

ΧΡΗΣΕΙΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΤΗΝ ΥΦΟΛΟΓΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΛΟΓΟΤΕΧΝΙΚΩΝ ΚΕΙΜΕΝΩΝ

Η στατιστική στην υφολογική ανάλυση των λογοτεχνικών κειμένων παρουσιάζεται σα «μια ποσοτική φιλοσοφία της γλώσσας»¹. Στηρίζεται στην καταγραφή στοιχείων, σε μετρήσεις και χρησιμοποιεί μαθηματικές μεθόδους. Ο ερευνητής οφείλει να συγκεντρώνει ενότητες προσδιορισμένες και ομοιογενείς, π.χ. ομοιοκαταληξίες, γραμματικές κατηγορίες, σχήματα, διαφόρους τύπους φράσεων, λέξεις-κλειδιά, λέξεις-θέματα κ.τ.λ.

Πώς μπορούμε να θεμελιώσουμε τις εργασίες της στατιστικής ανάλυσης του ύφους ενός λογοτεχνικού κειμένου; Αν για πεδίο παρατήρησης στη μελέτη μας λάβουμε ολόκληρο το κείμενο, η μέθοδος μας εξασφαλίζει περισσότερες εγγυήσεις. Αλλά αν το κείμενο είναι μεγάλης έκτασης, η εφαρμογή της μεθόδου μας είναι δυσκολότερη, διότι είμαστε υποχρεωμένοι να συγκροτήσουμε δείγματα των οποίων η επιλογή είναι προβληματική². Σχετικά με τη συγκρότηση των δειγμάτων, κατά τον Pierre Guiraud οι κανόνες που προσδιορίζουν τους στατιστικούς χαρακτήρες π.χ. του λεξιλογίου, εξακριβώνονται για δείγματα που περιλαμβάνουν από 10.000 ως 50.000 λέξεις - παραπομπές³. Ο Pierre Guiraud επισημαίνει επίσης τα εμπόδια που συναντούν οι στατιστικολόγοι με στοιχεία τα οποία δημιουργούν ασάφειες ως προς τον χαρακτηρισμό τους, όπως στοιχεία με διπλή γραμματική φύση (λ.χ. η μετοχή) ή με πολλαπλές συντακτικές χρήσεις (λ.χ. ορισμένα επιρρήματα, σύνδεσμοι, προθέ-

1. Herdan G., *The advanced theory of Language as Choice and Chance*, Springer Verlag, Berlin - Heidelberg, New York, 1966, σ. 9. (Ιδρυτής της στατιστικής στην ανάλυση του ύφους θεωρείται ο Boole χωρίς να αγνοείται η συμβολή του De Morgan: Βλ. Bailey W., Dolezel L., *Statistics and style: A historical Survey*, Elsevier, New York - Londres - Amsterdam, 1968, σ. 217-238).

2. Βλ. Yule G. Udney, *The statistical Study of literary vocabulary*, Archon Books, U.S.A. 1968, *passim*.

3. Μεταξύ του αριθμού των occurrences (λέξεων - παραπομπών) (N) και του αριθμού των λέξεων - λημμάτων (V), υπάρχει σχέση που εκφράζει τον πλούτο του λεξιλογίου ενός κειμένου. Ο Pierre Guiraud θεμελιώνει ως εξής το δείκτη πλούτου του λεξιλογίου: δείκτης πλούτου: $V/\sqrt{N}$ - Guiraud Pierre, *Problèmes et Méthodes de la statistique linguistique*, ed. PUF., Paris, 1960, σ. 20. (α) ένα κείμενο είναι γραμμένο με λέξεις, τη μία δίπλα στην άλλη, που καλύπτουν τη σελίδα: τις λέξεις - παραπομπές, που ορίζουν την έκταση του κειμένου και (β) για να γραφεί το κείμενο αυτό, χρησιμοποιήθηκαν ορισμένες λέξεις: είναι οι λέξεις - λήμματα, αυτές που αποτελούν το λεξιλόγιο του κειμένου». Βλ. Κοκόλη Ξ. Α., «Λέξεις-άπαξ»: *Στοιχείο ύφους. Θεωρητική εξέταση - καταγραφή στα «Ποιήματα» του Γ. Σεφέρη*, εξάντας, Αθήνα, 1975, σ. 25).

σεις και μόρια)⁴, εμπόδια όχι ανυπέρβλητα κατά τον Ερατοσθένη Γ. Καψωμένο⁵.

Ας εξετάσουμε στη συνέχεια μερικά συγκεκριμένα παραδείγματα χρήσης της στατιστικής στην ανάλυση του ύφους των λογοτεχνικών κειμένων.

Στη στατιστική ανάλυση του έργου του Paul Valéry, ο Pierre Guiraud χρησιμοποιεί τον όρο συχνότητα, του οποίου τους γενικούς κανόνες προσδιόρισε ο G. K. Zipf⁶, και απομονώνει λέξεις - κλειδιά και λέξεις - θέματα. Καταλήγει στο συμπέρασμα ότι στο έργο αυτού του συγγραφέα, «ombre» (σκιά) και «pur» (καθαρός) είναι οι δύο πρώτες λέξεις - κλειδιά που επανέρχονται με τη μεγαλύτερη συχνότητα⁷.

Ο Pierre Guiraud παρουσιάζει επίσης μια στατιστική ανάλυση στο ποιητικό έργο του Arthur Rimbaud. Εξετάζει σε κάθε ποίημα τη συχνότητα φωνητικών, λεκτικών και γραμματικών χαρακτηριστων, π.χ. του ουσιαστικού, της υποτακτικής, της ερώτησης, του πληθυντικού κτλ.⁸ για να καταλήξει στο συμπέρασμα ότι η ποιητική συλλογή *Illuminations* (Εκλάμψεις), στην καμπύλη της εξέλιξης του ύφους του ποιητή τοποθετείται χρονολογικά μετά το 1873 που είναι η ημερομηνία σύνθεσης του έργου *Une Saison en Enfer* (Μια Εποχή στην Κόλαση).

Αξιόλογη εξάλλου είναι η συμβολή του Charles Muller που πραγματοποίησε μια μελέτη συγκρίνοντας τα επίθετα στο έργο *Horace* (Οράτιος) του P. Corneille με τα επίθετα στα έργα *L' Illusion comique* (Κωμική Ψευδαίσθηση), *Le Cid* (Ο Σιντ), *Cinna* (Κίννας), *Polyeucte* (Πολύεκτος) του ίδιου συγγραφέα. Παίρνοντας για πρότυπο τη συχνότητα των επιθέτων στο σύνολο των έργων αυτών, συγκρίνει αυτό το πρότυπο με τη συχνότητα των επιθέτων στο θεατρικό έργο που αποτελεί κάθε φορά το πεδίο παρατήρησης και βρίσκει μια απόκλιση που οπωσδήποτε δεν είναι τυχαία⁹.

Ο ίδιος συγγραφέας εξετάζει επίσης την ποσοτική σχέση ανάμεσα

4. Guiraud Pierre, *Problèmes et Méthodes de la statistique linguistique*, ό.π., σ. 35, 47, 48.

5. Καψωμένος Ερατοσθένης Γ., *Η συντακτική δομή της ποιητικής γλώσσας του Σεφέρη*, Επιστ. Επετ. Φιλολ. Σχολ. Πανεπιστημίου Θεσσαλονίκης, παράρτ. αρ. 18, Θεσσαλονίκη, 1975, σ. 48.

6. Zipf G. K., *Selected studies of the principle of relative frequency in Language*, Harvard University Press, 1932, passim.

7. Guiraud Pierre, *Langage et Versification d' après l' œuvre de Paul Valéry*, éd. C. Klincksieck, Paris, 1953, σ. 155-162.

8. ID, *Problèmes et Méthodes de la statistique linguistique*, ό.π., σ. 134.

9. Muller Charles, *Quelques principes de linguistique statistique: expérience, calcul et prévision*, Bull. des jeunes romantistes, Stasbourg IX, 1965, passim. (Για το θέμα αυτό βλ. επίσης: Wells R., «Nominal and Verbal Style», στον τόμο *Style in Language*, επιμ. Th. A. Sebeok, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1964, σ. 215).

στο ρήμα και το ουσιαστικό στο έργο *L' Illusion comique* (Κωμική Ψευδαίσθηση) του P. Corneille και καταλήγει ότι στους ποσοτικούς χαρακτήρες του λεξιλογίου υπάρχει κυρίως μια ένδειξη δραματικής κατάστασης και λιγότερο μια ένδειξη ύφους¹⁰.

Χρησιμοποιώντας τη στατιστική στη διερεύνηση του λεξιλογίου, ο Arthur Bodson εντοπίζει ορισμένες αναλογίες αποκαλυπτικές για την ψυχολογική κατάσταση της ηρωίδας Διδούς, στο έργο *Enéide* (Αινειάδα) του Βιργιλίου. Επισημαίνει, μεταξύ άλλων, το γεγονός ότι ενώ στο πρώτο μέρος η Διδώ χρησιμοποιεί πολλά κύρια ονόματα, πράγμα που αποτελεί ένα χαρακτηριστικό του βασιλικού λόγου ή του επικού ύφους, δηλ. 14% των λέξεων - παραπομπών, στο δεύτερο μέρος, τα κύρια αυτά ονόματα δεν αποτελούν παρά μόνο το 4%¹¹.

Η στατιστική ανάλυση βέβαια με την ερμηνεία των δεδομένων της, συμπληρώνει την με άλλους τρόπους προσέγγιση ενός λογοτεχνικού κειμένου. Ο Georges Matoré π.χ. προβαίνει σε λεπτομερή εξέταση του λεξιλογίου των αισθήσεων σε πολλά γαλλικά μυθιστορήματα του Κ' αιώνα¹². Στο μυθιστόρημα *L' Etranger* (Ο Ξένος) του Camus για παράδειγμα, η φωτεινότητα (όροι: λάμψη, φως, λαμπρός κτλ.), παίζει σπουδαίο ρόλο: είναι το 24% του συνόλου των όρων των σχετικών με τις γενικές κατηγορίες χρώματα, φως, σκοτάδι. Τα αντικείμενα στο μυθιστόρημα αυτό είναι τόσο προικισμένα «με μια καθαρότητα που πληγώνει», που ο Meursault, ήρωας του έργου, «πληγωμένος από τη λάμψη τους αισθάνεται ότι ασυνείδητα βρίσκεται σε κατάσταση νόμιμης άμυνας»¹³.

Η Simone Monsonégó από την πλευρά της επιχειρεί μια διερεύνηση της κατανομής των λέξεων σε πεζά και έμμετρα κείμενα του ίδιου λογοτέχνη: διαπιστώνει ότι η επιλογή λεξιλογίου που κάνει ένας συγγραφέας όταν γράφει στίχους διαφέρει από την επιλογή που κάνει ο ίδιος όταν γράφει πεζό κείμενο¹⁴.

Τη στατιστική έχω χρησιμοποιήσει και εγώ ερευνώντας το λεξιλόγιο των χρωμάτων της προτίμησης του Jean-Paul Sartre, στο μυθιστορηματικό έργο *Les Chemins de la liberté*. Τα δεδομένα που συγκέντρωσα έδειξαν ότι η επιλογή των χρωμάτων δεν είναι τυχαία. Τα χρώματα που κυριαρχούν στον πρώτο τόμο είναι τα εξής: κόκκινο 16%, μαύρο 15%,

10. Muller Charles, *Essai de statistique lexicale. L' Illusion comique de P. Corneille*, Libr. Klincksieck, Paris, 1964, σ. 27, 30.

11. Bodson Arthur, «Etude littéraire de l' *Enéide*», στη συλλογική έκδοση *Statistique et Analyse linguistique, Colloque de Strasbourg*, éd. P.U.F., 1966, σ. 82.

12. Matoré Georges, *L' Espace humain*, éd. Nizet, Paris, 1976, σ. 216-235.

13. Id., *ibid.*, σ. 221.

14. Monsonégó Simone, *Etude stylo-statistique du vocabulaire des vers et de la prose dans la chantefable Aucassin et Nicolette*, Libr. Klincksieck, Paris, 1966, σ. 74.

λευκό 14%, πράσινο 8%. Στο δεύτερο τόμο τα εξής: λευκό 22%, μαύρο 18%, κόκκινο 17%, μπλε 6%. Στον τρίτο: μαύρο 24%, κόκκινο 18%, λευκό 13%, γκρι 7%. Στον τέταρτο¹⁵: μαύρο 37%, κόκκινο 25%, γκρι 10%, μπλε 8%¹⁶.

Ο κόσμος των χρωμάτων στα μυθιστορήματα αυτά κυριαρχείται από το μαύρο, το κόκκινο, το λευκό, το γκρι, το μπλε και κατά τον Jacques Salvan, που αφιερώνει ένα κεφάλαιο στον υπαρξιακό συμβολισμό, το μαύρο συμβολίζει κάθε τι το σταθερό, το συμπαγές (σύμβολο του *en soi*), το λευκό την αδιαφάνεια μιας συνείδησης προς κάποια αξία κτλ.¹⁷

Επίσης επιχειρήσα την καταγραφή στοιχείων, που έχουν σχέση με την κατανομή της δράσης (*action*) στην ίδια μυθιστορηματική σειρά, στις χρονικές περιόδους του ημερονυκτίου (πρωί, μεσημέρι, απόγευμα, βράδυ, νύχτα) και κατέληξα στα εξής συμπεράσματα: Στον πρώτο τόμο του μυθιστορηματικού κύκλου, η δράση κατά ένα μεγάλο μέρος πραγματοποιείται αργά το βράδυ ή τη νύχτα (εικόνα 1). Πρόκειται για θέματα που αφορούν ανθρώπινες σχέσεις κατά το διάστημα της ειρήνης πριν από την έκρηξη του πολέμου του 1940. Αντίθετα στους άλλους τρεις τόμους, η δράση, η οποία εκτυλίσσεται υπό την απειλή μιας πιθανής έκρηξης πολέμου (τόμος 2, εικόνα 2), ή κατά τη διάρκεια του πολέμου, της ήττας και της αιχμαλωσίας (τόμοι 3-4, εικόνες 3-4), τοποθετείται σε σημαντικό ποσοστό στο διάστημα από το πρωί ως το απόγευμα. Έτσι συμπεραίνουμε ότι η επιλογή μιας χρονικής περιόδου έχει άμεση σχέση με τη θεματική στα παραπάνω μυθιστορήματα. Η διαφορά στην επιλογή της χρονικής στιγμής υποδηλώνει και θεματική διαφοροποίηση ή διαφορετική διαπραγμάτευση του ίδιου θέματος.

Στον ελληνικό χώρο, ο Ερατοσθένης Γ. Καψωμένος επιχειρεί «ένα στατιστικό προσδιορισμό των ποσοτικών χαρακτήρων όλων των γλωσσικών κατηγοριών (των μερών του λόγου)», προκειμένου να τους συγκρίνει με τους αντίστοιχους χαρακτήρες μιας ορισμένης νόρμας με σκοπό να διαπιστώσει «την πιθανή παρέκκλιση που θα αντιπροσωπεύει την ποσοτική έκφραση της υφολογικής ιδιοτυπίας του ποιητή»¹⁸.

Με το ποιητικό έργο του Σεφέρη επίσης ασχολείται και ο Ξ. Α. Κοκόλης, ο οποίος συγκεντρώνει τις λέξεις που στο σύνολο του έργου

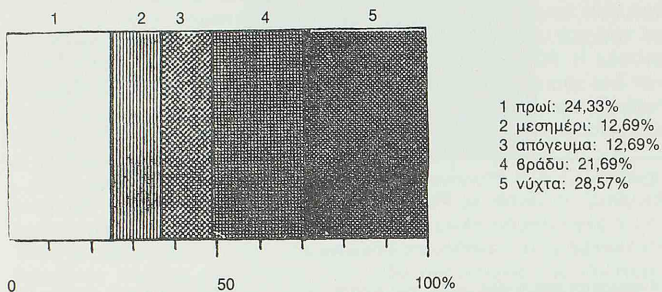
15. Μυθιστόρημα (με τίτλο *Drôle d' Amitié*) ατελείωτο, δημοσιευμένο στο περιοδικό *Les Temps Modernes* (Μοντέρνοι Καιροί).

16. Tabaki Frideriki, *La technique du roman dans les Chemins de la liberté de Jean-Paul Sartre*, Doctorat de 3^e cycle, Univ. de Paris IV, 1978, σ. 72-75.

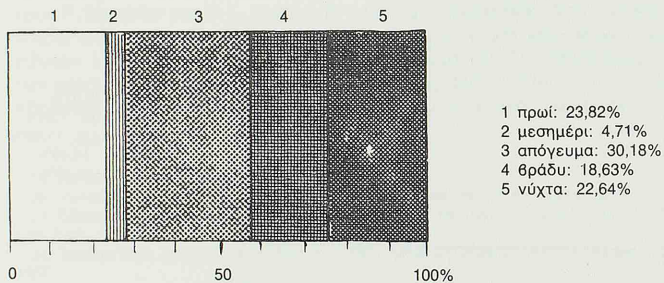
17. Salvan Jacques, *The Scandalous Ghost. Sartre's Existentialism as Related to Vitalism, Humanism, Mysticism, Marxism*, Détroit, 1967, σ. 94-117.

18. Καψωμένος Ερατοσθένης Γ., *Η συντακτική δομή της ποιητικής γλώσσας του Σεφέρη*, ό.π., σ. 51-52.

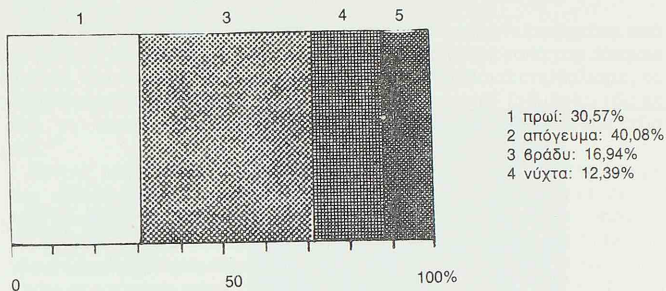
Ποσοστιαία αναλογία δράσης (action).



Εικόνα 1.

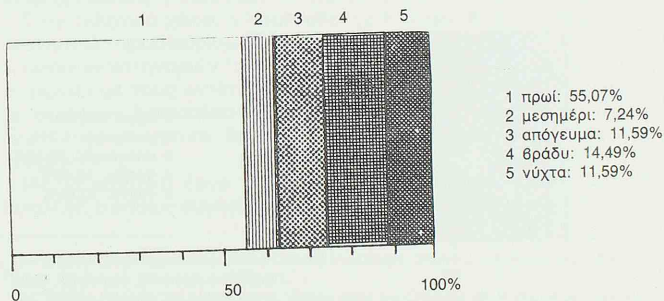


Εικόνα 2.



Εικόνα 3.

(Αναφέρεται στο πρώτο μέρος του τρίτου τόμου της μυθιστορηματικής σειράς. Στο δεύτερο μέρος του τρίτου τόμου η υπόθεση διαδραματίζεται κατά το μεγαλύτερο μέρος το πρωί. Η καταμέτρηση στοιχείων είναι αδύνατη διότι δεν υπάρχουν συγκεκριμένες ενδείξεις).



Εικόνα 4.

του ποιητή έχουν χρησιμοποιηθεί μια μόνο φορά. Τις άπαξ αυτές λέξεις εξετάζει ως στοιχείο του ύφους του ποιητή¹⁹.

Τα τελευταία χρόνια, η χρήση των ηλεκτρονικών υπολογιστών συνέβαλε στην πρόοδο των εργασιών καταγραφής και ταξινόμησης και στην απλοποίηση των υπολογισμών. Ο ηλεκτρονικός υπολογιστής μπορεί να συγκεντρώνει ακριβή στοιχεία ύφους, να συλλέγει λέξεις, φράσεις κτλ. και να καταγράφει ομοιότητες και συχνότητες.

Το 1957 έγινε σημαντική προσπάθεια στον τομέα αυτό από τον καθηγητή Stephen Parrish και τους συνεργάτες του, που άρχισαν να συλλέγουν στοιχεία με ηλεκτρονικό υπολογιστή IBM 704. Η έρευνα τους οδήγησε στη δημοσίευση του πρώτου τόμου μιας σειράς από τον εκδοτικό οίκο του Πανεπιστημίου του Cornell, με θέμα το λεξιλόγιο του ποιητή Matthew Arnold²⁰, ενώ ακολούθησαν ανάλογες εργασίες του J. W. Ellison²¹ και του Paul Tasman²².

Εξάλλου με τη βοήθεια ηλεκτρονικού υπολογιστή ο J. T. Donough ετοίμασε ένα πίνακα μετρικής της *Ιλιάδας* με βάση τη ρυθμική λειτουργία των 112.000 λέξεων στις 157 μετρικές παραλλαγές τους. Όταν ο ρυθμός των 15.693 στίχων κωδικοποιήθηκε και γράφτηκε σε διάτρητες κάρτες, μια μηχανή IBM 650 απομόνωσε τις ιδιαίτερα χρησιμοποιούμενες από τον ποιητή χαρακτηριστικές λέξεις και εκτύπωσε τα αποτελέσματα σε λίγες ώρες. Η στατιστική αυτή έρευνα απέδειξε ότι η *Ιλιάδα* του Ομήρου δεν αποτελείται από ξεχωριστά ποιήματα διαφορετικών ποιητών συνδεδεμένα μεταξύ τους, αλλά ότι γράφτηκε από ένα μόνο ποιητή²³.

Ανάλογες μελέτες έγιναν από τον καθηγητή Walker M. Gibson και τον George Petty, στο μυθιστόρημα *Daisy Miller* του Henry James· από τους F. Mosteller και D. L. Wallace²⁴, τη Sally Y. Sedelow²⁵. Από τον Karl Krober επίσης έγινε ανάλυση συγκεκριμένων συντακτικών δομών και ειδικών λεξιλογίων αντιπροσωπευτικών δειγμάτων 22 μυθιστορημάτων πέντε βρετανών μυθιστοριογράφων (της Jane Austen, των τριών αδελφών Brontë και του George Eliot), με τη συμβολή ηλεκτρονικού

19. Κοκόλης Ξ. Α., «Λέξεις - άπαξ»... ό.π., passim.

20. Parrish S. M., *A Concordance to the Poems of Matthew Arnold*, Ithaca 1959.

21. Ellison J. M., *Nelson's Complete Concordance to the Revised Standard Version Bible*, New York, 1957.

22. Tasman Paul, *Literary Data Processing*, IBM Journal of Research and Development, I, 1957.

23. McDonough J. T., «Homer, the Humanities and IBM», *Proceedings of the IBM Literary Data Processing Conference*, Yorktown Heights, N.Y., 1964, σ. 25-36.

24. Mosteller F., Wallace D. L., «Inference in an Authorship Problem», *Journal of the American statistical Association*, LVIII, 1962, σ. 275-309.

25. Sedelow Y. Sally, «A Preface to computational Stylistics», *The Computer and Literary Style*, ed. I. Leed, Kent, 1966, σ. 1-13.

υπολογιστή που του έδωσε τη δυνατότητα να εργαστεί με ένταση και σε μεγάλη έκταση²⁶.

Η στατιστική ανάλυση των δεδομένων ενός λογοτεχνικού κειμένου στην υφολογική έρευνα, προϋποθέτει μεθοδική εργασία με υπολογισμούς, μετρήσεις, απαριθμήσεις, κατηγοριοποιήσεις κτλ., μέρος των οποίων συχνά μένει ανεκμετάλλευτο ή χωρίς ενδιαφέρον. Τα παραδείγματα που αναφέραμε δίνουν μια ιδέα των δυσκολιών που συναντά ο ερευνητής. Ο R. Moreau αναρωτιέται αν μπορούμε να εφαρμόσουμε τα μαθηματικά σύνολα στην υφολογική ανάλυση, θεωρώντας ότι ο λογοτεχνικός λόγος δεν υπακούει ποτέ σε κανόνες, αλλά μόνο σε τάσεις, που είναι συνάρτηση σύνθετων προσωπικών και κοινωνικών παραγόντων²⁷. Συχνά η εσωτερική κίνηση ενός έργου διαφεύγει από τη στατική μέθοδο των μετρήσεων: «Η σειρά εμφάνισης των σημείων δεν είναι αδιάφορη προς τη σημασία του έργου και δίνει μια "πληροφορία" που διαφεύγει από τη στατιστική μελέτη», σημειώνει ο Tzvetan Todorov²⁸.

Ποιές είναι τελικά οι προοπτικές της στατιστικής ανάλυσης στην υφολογία; «Οπωσδήποτε, η στατιστική ανάλυση συμπληρώνει, δεν ανα-πληρώνει τη δομική ανάλυση»²⁹. Θα πρέπει να λάβουμε υπόψη μας τα όρια των δυνατοτήτων της «μηχανιστικής» μεθόδου συμπληρώνοντας την με άλλα δεδομένα (ιστορικά, κοινωνικά, ψυχολογικά, κτλ.). Για να συστηματοποιήσουμε όμως την έρευνα διαφόρων στοιχείων του ύφους ενός λογοτέχνη δεν μπορούμε να προθυμώμε στην ταξινόμηση ή στη σύγκριση πληροφοριών χωρίς προηγουμένως να μετρήσουμε.

26. Kroeber Karl, «Computers and research in literary analysis», *Computers in humanistic Research*, επιμ. Bowles A., ed. Prentice Hall, Englewood Cliffs, N. Jersey, σ. 138.

27. Moreau R., *Statistique et analyse linguistique*, Colloque de Strasbourg, éd. P.U.F., Paris, 1966, σ. 125-132.

28. Todorov Tzvetan, «Procédés mathématiques dans les études littéraires», *Annals*, n° 3, 1965.

29. Μπαμπινιώτης Γεώργιος, *Γλωσσολογία και Λογοτεχνία*, εκδ. Δ. Μαυρομάτη, Αθήνα, 1984, σ. 43.

RESUMÉ

Frideriki Tabaki-Ionas, *La statistique à l'analyse stylistique des textes littéraires*

La statistique à l'analyse stylistique des textes littéraires se présente comme «une philosophie quantitative du langage». Elle repose sur des inventaires et des dénombrements employant des méthodes mathématiques. Les unités qu'on rassemble doivent être bien définies et homogènes.

Les règles générales de la fréquence, établies par G. K. Gípf, sont utilisées dans les études statistiques de vocabulaire et ont abouti à des résultats relatifs aux traits caractéristiques des textes littéraires.

La statistique permet de répondre s'il existe une relation entre la fréquence d'un élément et celle de plusieurs d'autres; aussi peut-elle répondre à la question de la répartition des unités observées à l'intérieur d'une œuvre.

Les exemples que nous mentionnons (empruntés à P. Guiraud, Ch. Muller, A. Bodson etc.) donnent une première idée des possibilités de la stylistique statistique.

L'emploi des ordinateurs, ces dernières années, a contribué à la simplification des études de calculs de mots, de caractères morphologiques, syntaxiques, matériel utile à l'analyse linguistique et littéraire du style.

L'analyse quantitative est indispensable pour la systématisation de la recherche de différents éléments du style d'un écrivain mais elle doit être complétée nécessairement par d'autres données.