή των σημείων-λέξεων στο πρώτο παραγοντικό επίπεδο (άξονες 1 και 2).



( $F1 > 0$ ) εξαπλώνονται όλες οι απαντήσεις «2», ενώ οι «1» κατανέμονται καθ' όλο το μήκος του ημι-άξονα ( $F1 < 0$ ) και πολύ πλησίον του. Θα επιστρέψουμε στο διαχωρισμό αυτό λίγο παρακάτω, στην αυτόματη ιεραρχική ανάλυση των απαντήσεων.

Το νέφος των λέξεων (Σχήμα 2) χαρακτηρίζεται και αυτό από μια τριγωνική μορφή, με τις διάφορες λέξεις που αντιστοιχούν στο γνωστικό αντικείμενο να διαχέονται κατά μήκος του άξονα 2. Με σκοπό να συσχετίσουμε την έννοια με την οποία χρησιμοποιείται κάθε φορά μια λέξη με τη θέση των σημείων-λέξεων στο διάγραμμά μας, δημιουργήσαμε μέσω του προγράμματος μια νέα κατηγοριοποίηση στην οποία η κάθε λέξη εμφανίζεται στο «πλαίσιο» όπου χρησιμοποιήθηκε. Το «πλαίσιο» αυτό, που μας επιτρέπει να διακρίνουμε την έννοια με την οποία χρησιμοποιήθηκε η λέξη, το ορίσαμε για την περίπτωση αυτής της έρευνας στις πέντε λέξεις που προηγούνται της λέξης μας και στις πέντε που έπονται. Μ' αυτή την τεχνική, στις περισσότερες των περιπτώσεων προσδιορίζεται η ροή του λόγου, χωρίς να απαιτείται να ανατρέξει ο ερευνητής στο πλήρες κείμενο της κάθε απάντησης.

ΣΧΗΜΑ 3. Αυτόματη ιεραρχική ταξινόμηση των 354 απαντήσεων ως προς την αντιστοιχία τους με τις 122 λέξεις. Διακρίνεται ο χωρισμός του δενδρογράμματος σε δύο κλάδους που αντιστοιχούν στις ερωτήσεις 1 ( $F1 < 0$ ) και 2 ( $F1 > 0$ ).



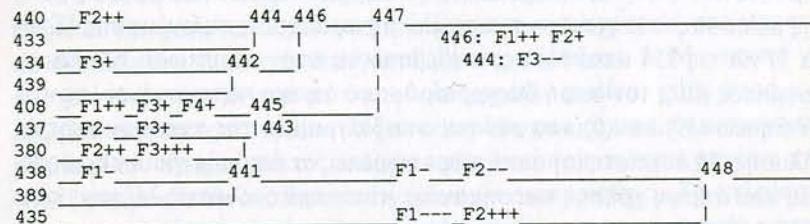
Ας δούμε, για παράδειγμα, τις χρήσεις της λέξης «ζωή», που άλλοτε μεν έχουν το χαρακτήρα γενικής αναφοράς, π.χ.: «σημαντικό ρόλο στην καθημερινή μας ζωή», και: «η τεχνολογία συνδέεται με τη ζωή», και άλλοτε αναφέρεται σ' ένα συγκεκριμένο εκπαιδευτικό σενάριο, π.χ.: «ενδυμασία και ζωή των Βυζαντινών», και: «να δείξουμε σε βίντεο τη ζωή της μέλισσας». Η χρήση της αυτόματης ιεραρχικής ταξινόμησης (Σχήμα 3) για τις 354 απαντήσεις, επιβεβαιώνει, στην περίπτωση του παραδείγματός μας, τον σαφή διαχωρισμό, από άποψη περιεχομένου, των απαντήσεων (1) και (2) που φάνηκε στο διάγραμμα της παραγοντικής ανάλυσης. Η διαπίστωση αυτή μάς επιτρέπει να δημιουργήσουμε ομογενείς από άποψη χρήσης και σημασίας κατηγορίες-κλάσεις λέξεων, κάτι πολύ ενδιαφέρον για την εξαγωγή συμπερασμάτων της έρευνας.

#### B'. Υφολογική Ανάλυση

Για να προχωρήσουμε στην υφολογική ανάλυση των απαντήσεων της έρευνας με τη χρήση H/Y και των προγραμμάτων μας, επιλέξαμε σαν εξωτερική αναφορά δύο κείμενα, και απορρίψαμε τις τηλεγραφικού τύπου απαντήσεις από την έρευνα. Τα δύο κείμενα ήταν μια αυτοβιογραφική νουβέλα (ένα μικρό μυθιστόρημα που διαιρείται σε 16 Κεφάλαια μεταξύ των οπίων έχουμε διαφορά τόνου: άλλοτε περιγραφή προσώπων και άλλοτε περιπέτειες και διάλογοι) και τα πρακτικά μιας συζήτησης στο ελληνικό Κοινοβούλιο.

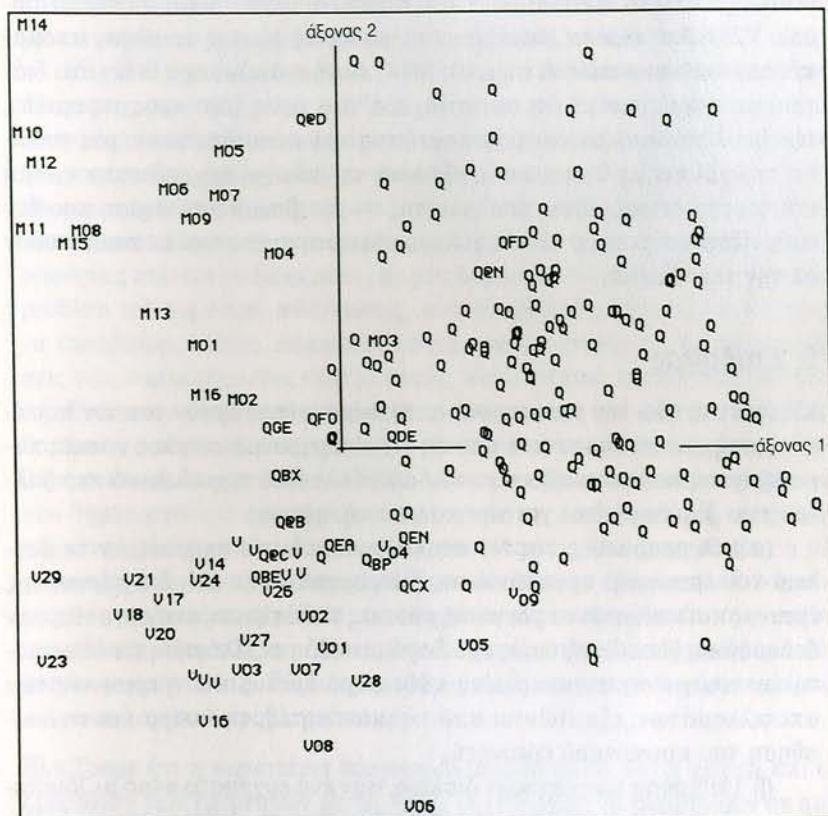
Κατασκευάσαμε, καταρχάς, και για τα τρία πλέον κείμενα, σύμφωνα με τη συχνότητα, ένα νέο λεξικό που περιελάμβανε 76 τύπους (λέξεις). Μια πρώτη ανάλυση έδειξε έναν άξονα που δημιουργείται από τις λέξεις «μου, εγώ, κάποια». Πράγματι, σε κάθε γλώσσα, η παρουσία του υποκειμένου στο λόγο του θέτει σε πρώτο επίπεδο κείμενα που, κατά τα άλλα, δεν παρουσιάζουν τίποτα το αξιοπερίεργο. Παραλείποντας τις τρεις αυτές μορφές, καθώς και τη λέξη «ενός» που εμφανίζεται με μικρή συχνότητα, τα αποτελέσματα της αυτόματης ιεραρχικής ταξινόμησης των 225 συνολικά κειμένων σε σχέση με τις 72 μορφές των λέξεων-εργαλείων δίνονται από ένα δενδρόγραμμα (Σχήμα 4) με μία αρχική διαίρεση σε δύο κλάδους: ο πρώτος αντιστοιχεί καθαρά στις απαντήσεις των εκπαιδευ-

**ΣΧΗΜΑ 4.** Αυτόματη ιεραρχική ταξινόμηση των 225 κειμένων ως προς την αντιστοιχία τους με τις 72 μορφές λέξεων. Διακρίνεται ο αρχικός χωρισμός του δενδρογράμματος σε δύο κλάδους που αντιστοιχούν στις απαντήσεις των εκπαιδευτικών Q ( $F1 > 0$ ) και τα υπόλοιπα κείμενα καθώς και ο αμέσως επόμενος ιεραρχικά χωρισμός αυτών στα Κεφάλαια του μυθιστορήματος M ( $F2 > 0$ ) και στις ομιλίες των βουλευτών V ( $F2 < 0$ ).



τικών (περιέχει ένα μόνο Κεφάλαιο του μυθιστορήματος κι έναν λόγο στο Κοινοβούλιο), ενώ ο δεύτερος μοιράζεται στο μυθιστόρημα και στους λόγους (ενώ περιέχει και μερικές από τις απαντήσεις των εκπαιδευτικών). Έχοντας συμβολίσει τα Κεφάλαια του μυθιστορήματος με {M01, M02, ..., M16}, τους λόγους των διαφόρων βουλευτών-ομιλητών με {V01, V02, ..., V29} και τις 225 απαντήσεις με το γράμμα Q, βλέπουμε στο Σχήμα 5, που μας δίνει την προβολή των νέφους των διαφόρων σημείων στο επίπεδο των δύο πρώτων παραγοντικών αξόνων, έναν σαφή καταρχάς διαχωρισμό των τριών συνόλων σημείων M, V, Q. Τα σημεία M02 και V09 που έχουν τοποθετηθεί από το πρόγραμμα στα όρια των υπολοίπων σημείων των συνόλων τους και πλησιάζουν αμφότερα το σύνολο των σημείων των απαντήσεων, αποδεικνύεται ότι έχουν ιδιαίτερο ενδιαφέρον και για την υφολογική ανάλυση. Πράγματι, επιστρέφοντας στα δεδομένα, βλέπουμε ότι το M02 αντιστοιχεί στο Κεφάλαιο του μυθιστορήματος που είναι αφιερωμένο στον πατέρα και είναι το πιο λιτό και ολιγόλογο, θυμίζοντας περισσότερο ένα curriculum vitae παρά τη διήγηση κάποιου επεισοδίου. Η περίπτωση της ομιλίας V09 είναι επίσης αξιοσημείωτη. Προκειται για την ομιλία ενός βουλευτή, πρώην υπουργό Οικονομικών, που μιλάει για τις ιδιωτικοποιήσεις, ιδιαίτερα δε στον τομέα της ναυπηγεπισκευαστικής. Αυτός ο βουλευτής αποδεικνύεται (από τα δεδομένα) ότι έχει σπουδάσει μαθηματικά, έχει στο παρελθόν διδάξει και, όπως φαίνεται (από την ανάλυση), έχει διατηρήσει στο λόγο του, κατά κάποιο τρό-

**ΣΧΗΜΑ 5.** Παραγοντική ανάλυση αντιστοιχιών των 225 κειμένων και των 72 μορφών λέξεων. Προβολή των σημείων (C) απαντήσεων των εκπαιδευτικών (Q), των Κεφαλαίων του μυθιστορήματος (M...) και των ομιλιών των βουλευτών (V...) στο πρώτο παραγοντικό επίπεδο (άξονες 1 και 2).



πο, το ύφος των πάλαι ποτέ συναδέλφων του (ένας άλλος λόγος του ίδιου βουλευτή, ο V28, δεν απομακρύνεται κι αυτός της περιοχής των εκπαιδευτικών). Αντιθέτως, η ομιλία ενός άλλου βουλευτή, που αποδεικνύεται ότι έχει ακολουθήσει μια εξαρχής πορεία κομματικού στελέχους, αντιστοιχεί από το πρόγραμμα σ' ένα σημείο αρκετά απομακρυσμένο (σημείο V29). Απ' όλα τα Κεφάλαια του μυθιστορήματος το πλέον απομακρυσμένο είναι αυτό του σημείου M14. Αυτή η απόμακρη θέση στο διάγραμμα παρατηρούμε ότι αντιστοιχεί σ' ένα ύφος ξένο προς τις ομιλίες των βουλευτών αλλά και τις απαντήσεις των εκπαιδευτικών: μια γυναικεία περιγράφει με θέρμη και ανθρωπιά τον ευτυχή ερχομό στον κόσμο και τους πρώτους μήνες του γιου της — μια βιωματική σχέση που δεν εμφανίζεται ούτε στον πολιτικό λόγο ούτε στη σχέση των εκπαιδευτικών με την τεχνολογία.

## 5. Επίλογος

Κλείνοντας εδώ την εισαγωγική αυτή παρουσίαση τριών τομέων που έχουν επηρεαστεί σημαντικά από τις NT, επιχειρούμε κάποιες γενικές παρατηρήσεις που συνοψίζουν το νέο μεθοδολογικό τεχνολογικό περιβάλλον που δημιουργείται για την κοινωνική έρευνα:

(α') Οι εφαρμογές της NT στην Κοινωνιολογία επηρεάζουν τη δουλειά του ερευνητή, οργανώνοντας διαφορετικά την όλη διαχείριση της έρευνας και καθιστώντας δυνατές κάποιες πολύπλοκες επεξεργασίες των δεδομένων. Η επιλογή όμως του λογισμικού, η αξιολόγηση των δυνατοτήτων του για τη συγκεκριμένη κάθε φορά έρευνα και η ερμηνεία των αποτελεσμάτων, εξαρτώνται από τις ικανότητες, την πείρα και τη διαίσθηση του κοινωνικού ερευνητή.

(β') Η χρήση των οπτικών δίσκων, σαν ένα εύχρηστο μέσο με ιδιαίτερα μεγάλες αποθηκευτικές ικανότητες όχι μόνο κειμένων αλλά και άλλου τύπου υλικού, από κοινού με τις εξελίξεις στο χώρο των μικροϋπολογιστών, δίνει νέες δυνατότητες στον κοινωνικό ερευνητή. Στην εθνογραφική, για παράδειγμα, έρευνα, εικόνες και μαγνητοσκοπήσεις μιας φυλής επιτρέπουν στον αναγνώστη να δει εικόνες και ν' ακούσει ήχους μιας άλλης κουλτούρας, ενώ συγχρόνως διαβάζει την ερμηνεία του ερευνητή.

(γ') Στην περίπτωση των ερευνών πεδίου και της στατιστικής τους επεξεργασίας, είναι τώρα δυνατόν σ' ένα CD-ROM να βρίσκονται όχι μόνο τα δεδομένα που συλλέχθηκαν αλλά και οι διάφορες φάσεις της ανάλυσής τους, καθώς και τα επιμέρους αποτέλεσματα, επιτρέποντας στο μελετητή να κατανοήσει καλύτερα την επεξεργασία που έγινε αλλά και να επιχειρήσει νέες, με δική του πρωτοβουλία, επεξεργασίες.

(δ') Η νέα κουλτούρα των ηλεκτρονικών βιβλίων και των multimedia, που χαρακτηρίζονται από μια μη γραμμική λογική, έχει μια κατεύθυνση όχι πια αφηγηματική (narrative culture) αλλά περισσότερο τεχνική-πληροφοριακή, που εφοδιάζει τον αναγνώστη με νέες δυνατότητες. Έτσι ο αναγνώστης-μελετητής επιλέγει τη δική του διαδρομή για την ανάγνωση του «ηλεκτρονικού κειμένου», ταξιδεύοντας στα κομμάτια πληροφορίας που τον ενδιαφέρουν, με μια λογική περισσότερο κατεύθυνσης problem solving παρά ανάγνωσης, ενώ παράλληλα έχει τη δυνατότητα να επεμβαίνει ο ίδιος, συμπληρώνοντας το «κείμενο» με τις παρατηρήσεις του, συμμετέχοντας έτσι κι αυτός, κατά κάποιο τρόπο, στη συγγραφή. Ενώ στην αφηγηματική γνώση ο συγγραφέας είν' εκείνος που δίνει δομή στη σκέψη του αναγνώστη, στην τεχνική κουλτούρα γνώσης η δομή του κειμένου φαίνεται να είναι λιγότερο σημαντική από το κείμενο που δημιουργούν οι ερωτήσεις του αναγνώστη. Από την άλλη, αν η αφηγηματική γνώση απαιτεί να «παραδοθεί» κανείς στο συγγραφέα, η μη γραμμική λογική τύπου hypertext προτείνεται σαν ένα εργαλείο «ισοτιμίας», μια και είναι εργαλείο, ταυτόχρονα, για το συγγραφέα και για τον αναγνώστη. Πομπός και δέκτης δεν διαχωρίζονται σ' ένα αμφίδρομο σύστημα επικοινωνίας. Είναι προφανές ότι εδώ ανοίγει ένα νέο κεφάλαιο κοινωνιολογικού προβληματισμού.

Ελπίζουμε ότι η περαιτέρω θεωρητική μελέτη αλλά και η χρήση και εξοικείωση των ερευνητών με τις Νέες Τεχνολογίες θα συμβάλουν σε μια δημιουργική ένταξη των Ελλήνων κοινωνιολόγων στο νέο τεχνολογικό μεθοδολογικό περιβάλλον και τη συμμετοχή τους στον αντίστοιχο προβληματισμό.

ΒΙΒΛΙΟΓΡΑΦΙΑ ΕΠΙΜΕΤΡΟΥ

- ΑΘΑΝΑΣΙΑΔΗΣ, Η. (1995), *Παραγοντική Ανάλυση Αντιστοιχιών και Ιεραρχική Ταξινόμηση*. Εκδόσεις Νέων Τεχνολογιών, Αθήνα.
- BLANK, G., MCGARTNEY, J.L., & BRENT, E. (επιμ.) (1989), *New Technology in Sociology*. Transaction Publishers, New Jersey.
- LEBART, L., SALEM, A. (1988), *Analyse statistique des données textuelles*. Dunod, Paris.
- LEE, R.M. (επιμ.) (1995), *Information Technology for the Social Scientist*. UCL Press, London.
- MEIMARIS, M. (1996), «Attitudes des professeurs grecs face aux nouvelles technologies: textes des réponses libres à un questionnaire». Στο *Les Cahiers de l'Analyse des Données*, τ. XXI, αρ. 2.