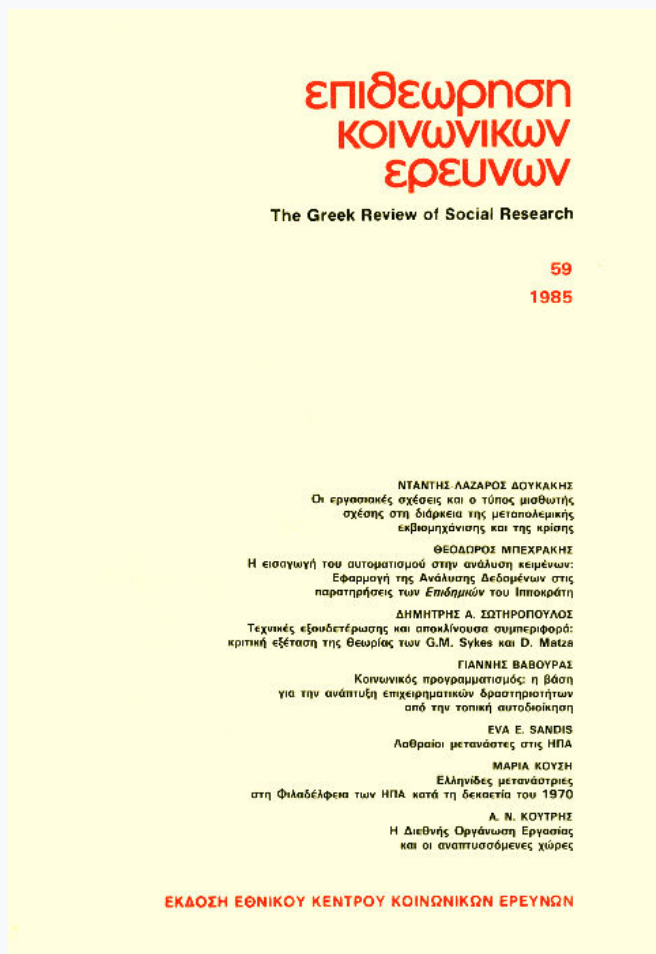


The Greek Review of Social Research

Vol 59 (1985)

59



Η εισαγωγή του αυτοματισμού στην ανάλυση κειμένων: εφαρμογή της ανάλυσης δεδομένων στις παρατηρήσεις των *Επιδημιών* του Ιπποκράτη

Θεόδωρος Μπεχράκης

doi: [10.12681/grsr.770](https://doi.org/10.12681/grsr.770)

Copyright © 1985, Θεόδωρος Μπεχράκης



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

To cite this article:

Μπεχράκης Θ. (1985). Η εισαγωγή του αυτοματισμού στην ανάλυση κειμένων: εφαρμογή της ανάλυσης δεδομένων στις παρατηρήσεις των *Επιδημιών* του Ιπποκράτη. *The Greek Review of Social Research*, 59, 41–57.
<https://doi.org/10.12681/grsr.770>

Θεόδωρος Μπεχράκης

Η ΕΙΣΑΓΩΓΗ ΤΟΥ ΑΥΤΟΜΑΤΙΣΜΟΥ
ΣΤΗΝ ΑΝΑΛΥΣΗ ΚΕΙΜΕΝΩΝ

Εφαρμογή της Ανάλυσης Δεδομένων στις
παρατηρήσεις των *Επιδημιών* του Ιπποκράτη

I. LA MACHINA SAPIENS: ΜΥΘΟΣ ΚΑΙ ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ

Απ' τη στιγμή της κατασκευής του πρώτου ηλεκτρονικού υπολογιστή γύρω στο 1944 από τον Αμερικανό Αϊκεν, οι επιστήμονες της εποχής έθεσαν πολύ σοβαρά το ερώτημα αν κάποια μέρα οι μηχανές αυτές θα μπορούσαν να σκέπτονται. Η ιδέα που προϋπάρχει του εργαλείου του Αϊκεν είναι η κατασκευή μηχανών κατ' εικόνα του ανθρώπου, ιδέα η οποία ενεργοποιεί παγκόσμια εδώ και σαράντα χρόνια τους ερευνητές του τομέα της τεχνητής νοημοσύνης.

Η γενική υπόθεση που επικρατεί και καθορίζει μέχρι σήμερα την πορεία των ερευνών είναι ότι η σκέψη λειτουργεί σύμφωνα με αφηρημένους κανόνες και λογικές διαδικασίες και άρα είναι δυνατόν να υπάρξουν προγράμματα που να την αναπαράγουν.

Η υπόθεση αυτή, που έχει τις ρίζες της στις φιλοσοφικές θέσεις του Πλάτωνα, παραμένει αναπόδεικτη, της οφείλουμε πάντως σε μεγάλο βαθμό την τεράστια ανάπτυξη της πληροφορικής.

Σε ποιο στάδιο βρίσκεται σήμερα η πραγματοποίηση της ιδέας της τεχνητής νοημοσύνης; Για να απαντήσουμε θα προσπαθήσουμε να δώσουμε μια συνοπτική εικόνα της κατάστασης της πληροφορικής το 1984. Στην περιοχή του hardware υπάρχει πλήρης άθηση: Υπολογιστές με τεράστια και αναπτυσσόμενη δυνατότητα εφαρμογών, τσιπς όλο και πιο ελαχιστοποιημένα, κυκλώματα όλο και πιο ολοκληρωμένα. Στην περιοχή του software, δηλαδή της δημιουργίας προγραμμάτων, η κατάσταση είναι αμφισβητούμενη. Στον τομέα της τεχνητής

* Ερευνητής - στατιστικός ΕΚΚΕ.

νοημοσύνης το ανώτερο σημείο είναι τα «συστήματα εμπειρογνώμονες» (systems experts). Τα συστήματα αυτά, θεωρητικά, είναι ικανά να λειτουργούν ακολουθώντας μια διαδικασία που μπορεί να συγκριθεί μ' αυτήν ενός ειδικού (γιατρού, μηχανικού, γεωλόγου...) όταν επιλύει ένα πρόβλημα της ειδικότητάς του.

Ο κατασκευαστής του «συστήματος εμπειρογνώμονα» αρχίζει με τη συλλογή και τυποποίηση των γνώσεων που χρησιμοποιεί ο ειδικός: γνώσεις θεωρητικές βέβαια, αλλά κυρίως γνώσεις που έχουν αποκτηθεί από την εμπειρία. Ένα πρώτο πρόβλημα που γεννιέται άμεσα με τον ορισμό των «συστημάτων εμπειρογνώμωνων» είναι ότι ο ειδικός συχνά είναι ανίκανος να τυποποιήσει μια γνώση την οποία αυτός αισθάνεται ως προφανή. Άλλο ζήτημα είναι ο περιορισμός του πεδίου έρευνας σε μια μικρή περιοχή γνώσης. Μια προοδευτική επέκταση του πεδίου αυτού δεν είναι καθόλου απλή υπόθεση. Δεν είναι δηλαδή δυνατό να επεκτείνουμε ένα σύστημα με εκατό κανόνες σε ένα σύστημα με χίλιους κανόνες και εν συνεχεία σε ένα σύστημα με δέκα χιλιάδες κανόνες, για τον απλό λόγο ότι ο χρόνος των υπολογισμών γίνεται υπερβολικός.

Εξάλλου ένα πρόγραμμα που προβλέπει την αμετάβλητη σειρά μιας διαδικασίας, σημαίνει ότι το πρόβλημα πρέπει να επιλυθεί εντελώς πριν το εμπιστευθούμε στη μηχανή.

Έτσι αν και τα συστήματα αυτά θεωρούνται χρήσιμα και υπάρχει μια αναπτυσσόμενη αγορά γύρω απ' αυτά, πολλοί είναι αυτοί οι οποίοι αμφισβητούν τη νοημοσύνη τους και την ευρετική τους ικανότητα.

Τη θέση αυτή μεταξύ άλλων υποστηρίζει και ο Hubert Dreyfus, φιλόσοφος της τεχνητής νοημοσύνης του Πανεπιστημίου του Barkley, στο βιβλίο του που πρόσφατα κυκλοφόρησε με τίτλο *What computers can't do*. Με τις υποθέσεις που έχουμε ξεκινήσει, υποστηρίζει ο Dreyfus, δεν θα καταφέρουμε ποτέ να κατασκευάσουμε έξυπνες μηχανές. Και συμπληρώνει ότι για να προχωρήσει η έρευνα της τεχνητής νοημοσύνης είμαστε αναγκασμένοι ν' αλλάξουμε ριζοσπαστικά άξονα και θεωρία.

Ποια είναι η σχέση της Ανάλυσης Δεδομένων με τα θέματα τεχνητής νοημοσύνης;

Αν και η μεθοδολογία που ακολουθείται είναι εντελώς διαφορετική από αυτήν των «συστημάτων εμπειρογνώμωνων», τα αποτελέσματα που μας δίνει —πιστεύουμε ότι θα διαφανεί και στη συνέχεια του άρθρου— είναι ενθαρρυντικά ως προς την ικανότητα της μεθόδου να αναγνωρίζει μορφές, πράγμα το οποίο αποτελεί έναν από τους βασικούς στόχους των προγραμμάτων της τεχνητής νοημοσύνης.

Η μεθοδολογία της Ανάλυσης Δεδομένων συνίσταται στη συγκέντρωση ενός εκτεταμένου πλήθους πληροφοριών που αφορούν το εξεταζόμενο φαινόμενο —το οποίο μπορεί να είναι φυσικό, κοινωνικό ή γλωσσολογικό...— και την οργάνωση του συνόλου των πληροφοριών σε ένα σύστημα με τη βοήθεια αλγε-

βρικών μεθόδων γενικά αποδεκτών, οι οποίες δεν σχετίζονται με οποιαδήποτε εκ των προτέρων θεωρία επί του φαινομένου. Από την οργάνωση αυτή των πληροφοριών προκύπτουν οι πραγματικές διαστάσεις του φαινομένου, οι κυριότερες συνιστώσες του, πράγμα το οποίο κάνει δυνατό τον εντοπισμό και την αναγνώριση των μορφών, τις οποίες είναι δύσκολο να παρατηρήσουμε στο εκτεταμένο και ακατέργαστο αρχικό σύνολο πληροφοριών.

Διευκρινίζουμε ότι μια μορφή μπορεί να είναι διάφορα πράγματα όπως ένας τύπος κοινωνικής συμπεριφοράς σε μια κοινωνιολογική έρευνα ή μια σημασιολογική οντότητα στην ανάλυση ενός corpus κειμένων κ.λπ.

Στην εισαγωγή του θεμελιώδους έργου του (*L'analyse des données*, T. 1) και αναφερόμενος στη σχέση της Ανάλυσης Δεδομένων με την αυτόματη αναγνώριση μορφών ο καθηγητής Benzecri υποστηρίζει: «Δεν θα επιλύσουμε με τις μηχανές τέτοια προβλήματα (βλέπε προβλήματα αναγνώρισης μορφών) παρά μόνο με αλγόριθμους ταξινόμησης και αναγωγής του αριθμού των διαστάσεων, αλγόριθμοι οι οποίοι, ξεκινώντας από ένα σύνολο παρατηρήσεων ή σχέσεων, να αναγνωρίζουν τις δομικά σημαντικότερες ιδιότητες και τις διαστάσεις ως προς τις οποίες κατανέμονται τα στοιχεία του υπό μελέτη συνόλου».

Από τη στιγμή που διαθέτουμε τέτοιους αλγόριθμους ταξινόμησης και αναγωγής των διαστάσεων, μπορούμε να τους εφαρμόσουμε σε προβλήματα που άπτονται της νοημοσύνης, όπως για παράδειγμα την αναγνώριση του θέματος ενός κειμένου, το στυλ ή το συγγραφέα ενός ποιήματος, την αυθεντικότητα ενός κειμένου, τα κυριότερα σημασιολογικά του στοιχεία κ.λπ.

Πρέπει να επισημάνουμε τη διάκριση μεταξύ των ερευνών όπου μια ταξινόμηση είναι δεδομένη a priori και οφείλουμε να την αναπαράγουμε μηχανικά και τις έρευνες —μεταξύ των οποίων και αυτή εδώ— όπου το μόνο a priori δεδομένο περιλαμβάνει πολυάριθμες παρατηρήσεις.

Αυτή είναι και η σπουδαιότητα των πολυδιάστατων μεθόδων της Ανάλυσης Δεδομένων, η οποία μπορεί αφ' ενός να μας απαλλάξει από απόψεις απλοϊκές ή προπαρασκευασμένες και αφ' ετέρου μπορεί να αποτελέσει ένα βήμα προς την κατεύθυνση των έξυπνων μηχανών.

2. ΑΝΑΛΥΣΗ ΓΛΩΣΣΟΛΟΓΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Η ανάλυση δεδομένων, βασισμένη στη χρήση του ηλεκτρονικού υπολογιστή, είναι μια νέα μεθοδολογία που η στατιστική προσφέρει στις επιστήμες και ειδικότερα στις επιστήμες του ανθρώπου.¹

1. J.P. Benzecri, *L'analyse des données: l'analyse des correspondances*, τ. 2, σ. 3, Dunod, Paris 1980.

Το αντικείμενο της ανάλυσης δεδομένων είναι να παρουσιάσει με γεωμετρική μορφή το σύστημα των σχέσεων που υπάρχουν μεταξύ των στοιχείων δύο συνόλων I και J ή ακόμη του ενός μόνο από αυτά. Για το σκοπό αυτό αποκαλούνται πίνακες αντιστοιχιών (*tableau des correspondances*). Η μαθηματική επεξεργασία των πινάκων αυτών σκοπό έχει για μεν την παραγοντική ανάλυση (*analyse factorielle*) την εξαγωγή των κυριότερων παραγόντων (*facteurs*) του φαινομένου που εξετάζουμε, και την προβολή στα επίπεδα που οι παράγοντες αυτοί ορίζουν των στοιχείων των δύο συνόλων, για δε την αυτόματη ταξινόμηση (*classification automatique*) τη δημιουργία ομογενών ομάδων στο εσωτερικό των συνόλων I ή J.

Είναι ενδιαφέρουσα η παρατήρηση ότι η απόσταση του X^2 —που ονομάζεται επίσης κατανομιστική απόσταση— εισήχθη στην παραγοντική ανάλυση και κατόπιν στην αυτόματη ταξινόμηση μετά από θεωρήσεις γλωσσολογικές.

Γύρω στα 1960, στα πλαίσια των συναντήσεων της γαλλικής εταιρείας για την αυτόματη μετάφραση (A.T.A.L.A.), ο καθηγητής Benzecri πρότεινε την παραγοντική ανάλυση των αντιστοιχιών (*analyse factorielle des correspondances*) σαν μια επαγωγική μέθοδο ανάλυσης γλωσσολογικών δεδομένων.

Την εποχή αυτή στο χώρο της μαθηματικής γλωσσολογίας κυριαρχούσε το όνομα του N. Chomsky, με το έργο του *Syntactic structures*. Μεταξύ άλλων θέσεων ο N. Chomsky υποστήριξε ότι δεν μπορεί να υπάρξει μια συστηματική διαδικασία για τον καθορισμό της γραμματικής μιας γλώσσας, ή πιο γενικά των γλωσσολογικών δομών της, που να ξεκινά από ένα σύνολο δεδομένων, όπως είναι για παράδειγμα μια συλλογή κειμένων, που συνήθως αποκαλείται *corpus*. Για τον N. Chomsky, λοιπόν, η γλωσσολογία δεν μπορεί να είναι επαγωγική. Αντίθετα οφείλει να είναι απαγωγική, υπό την έννοια ότι ξεκινώντας από αξιώματα, παράγει μοντέλα των φυσικών γλωσσών (ακολουθεί δηλαδή την πορεία από τη θεωρία προς τα πράγματα).

Ενώπιον στη θέση αυτή ο J. P. Benzecri πρότεινε, ελλείψει ενός αλγορίθμου για τη δημιουργία ενός συντακτικού από ένα κείμενο 10.000 σελίδων μιας γλώσσας, μια αποτελεσματική επαγωγική μέθοδο για την επεξεργασία πινάκων δεδομένων που μπορούσαν να συγκεντρωθούν άμεσα. Έκτοτε μία σειρά από ενδιαφέρουσες εφαρμογές της μεθόδου αυτής σε γλωσσολογικά δεδομένα² σε συνδυασμό με ποικίλες τεχνικές εξελίξεις δεν σταμάτησαν να δίνουν ενθαρρυντικά αποτελέσματα. Θα παρουσιάσουμε ενδεικτικά μερικές από τις εφαρμογές αυτές, στις οποίες ο Πίνακας που αναλύεται διασταυρώνει ένα σύνολο I κειμένων με ένα σύνολο J λέξεων.

Μπορούμε να βρούμε στους νόμους I και II του *Traite sur l' A. des D.* δύο

2. Υπάρχει επίσης ένα πλήθος εφαρμογών της μεθόδου σε άλλες επιστήμες πλην της γλωσσολογίας.

απ' τα πρώτα παραδείγματα ανάλυσης κειμένων. Το ένα είναι «Οι ρόλοι των πρωταγωνιστών στη *Φαίδρα* του Ρακίνα» (Τ.ΙΑ Νο 2 § 3), όπου μελετάται το λεξιλόγιο που χρησιμοποιούν τα κύρια πρόσωπα του έργου. Το άλλο είναι «Οι πολιτικές διακηρύξεις των βουλευτών της Γαλλίας το 1881» (Τ. ΙΙ C Νο 2), στο οποίο μελετάται η διαφορά λεξιλογίου μεταξύ αριστεράς - δεξιάς, όπως προκύπτει από κείμενα της εποχής.

Κατόπιν εκδόθηκε ένας τόμος της σειράς *Pratique de l'analyse des donnees* αφιερωμένος αποκλειστικά σε μελέτες γλωσσολογικές και λεξικολογικές. Μερικοί τίτλοι από τις μελέτες που μπορούμε να βρούμε είναι: Το βιβλίο του Ησαΐα και η κριτική ανάλυση της ομοιογένειας του κειμένου, το λεξιλόγιο της δουλειάς στους αττικούς ρήτορες, το λεξιλόγιο των Ευαγγελίων και η ανάλυση των ομοιοτήτων των κεφαλαίων του Ιωάννη, τα πολιτικά κείμενα της Γαλλίας του 1793, η τυπολογία των κύριων τίτλων του βρετανικού τύπου...

Επίσης τακτικά δημοσιεύονται στα *Cahiers de l'A. des D.* νέες αναφορές εκ των οποίων προσφάτως μία με τίτλο «Analyse des repetitions et indexation automatique des documents» (C.A.D 1984, No. 2), που αναφέρεται στην αυτόματη δημιουργία mots-cle για την περιγραφή ενός συνόλου ντοκουμέντων και με δυνατότητα εφαρμογής σε βάσεις δεδομένων.

Τέλος, ένας άλλος τομέας όπου η εφαρμογή της μεθόδου είναι εκτεταμένη, είναι στην ανάλυση των ανοιχτών ερωτήσεων των ερωτηματολογίων.³ Αλλά στο θέμα αυτό θα αναφερθούμε και παρακάτω.

3. Ο ΑΥΤΟΜΑΤΙΣΜΟΣ ΣΤΗΝ ΚΑΤΑΓΡΑΦΗ ΛΕΞΙΚΩΝ ΤΥΠΩΝ

Είναι γενικά αποδεκτό ότι η τελική μορφή των δεδομένων που αναλύονται με τις τεχνικές της Α.Δ. είναι ο πίνακας I X J, καλούμενος και πίνακας αντιστοιχιών. Υπάρχει όμως το πρόβλημα ποιος πίνακας πρέπει να αναλυθεί. Αν υποθέσουμε ότι το σύνολο I είναι ένα σύνολο κειμένων, μένει να διαλέξουμε το J και να ορίσουμε τι σημαίνουν τα K (I, J).

Είναι συνηθισμένο να παίρνουμε σαν J ένα σύνολο λέξεων. Αλλά ακόμη και όταν ορίσουμε το σύνολο αυτό, πολλά ερωτήματα τίθενται: θα πάρουμε τις λέξεις όπως είναι στο λεξικό: αγαπώ, ωραίος... Τότε θα πρέπει με κάποιο τρόπο να προσαρτήσουμε το αγαπούμε στο αγαπώ και το ωραία στο ωραίος. Πολλές από τις «λεπτομέρειες» αυτές δεν διακρίνονται εύκολα από ένα αυτόματο. Κανέ-

3. Το CREDOC (Centre de Recherche pour l' Etude et l' Observation des Conditions de vie) εφαρμόζει τη μεθοδολογία αυτή από το 1978.

να πρόγραμμα μορφολογικής ανάλυσης ή εύρεσης ομωνυμίων δεν έχει ακόμα καταφέρει να διακρίνει χωρίς λάθη.

Αυτό που μπορούμε να κάνουμε είναι να εφαρμόσουμε την τεχνική που πρότεινε ο L. Lebart (C.A.D., τ. VI, Νο 2 (ANA.LEX) 1981), με την οποία απαριθμούμε απλώς λεξικούς τύπους, δηλαδή σειρές γραμμάτων που βρίσκονται μεταξύ δύο λευκών. Η τεχνική αυτή εφαρμόστηκε και στην ανάλυση των παρατηρήσεων του Ιπποκράτη καθώς επίσης και σε μια σειρά άλλων ερευνών και έδωσε θετικά αποτελέσματα. Εξάλλου ας σημειωθεί ότι η ανάλυση δεδομένων σήμερα κάνει δυνατό τον επανασχηματισμό μη κωδικοποιημένων πληροφοριών, όπως είναι για παράδειγμα η αναπαραγωγή φράσεων του κειμένου (Sprad version 84).

Κλείνοντας το κεφάλαιο αυτό θα θέλαμε να αναφερθούμε σε μια ιδέα, την οποία μας υποβάλλει ένα γεγονός. Το γεγονός της εξάπλωσης των προσωπικών υπολογιστών και γενικότερα των μικρών (micro) και των αντίστοιχων προγραμμάτων των σχετικών με την επεξεργασία κειμένων. Τα προγράμματα αυτά μετατρέπουν τον υπολογιστή σε μια καλή γραφομηχανή που έχει τη δυνατότητα να αποθηκεύει το κείμενο στη μνήμη. Όμως κάθε στατιστική επεξεργασία πραγματοποιείται σε μια παρόμοια εγγραφή. Έτσι είναι νόμιμο να θεωρήσουμε ότι το αντικείμενο των γλωσσολογικών ερευνών είναι ένα σύνολο κειμένων που έχει ήδη εισαχθεί. Τότε στα κείμενα αυτά μπορούμε να πραγματοποιήσουμε μια «ατελή» ανάλυση λεξικών τύπων, αφήνοντας τις όποιες αμφιβολίες να λυθούν από τον ερευνητή με το σύστημα των ερωτήσεων στο ηλεκτρολόγιο και των απαντήσεων στην οθόνη.

4. ΑΥΤΟΜΑΤΗ ΑΝΑΛΥΣΗ ΤΟΥ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ ΕΝΟΣ ΚΕΙΜΕΝΟΥ ΤΟΥ ΙΠΠΟΚΡΑΤΗ

Στο σημείο αυτό θα παρουσιάσουμε τη μεθοδολογία και τα αποτελέσματα της ανάλυσης του σημασιολογικού περιεχομένου ενός κειμένου του Ιπποκράτη. Το κείμενο αυτό αποτελείται από σαράντα δύο παρατηρήσεις.⁴

4.1. Το κείμενο και η προβληματική για το περιεχόμενό του

Τι είναι οι παρατηρήσεις αυτές; Κάθε παρατήρηση είναι η περιγραφή της εξέλιξης της κατάστασης ενός ασθενούς. Θα λέγαμε ότι πρόκειται για τα πρώτα ιστο-

4. Οι 42 παρατηρήσεις περιέχονται στα βιβλία *Επιδημιών Ι* και *Επιδημιών ΙΙΙ*. Τα βιβλία αυτά θεωρούνται από την εποχή του Γαληνού από τα πλέον αυθεντικά.

ρικά ασθeneίας, όπου δίνεται ιδιαίτερη προσοχή στα εξωτερικά χαρακτηριστικά που παρουσιάζει ο ασθενής, με αποτέλεσμα να έχουμε μάλλον μια εικόνα του εξεταζόμενου ανθρώπου παρά μια στείρα καταγραφή κλινικών σημείων.

Δύο από τις παρατηρήσεις των *Επιδημιών*.

Πρόκειται για την πρώτη και τρίτη παρατήρηση των *Επιδημιών Ι*, όπως τις βρίσκουμε στο έργο του E. Littré. Στην πρώτη περίπτωση ο ασθενής πέθανε, ενώ στην τρίτη θεραπεύθηκε. Δύσκολα μπορούμε να διακρίνουμε μια διαφορά του λεξιλογίου.

Μεταξύ των σημείων που εξετάζονται είναι η κατάσταση του προσώπου, η θέση του στο κρεβάτι, οι κινήσεις των χεριών. Ο συγγραφέας χρησιμοποιεί το λόγο για την απλή καταγραφή γεγονότων: «άκρεια πάντοθεν ψυχρά». Είναι μια πληροφορία που έχει ενδιαφέρον βασικά ως προς το λεξιλόγιο. Πρόκειται δηλαδή για ένα περιγραφικό κείμενο στο οποίο η πληροφορία βρίσκεται συνήθως στην καθαρή της μορφή.

Στο σύνολο των 42 παρατηρήσεων, υπάρχουν είκοσι πέντε περιπτώσεις όπου η ασθένεια κατέληξε στο θάνατο. Αυτή η αναλογία περιπτώσεων θανάτου είναι εντυπωσιακή. Κατά την αρχαιότητα, ο Ασκληπιός αποκαλούσε τις *Επιδημίες* πραγματεία θανάτου. Αλλά δεν μπορούμε να δούμε την αναλογία αυτή των αριθμών σαν έναν στατιστικό ισολογισμό της θεραπευτικής πρακτικής του Ιπποκράτη. Είναι προφανές ότι οι 42 αυτές παρατηρήσεις είναι μια επιλογή μεταξύ πολλών άλλων. Με ποια λογική έγινε τότε η συγκεκριμένη αυτή επιλογή;

Αν ανατρέξουμε στο *Προγνωστικό*⁵ και παρατηρήσουμε τους δογματικούς κανόνες, που ο ίδιος ο Ιπποκράτης είχε εκεί θέσει, θα πειστούμε ότι κινήθηκε από την επιθυμία να φωτίσει το βιβλίο αυτό και ότι χρησιμοποίησε τις παρατηρήσεις που του φάνηκαν οι πλέον κατάλληλες.

Η σκοπιμότητα των κανόνων αυτών δεν είναι η διάγνωση της γενικής κατάστασης του ασθενούς. Όμως ο σύγχρονος αναγνώστης των «παρατηρήσεων» δεν μπορεί να αποφύγει την προσπάθεια να ταυτοποιήσει τις ασθένειες που περιγράφονται με σημερινές. Εξάλλου ο ίδιος ο Ιπποκράτης ονομάζει τη φθίση, τη φρενίτιδα και τον καύσο.

Ο E. Littré, ο οποίος παραμένει ο βασικότερος σχολιαστής του Ιπποκράτη, είχε θέσει το ερώτημα⁶ με ποιες ασθένειες μπορούμε να συσχετίσουμε τις ιστορίες των ασθενών που περιγράφονται στις *Επιδημίες*. Σε μια επανέκδοση (Jouval et Bourdeaux, Παρίσι 1932) της μετάφρασης του Littré, ο καθηγητής Roger κάνει το παρακάτω σχόλιο για τις παρατηρήσεις των *Επιδημιών*: «Βρίσκουμε την

5. Ένα άλλο κείμενο από τα αυθεντικά, γραμμένο λίγο πριν από τις *Επιδημίες*.

6. E. Littré. Hippocrate: *Oeuvres completes*. 1839 - 1861 Τόμος II, σελ. 538.

13. ἌΡΡΩΣΤΟΙ ΤΕΣΣΑΡΕΣΚΑΙΔΕΚΑ.

Ἄρρωστος πρῶτος.

Φιλίσκος ἢ φεικὸς πρὸς τὸ τεῦχος, ὁ κατακλιθῆναι ἢ τῆ
 ἢ πρῶτη πυρετός δέξυς ἔρωσεν. ἢ ἔξ νόκτα, ἐπιπνῶς.
 ἢ Δευτέρῃ, πάντα ἢ παρωλύθῃ. ἢ δὲ ἀπὸ 14 δὲ ἀπὸ 15 κλυ-
 σματίου καλῶς διήλθε νόκτα δὲ ἡσυχίῃ. 16 Τρίτῃ, πρῶτῃ, καὶ
 μῆτρῃ 17 μέσον 18 ἡμέρας, εὖδως γενέσθαι 19 ἀπέρστος πρὸς δει-
 λην δὲ πυρετός δέξυς, 20 μετὰ ἰσθμῶς διηλώδης γλώσσα 21 ἐπι-
 κηράνεται 22 μέλινα 23 ὄργισεν 24 νόκτα 25 δυσπόρφως οὐκ ἐκοι-
 μήθη πάντα παρέφρουσε. 26 Τετάρτῃ, πάντα 27 παρωλύθῃ.
 ὄργα μέλινα 28 νόκτα 29 εὐφοροτέρῃν 30 ὄργα 31 εὐγρούτερα.
 32 Πέμπτῃ, πρὸς 33 μέσον 34 ἡμέρας, 35 σμικρὸν ἀπὸ βίνων 36 ἐτα-
 ξεν 37 ἀφῆρτον. ὄργα δὲ ποικίλα, ἔχοντα 38 ἐναυορῆματα 39 στρω-
 γίλα, γονοσίαι, 40 διασπαρμένα, 41 οὐχ 42 ἔρωτο. 4 προσημνίῳ
 5 δὲ βέλων, 6 φουδῶδες σμικρὰ 7 διήλθεν νόκτα 8 ἐπιπνῶς
 9 ὄργα 10 σμικρὸς λόγος, 11 λήρος 12 ἀρεα 13 πάντοθεν ψυχρὸς,
 καὶ 14 οὐκ ἔτι 15 ἐπιναθερμαινόμενα. ὄργα 16 μέλινα. ἐκοι-
 μήθη σμικρὰ πρὸς ἡμέραν. ἄφρωνος ἔρωσε 17 ψυχρὸν, 18 ἀρεα
 19 πελιῶν πρὸς 20 δὲ μέσον 21 ἡμέρας, 22 ἔκτατος 23 ἀπέθανεν.
 24 Τουτέῳ 25 πινόμενα 26 διὰ τέλους, ὅσπερ 27 ἀνεκαλωμένω 28 ἀρατῶν,
 μέγα. 29 σπλῆν ἐπὶ ἔρωθι 30 περιερείει κυρτώματι. ἔρωτες ψυχροὶ
 31 διὰ τέλους 32 οἱ παρωλύοι 33 ἐν ἀρτίῳ.

Ἄρρωστος τρίτος.

Ἡρώωντι πυρετός δέξυς ἀπὸ 31 κοιλίης θλίγα, 32 τεινε-
 σμῶδα 33 κατ' ἀργέας 34 μετὰ δὲ, 35 λεπτὰ διήει γολώδης, ὑπό-
 σφυγα 36 ὄργα οὐκ 37 ἐνώσεν. ὄργα 38 μέλινα, λεπτὰ.
 39 Πέμπτῃ, πρῶτῃ κόφωσις 40 παρωλύθῃ πάντα σπλῆν ἐπὶ ἄρ-
 θῃ. ὄργονόριου ζώντασις ἀπὸ 41 κοιλίης 42 θλίγα, 43 μέλινα
 44 διήλθεν. παρέρωσεν. 45 ἔρωτο, 46 ἐλήγει. 47 ἔξ νόκτα ἰσθμῶς
 ψύχῃ 48 λήρος 49 παρέμεινεν. 50 Ἐξόμῃ, 51 περιέμικτο. διηλώδης
 παρέφρουσεν νόκτα κατενόησι: κατεκοιμήθη. 3 Ὀρθῶν 4 ἐπύ-
 ρασεν 5 σπλῆν ἡμεῖοτο: κατενόησι πάντα. 6 ἤγχεσε ἢ τὸ πρῶτον
 κατὰ βουδῶνα, σπλῆνός κατ' ἔξιν ἔπειτα 8 ἐπῶνει 9 ἔξ ἀμφοτέ-
 ρας κνήμας 10 νόκτα 11 εὐφόρος ὄργα 12 εὐγρούτερα. ὑπό-
 σταςιν 13 εἶγξ 14 σμικρὸν, 15 ἔρωτο 16 ἔρωσεν. ἐπὶ ἔρωθι. 17 διέ-
 λπεν. 18 Πέμπτῃ 19 διήστρωσεν. αὐτίκα 20 δὲ σπλῆν ἐπὶ ἄρθῃ.
 πυρετός δέξυς κόφωσις πάλιν. Μετὰ δὲ τὴν ὑποτροπὴν 21 τρίτῃ,
 σπλῆν ἡμεῖοτο: κόφωσις 22 ἦσσαν σκέλεα 23 ἐποδῶνος 24 νόκτα
 ἔρωσεν. ἐπὶ ἔρωθι 25 ἐπιτακτικῶδες. οὐδὲ παρέφρουσεν 26 ἐπὶ τῇ
 ὑποτροπῇ.

Διὸ ἀπὸ τῆς παρατηρήσεως τῶν Ἐπιδημιῶν.

Πρόκειται γὰρ τὴν πρώτην καὶ τρίτην παρατήρησιν τῶν Ἐπιδημιῶν I, ὅπως τὶς βρισκόμενος ἐν τῷ ἔργῳ τοῦ E. Littre. Στὴν πρώτην οὐ σαφηνεῖ ἀπέθανε, ἐνὶ τῇ τρίτῃ εβραευσθῆναι. Δυσκόλια μποροῦμε νὰ διακρίνομε μὴ διαφορὰ τοῦ λέξιολογίου.

περιγραφή κλινικών τύπων που μας είναι ήδη γνωστοί: λοιμώξεις με νευρικά συμπτώματα, καύσους, φρενίτιδες, λήθαργο και μια σειρά από ασθένειες με κύριο χαρακτηριστικό τον πυρετό, τις οποίες προσπαθήσαμε να ταυτίσουμε με σημερινές αρρώστιες. Η προσπάθεια αυτή αποβαίνει άσκοπη, γιατί την εποχή εκείνη η νοσογραφία μόλις ξεκινούσε να υπάρχει και όλες οι λοιμώξεις ήταν συγκεχυμένες».

Έχοντας υπόψη μας όλη αυτή την προβληματική γύρω από τις «παρατηρήσεις» των *Επιδημιών*, θελήσαμε να δούμε τι προκύπτει από τη μελέτη του λεξιλογίου που χρησιμοποιείται. Θεωρούμε ότι η μελέτη του λεξιλογίου στη συγκεκριμένη περίπτωση είναι καθοριστική για μια απόφαση του τι περιγράφεται στο κείμενο, εφόσον μια ασθένεια, ή γενικότερα μια κατάσταση, είναι ορισμένη σ' ένα κείμενο από τη στιγμή που ένα λεξιλόγιο τη σημαίνει.

Η χρήση των προγραμμάτων της «ανάλυσης δεδομένων» υπήρξε ιδιαίτερα διαφωτιστική για τη μελέτη του συγκεκριμένου προβλήματος, όχι γιατί δόθηκε μια καινούργια απάντηση, αλλά γιατί τεκμηρίωσε μια θέση ήδη γνωστή. Η επεξεργασία του κειμένου με την τεχνική αυτή συνίσταται στην απόσπαση της πληροφορίας που εμπεριέχεται, συμπυκνωμένη στα υποβλητικότερα στοιχεία της, τα οποία είναι ταυτόχρονα και τα πλέον σημαντικά.

4.2 Το περιεχόμενο της ανάλυσης

Το ελληνικό κείμενο εισήχθη στον ηλεκτρονικό υπολογιστή IRIS 80 του Πανεπιστημίου P. et M. Curie χωρίς καμία κωδικοποίηση. Ξεκινώντας από το κείμενο το πρόγραμμα ANA...EX. υπολογίζει τον πίνακα συναφείας, που διασταυρώνει τις παρατηρήσεις με τις λέξεις που εμφανίζονται σ' αυτές.

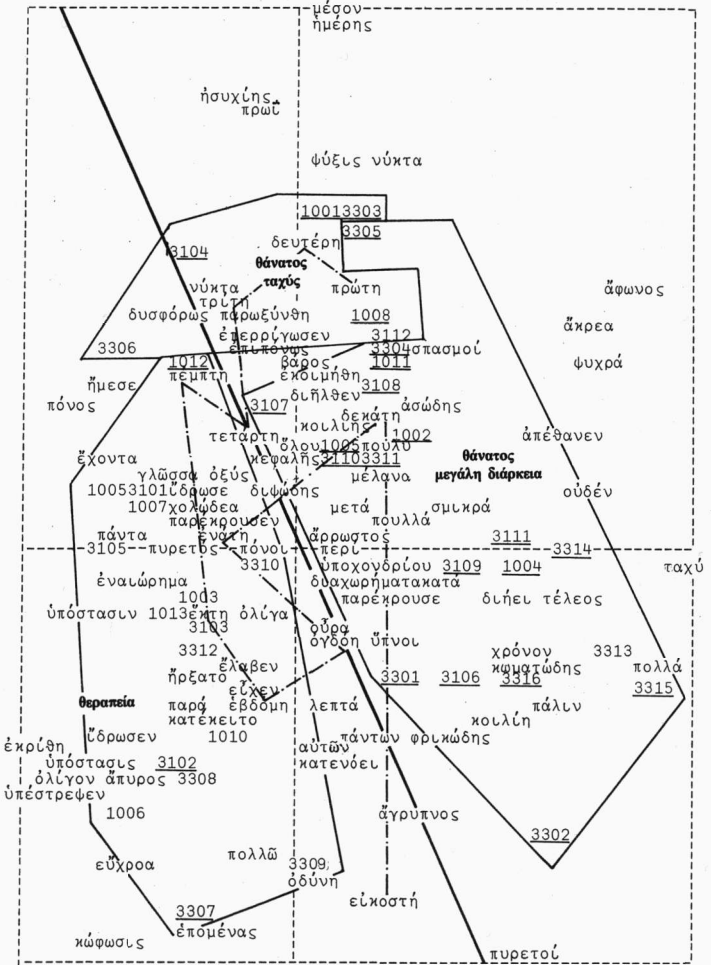
Η παραγοντική ανάλυση που πραγματοποιήθηκε στον προηγούμενο πίνακα μας δίνει τα πλάνα τα οποία αποτελούν το σημαντικότερο αποτέλεσμα της διαδικασίας και στα οποία θα αναφερθούμε αναλυτικά.

Η αυτόματη ταξινόμηση των παρατηρήσεων μας παρουσιάζει τις σημαντικότερες ομαδοποιήσεις τους. Ο εντοπισμός των ομάδων παρατηρήσεων στα πλάνα της παραγοντικής ανάλυσης αποτελεί το τελικό αποτέλεσμα της όλης διαδικασίας.

Τι παρουσιάζει ένα τέτοιο πλάνο;

Ας αναφερθούμε συγκεκριμένα στο πλάνο του Γραφήματος 1. Πρώτα απ' όλα παρατηρούμε λέξεις και αριθμούς. Οι αριθμοί δεν είναι τίποτε άλλο από την κωδικοποίηση των κειμένων μας, στη συγκεκριμένη περίπτωση των παρατηρήσεων. Για παράδειγμα, ο αριθμός 3303 είναι η τρίτη παρατήρηση του τρίτου μέρους του βιβλίου *Επιδημιών III*. Όσον αφορά τις λέξεις, είναι οι χαρακτηριστικότερες από αυτές που εμφανίζονται στο κείμενο (και στη συγκεκριμένη περίπτωση έχουν μια συχνότητα εμφάνισης άνω του 10). Όμως η κατανομή των ση-

ΓΡΑΦΗΜΑ 1: Πλάνο της ανάλυσης του κειμένου (αρ. λέξεων 97).
Υπογραμμισμένα σημεία: Παρατηρήσεις ασθενών που υπέκυψαν.
Μη υπογραμμισμένα σημεία: Παρατηρήσεις ασθενών που θεραπεύτηκαν.



μείων (λέξεων ή κειμένων) δεν είναι τυχαία πάνω στο επίπεδο. Κάθε άλλο. Θα μπορούσαμε να πούμε ότι η ομοιότητα εκφράζεται εδώ με σχέσεις απόστασης. Πιο συγκεκριμένα δύο κείμενα προσομοιάζουν περισσότερο ως προς το λεξιλογικό περιεχόμενο όσο πλησιέστερα βρίσκονται πάνω στο πλάνο.

Όσον αφορά την ερμηνεία της θέσης των λέξεων στο πλάνο, μπορούμε να πούμε ότι μια λέξη που προσεγγίζει ένα κείμενο, ή μια ομάδα κειμένων, είναι χαρακτηριστική για το κείμενο, ή την ομάδα αυτή, αφού η προσέγγιση προϋποθέτει μια υψηλή σχετική συχνότητα εμφάνισης.

Οι ερμηνείες αυτές της θέσης των σημείων (λέξεων ή κειμένων) πάνω στο επίπεδο είναι επιτρεπτές μιας και το μεθολογικό πρόβλημα που επιλύεται από την παραγοντική ανάλυση των αντιστοιχιών είναι: η ταυτόχρονη γραφική παρουσίαση των δύο συνόλων I και J με τις εξής συνθήκες: α) η θέση του κάθε $i \in I$ στο παραγοντικό επίπεδο να πλησιάζει το κέντρο βάρους των στοιχείων του J· α) η θέση του κάθε $j \in J$ αντίστοιχα να πλησιάζει το κέντρο βάρους των στοιχείων του I.

Τι γίνεται, όμως, με τις γραμμές που υπάρχουν στο πλάνο (εννοούμε τις συνεχείς, αφού υπάρχουν και διακεκομμένες). Οι γραμμές αυτές προκύπτουν από την αυτόματη ταξινόμηση (Γράφημα 2). Είναι ο σχεδιασμός των ομάδων της ταξινόμησης πάνω στο παραγοντικό επίπεδο. Η παχιά γραμμή, η οποία χωρίζει το πλάνο στα δύο, είναι η πρώτη και βασικότερη διάκριση του συνόλου των παρατηρήσεων. Οι κλειστές γραμμές, οι οποίες δημιουργούν τα τρία πολύγωνα, είναι οι ομάδες που δίνει η ταξινόμηση στο δεύτερο επίπεδο. Στην πραγματικότητα, η μία από τις δύο πρώτες αρχικές ομάδες χωρίζεται σε δύο υποομάδες.

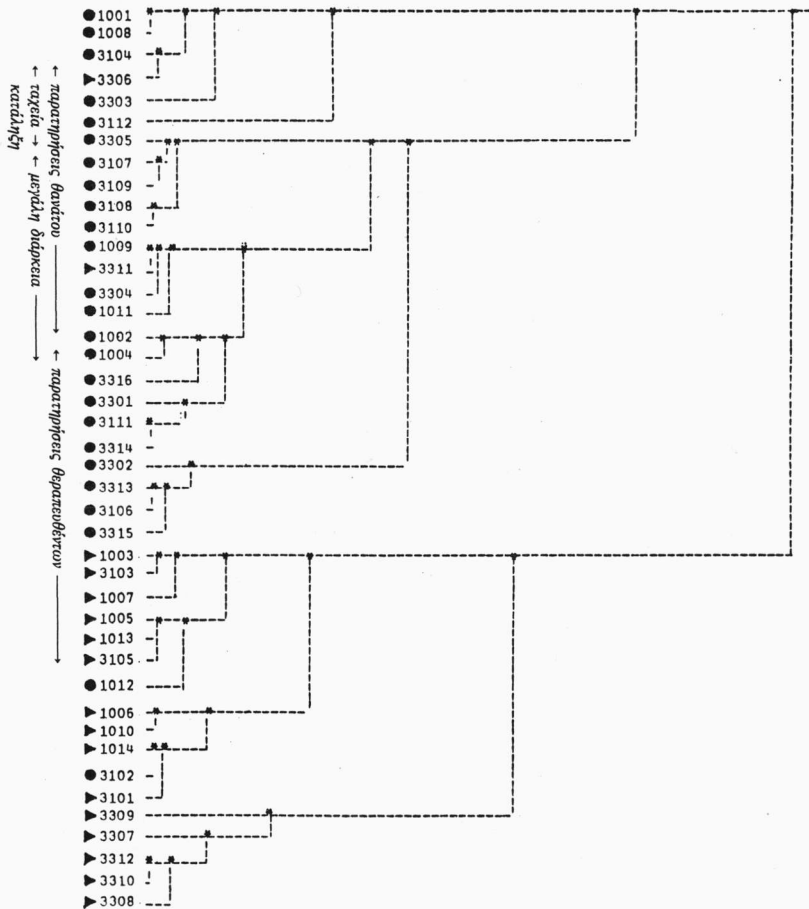
Αν τώρα εξετάσουμε τι είδους είναι οι παρατηρήσεις που περιλαμβάνονται στις επιφάνειες, όπως ορίζονται από τις παραπάνω γραμμές, θα δούμε ότι η μεν παχιά γραμμή χωρίζει τις παρατηρήσεις των θεραπευθέντων ασθενών από αυτές των αποθανόντων, τα δε πολύγωνα διαχωρίζουν το σύνολο των αποθανόντων ως προς το χρόνο της κατάληξης της ασθένειας.

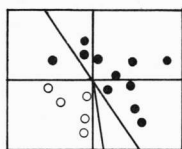
Παρατηρούμε ότι οι επιφάνειες όπως ορίζονται στο πλάνο αποτελούν αντανάκλαση σημασιολογικών διαφορών του κειμένου, για το λόγο αυτό θα μπορούσαν να χαρακτηριστούν σημασιολογικές επιφάνειες. Για να ολοκληρώσουμε την περιγραφή του πλάνου δεν θα πρέπει να παραλείψουμε την εξήγηση των διακεκομμένων γραμμών. Η μεν —●— συνδέει τις ημέρες της ασθένειας ξεκινώντας από την «πρώτη ημέρα» πάνω δεξιά για να καταλήξει στην «εικοστή ημέρα» κάτω δεξιά. Η γραμμή αυτή μπορεί να θεωρηθεί ως μια κλίμακα της διάρκειας της ασθένειας. Τέλος οι δύο κάθετες γραμμές είναι οι δύο πρώτοι άξονες που προκύπτουν απ' την παραγοντική ανάλυση.

Στην Εικόνα 3 έχουμε μια σύνοψη όλων των αναλύσεων που πραγματοποιήθηκαν. Η μόνη διαφορά είναι τα σύνολα των λέξεων που υπεισέρχονται

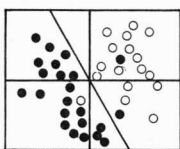
ΓΡΑΦΗΜΑ 2: Αυτόματη ταξινόμηση των παρατηρήσεων του κειμένου

- : Παρατηρήσεις ασθενών που υπέκυψαν
- ▲ : Παρατηρήσεις ασθενών που θεραπεύτηκαν

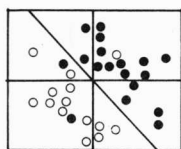




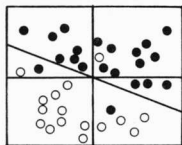
S = 2



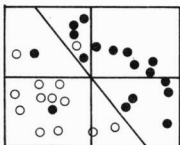
S = 3



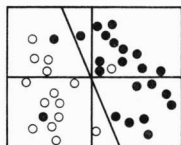
S = 4



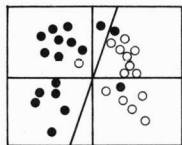
S = 6



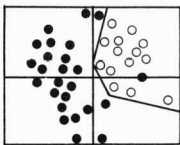
S = 8



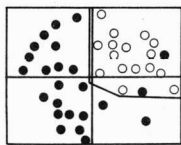
S = 10



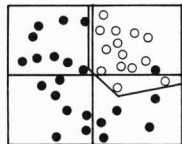
S = 12



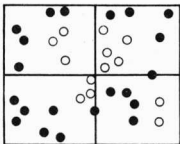
S = 14



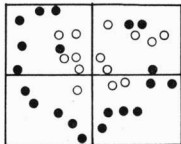
S = 16



S = 18



S = 25



S = 40

ΕΙΚΟΝΑ 3: Σύνοψη των αναλύσεων.

Από την ανάλυση στο S = 12 (πλήθος διακεκριμένων λέξεων = 79) μέχρι την ανάλυση στο S = 2 (πλήθος διακεκριμένων λέξεων = 370) παρατηρείται σταθερότητα του αποτελέσματος.

στην ανάλυση. Από το $S = 12$ και άνω παρατηρείται σταθερότητα του αποτελέσματος.

4.3 Συμπεράσματα

4.3.1 Έννοια της πρόγνωσης στις Επιδημίες

Έχουμε λοιπόν τρεις βασικές ομάδες παρατηρήσεων που αφορούν τη θεραπεία, το θάνατο και τη διάρκεια της ασθένειας. Για καθεμία από τις ομάδες αυτές έχουμε το αντίστοιχο λεξιλόγιο.

Το αποτέλεσμα αυτό, που αφορά το σημασιολογικό περιεχόμενο του εξεταζόμενου κειμένου, δικαιώνει τη θέση ότι στις «παρατηρήσεις» των *Επιδημιών* — βέβαια το συμπέρασμα είναι γενικότερο αφού οι παρατηρήσεις αυτές αποτελούν την καταγραφή της πρακτικής του— ο Ιπποκράτης ενδιαφέρεται για την πρόγνωση της γενικής κατάστασης του ασθενούς και όχι για τη διάγνωση ειδικών ασθενειών. Είναι δε αυτό φυσικό, γιατί η κλινική σημειολογία της εποχής του βρίσκεται στο στάδιο αυτό.

Στο βιβλίο *Επιδημιών I*, παράγραφος 9, ο Ιπποκράτης παραθέτει προγνωστικά σημεία, περιγράφει συμπτώματα που οδηγούν στο θάνατο και άλλα που οδηγούν στη θεραπεία. Στο Γράφημα 4 παρουσιάζουμε τη σημαντική αντιστοιχία μεταξύ των προγνωστικών σημείων του Ιπποκράτη και των αποτελεσμάτων της ανάλυσης και σε επίπεδο λεξιλογίου.

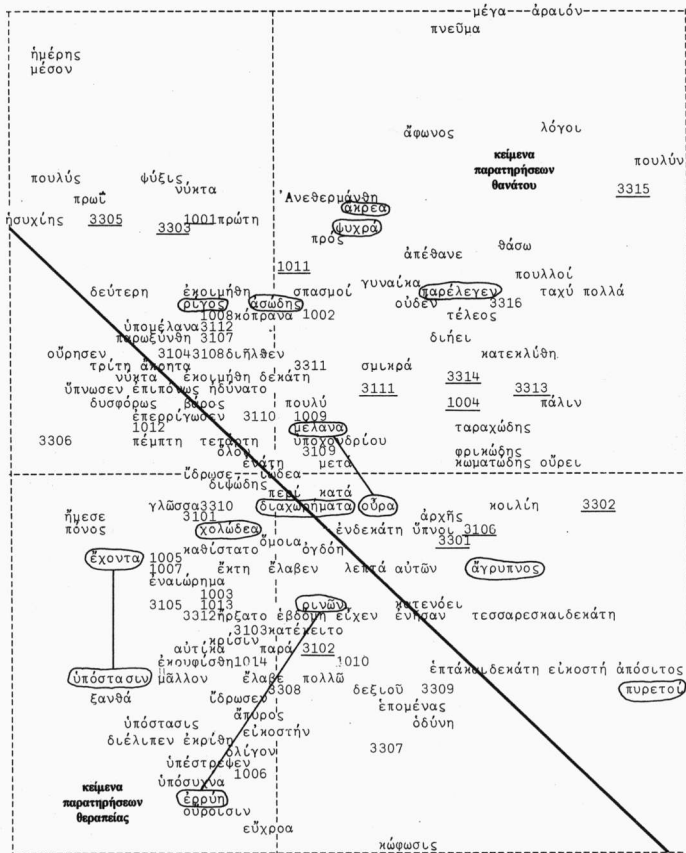
Αλλά η επιτυχία της ανάλυσης διαφαίνεται και από τις προθέσεις του συγγραφέα, όπως αυτές προκύπτουν από την τελευταία παράγραφο του «Προγνωστικού». Γράφει λοιπόν ο Ιπποκράτης: «Αυτός που θέλει να μάθει να προβλέπει σωστά ποιοι ασθενείς θα θεραπευθούν και ποιοι θα πεθάνουν, σε ποιους η ασθένεια θα διαρκέσει περισσότερες μέρες και σε ποιους λιγότερες, πρέπει να κρίνει το καθετί με τη μελέτη των σημείων, όπως αυτά περιγράφηκαν. Πρέπει να έχει βαθιά γνώση των σημείων και να μην ξεχνάει ότι τά άσχημα σημεία αναγγέλουν το κακό και τα καλά το καλό».

4.3.2 Η τυπολογία των παρατηρήσεων: Από το λεξιλογικό περιεχόμενο στο σημασιολογικό

Μια άλλη οπτική της έρευνας αυτής έχουμε με την εισαγωγή μερικών σημειολογικών θεωρήσεων. Το πρόβλημα των παρατηρήσεων τότε τίθεται ως εξής: μέσα στο *corpus* των 42 παρατηρήσεων μπορούμε να θεωρήσουμε την κάθε παρατήρηση με ένα δυσυπόστατο ρόλο, δηλαδή ως σημείο και ταυτόχρονα ως σύστημα σημείων.

Πράγματι, η κάθε παρατήρηση είναι ένα σημείο του συστήματος των παρατηρήσεων, που αποτελούν το *corpus*, αλλά είναι επίσης ένα σύστημα του οποίου

ΓΡΑΦΗΜΑ 4: Σύγκριση προγνωστικών σημείων του Ιπποκράτη (Επιδημιών Ι, παρ. 9) και λεξικών τύπων που προκύπτουν από την ανάλυση του κειμένου (αρ. λέξεων 145).



Αρνητικά σημεία

πυρετός οξύς
επεφρίγεον
άγρυπνοι
αδήμονες
διψώδεις
ασώδεις

εφίδρουν
παρέλεγον
φόβοι
δυσθυμία
άκρα περίψυχα
ούρα μέλανα ολίγα

Θετικά σημεία

ρινών αιμορραγία
ούρα πουλλά και πουλλήν
υπόστασιν έχοντα
κοιλίην ταραχώδεια και
χολώδεια
δυσεντερικοί γενοίατο

τα σημεία είναι οι λέξεις που την αποτελούν. Είναι λογικό κάθε σημείο - παρατήρηση να μπορεί να περιγράφεται και να χαρακτηρίζεται από τα σημεία - λέξεις τα οποία περιλαμβάνει. Εφαρμόζοντας την προαναφερθείσα μεθολογία προκύπτει αυτόματα μια τυπολογία των παρατηρήσεων βασισμένη στο λεξιλόγιό τους. Η συγκεκριμένη δε αυτή τυπολογία μας υποβάλλει το σημασιολογικό αποτέλεσμα που ήδη αναφέρθηκε.

Αν θεωρήσουμε ότι κάθε κείμενο μεταδίδει ένα σήμα, το σήμα αυτό για να μετατραπεί σε μήνυμα προϋποθέτει την ύπαρξη ενός κώδικα. Ο κώδικας αυτός είναι ένας τρόπος σχηματοποίησης της πληροφορίας ώστε να γίνει κοινοποιησιμη. Έτσι αν στη συγκεκριμένη περίπτωση το σήμα προκύπτει από το σύνολο των παρατηρήσεων, ο κώδικας που χρησιμοποιήθηκε για την ερμηνεία του σήματος είναι οι τεχνικές της «ανάλυσης δεδομένων», και σε τελευταία ανάλυση της ευκλείδειας γεωμετρίας που βρίσκεται στη βάση των τεχνικών αυτών.⁷

Τα πλάνα που προκύπτουν από την ανάλυση δεν είναι παρά η σχηματοποίηση της πληροφορίας. Το δε μήνυμα προκύπτει αναπόφευκτα. Η συμφωνία του αποτελέσματος με τη γενική φιλοσοφία της σχολής της Κω αποτελεί απόδειξη της ορθότητας του κώδικα ερμηνείας, δηλαδή τελικά της μεθοδολογίας που εφαρμόστηκε.

Η περίπτωση των «παρατηρήσεων» του Ιπποκράτη δεν είναι παρά μία από τις περιπτώσεις, όπου η επεξεργασία ενός κειμένου μας οδήγησε από το λεξιλογικό περιεχόμενο στο σημασιολογικό. Πρέπει να υπενθυμίσουμε ότι μεταξύ της εισαγωγής του κειμένου στον ηλεκτρονικό υπολογιστή και των πλάνων των αναλύσεών μας, δεν υπάρχει καμία παρέμβαση ή απόφαση του αναλυτή. Έτσι μια επεξεργασία αυτόματη, στην οποία δεν υπάρχει υποκειμενικότητα, είναι δυνατό να συνεισφέρει στην έρευνα της σημασίας ενός κειμένου.

5. Προοπτικές

Όλες οι γλωσσολογικές έρευνες που αναφέρθηκαν, συμπεριλαμβανομένης και αυτής του Ιπποκράτη (1983), είναι παλαιότερες. Όμως σ' έναν χώρο που επηρεάζεται από ραγδαίες τεχνολογικές εξελίξεις, η ενημέρωση δεν πρέπει να περιορίζεται στην αναφορά ερευνών ήδη πραγματοποιημένων. Δεν θα αναφερθούμε βέβαια εδώ στις δυνατότητες τεχνικών εξελίξεων της Α.Δ. ούτε στη δυνατότητα προσαρμογής των προγραμμάτων για τους μικρούς υπολογιστές, μιας και αποτελούν θέματα ξένα προς τη φύση του περιοδικού. Ούτε καν σε εφαρμογές που πιστεύουμε ότι είναι δυνατές σε κείμενα κάθε είδους: λογοτεχνικά, επιστημονικά ή τεχνικά και σε γλώσσα αρχαία ή νεοελληνική.

7. Το μαθηματικό πρόβλημα της όλης μεθοδολογίας είναι η μελέτη ενός νέφους σημείων σ' έναν ευκλείδειο χώρο μεγάλης διάστασης.

Θα αναφερθούμε μόνο σε δύο εφαρμογές που έχουν σχεδιαστεί και εγκριθεί για το '85. Πρόκειται για την εφαρμογή στην επεξεργασία των ανοιχτών ερωτήσεων στα πλαίσια των ερευνών του ΕΚΚΕ. Εφαρμογή, η οποία θα απαλλάξει τους ερευνητές από την κοπιώδη και αμφισβητούμενη εργασία της ανίχνευσης των θεμάτων μεγάλου αριθμού απαντήσεων (από 1.000 έως 5.000 απαντήσεις) και ταυτόχρονα θα μειώσει το κόστος της έρευνας.

Μια άλλη πρωτότυπη εφαρμογή, που έχει προγραμματιστεί σε συνεργασία με το Ψυχολογικό Τμήμα του Πανεπιστημίου Κρήτης, αφορά στην ανάλυση συνεντεύξεων μεγάλης διάρκειας με θέμα τη διερεύνηση των ψυχολογικών παραγόντων στο φαινόμενο έκτρωσης.

Πιστεύουμε ότι η Ανάλυση Δεδομένων, μη θεωρώντας αποκλειστικά τον ηλεκτρονικό υπολογιστή σαν μια βοηθητική - μνήμη, προσφέρει στον ασχολούμενο με τις επιστήμες του ανθρώπου ένα ισχυρό εργαλείο εργασίας.

ΒΙΒΛΙΟΓΡΑΦΙΑ - ΑΝΑΦΟΡΕΣ

- T. Behpakis, «Vers une analyse automatique des textes: le Traitement des 42 observations contenues dans les livres *Epidemies I et III* d' Hippocrate», *Cahiers de l'analyse des donnees*, τ. viii (4), 1983, σ. 475 - 489.
- J.P. Benzecri, *L'analyse des donnees, l'analyse des correspondances*, Dunod, Paris 1980.
- J.P. Benzecri κ.ά., «Pratique de l'analyse des donnees», τ. III, *Linguistique et lexicologie*, dunod, 1980.
- J.P. Benzecri, «Descriptions des textes et analyse documentaire», *Cahiers de l'analyse des donnees*, τ. IX (2), 1984, σ. 205 - 211.
- H. Dreyfus, *Intelligence artificielle, mythes et limites*, Flammarion 1985.
- L. Lebart, «Une procedure d'analyse lexical ecrite en langage fortran (Ana. Lex.)», *Cahiers de l'analyse des donnees*, 1981 (2), 1981 σ. 229 - 243.
- L. Lebart - A. Morineau, Spad - 1984. *Analyses de donnees textuelles*. Παρίσι C.E.S.I.A., 1984.
- E. Littré, Hippocrate, *Oeuvres completes*, 1839 - 1861. Τόμος II, σελ. 538.