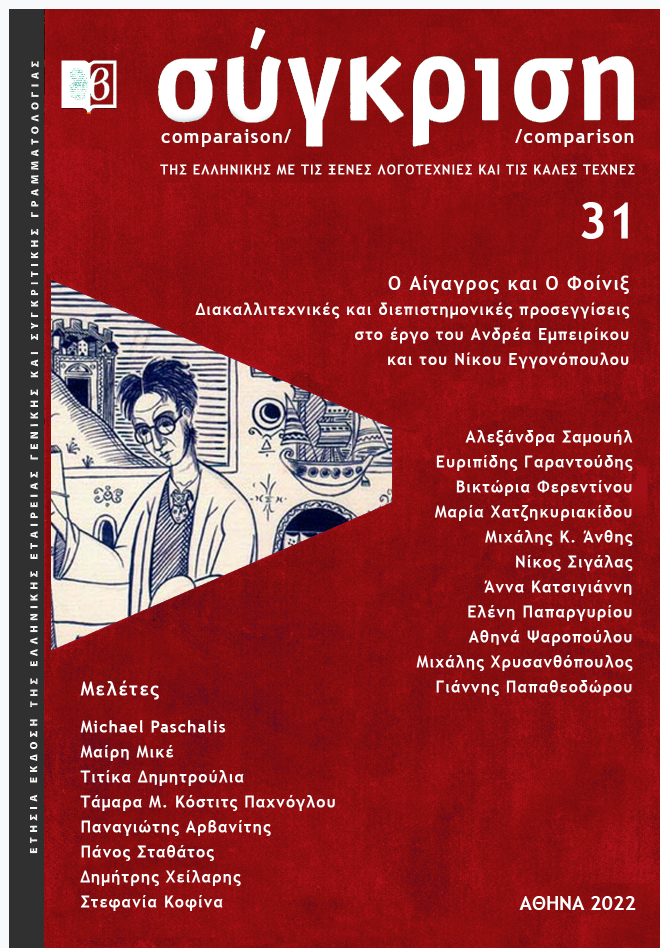


Σύγκριση

Τόμ. 31 (2022)



ΤΙΤΙΚΑ ΔΗΜΗΤΡΟΥΛΙΑ, Τα ηλεκτρονικά σώματα κειμένων και η ανάλυσή τους. Ερευνητικές προοπτικές για τη μελέτη της λογοτεχνίας

Τιτίκα Δημητρούλια

doi: [10.12681/comparison.32406](https://doi.org/10.12681/comparison.32406)

Copyright © 2022, Τιτίκα Δημητρούλια



Άδεια χρήσης [Creative Commons Attribution-NonCommercial-ShareAlike 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Βιβλιογραφική αναφορά:

Δημητρούλια Τ. (2022). ΤΙΤΙΚΑ ΔΗΜΗΤΡΟΥΛΙΑ, Τα ηλεκτρονικά σώματα κειμένων και η ανάλυσή τους. Ερευνητικές προοπτικές για τη μελέτη της λογοτεχνίας. *Σύγκριση*, 31, 160–184. <https://doi.org/10.12681/comparison.32406>

**Τα ηλεκτρονικά σώματα κειμένων και η ανάλυσή τους.
Ερευνητικές προοπτικές για τη μελέτη της λογοτεχνίας**

1. Εισαγωγή: ανασκοπήσεις, ορισμοί και προβληματισμοί

Ψηφιακά εργαλεία χρησιμοποιούνται στις ανθρωπιστικές επιστήμες εδώ και πολλές δεκαετίες: από τον πρώτο ηλεκτρονικό συμφραστικό πίνακα των έργων του Θωμά του Ακινάτη που δημιούργησε ο ιταλός ιερέας Roberto Busa σε συνεργασία με την IBM (Busa 1980· Winter 1999) στις αρχές της δεκαετίας του '50· αλλά και το πρώτο συστηματικό εγχείρημα αυτόματης μετάφρασης, την ίδια περίοδο, η αποτυχία του οποίου αναπροσανατόλισε ολόκληρο κλάδο της έρευνας που αφορούσε την ηλεκτρονική μελέτη της γλώσσας σε λύσεις που περιλάμβαναν την ανθρώπινη παρέμβαση (Δημητρούλια 2005). Το έργο του Busa αναφέρεται ως ιδρυτική πράξη του αγγλοσαξονικής σύλληψης διεπιστημονικού γνωστικού πεδίου των Digital Humanities (DH, ψηφιακές ανθρωπιστικές επιστήμες) (Hockey 2004), που έκανε την εμφάνισή του στις αρχές της νέας χιλιετίας (Schreibman, Siemens & Unsworth 2004). Το εν λόγω πεδίο πάντως έχει μια μακρά και πολύπλοκη ιστορία, που συχνά αναφέρεται αποσπασματικά και παρουσιάζει μεγάλο ενδιαφέρον για την ίδια την επιστημολογική θεώρησή του. Στην ιστορία αυτή εντάσσονται τα αποτελέσματα της έρευνας με ψηφιακές μεθόδους σε χώρες που είχαν πρωτοστατήσει στην αλλαγή του ερευνητικού παραδείγματος στην υπολογιστική γλωσσολογία από τη δεκαετία του '60 ήδη, όπως η Γαλλία και η Γερμανία, οι οποίες στη συνέχεια ανέπτυξαν περαιτέρω σημαντική έρευνα στη χρήση των εργαλείων στη γλωσσολογία, τη μετάφραση και τη λογοτεχνία (Brunet 2014· Δημητρούλια 2005· Krollmann 1981)¹ αλλά και η χρήση της ψηφιακής τεχνολογίας στις λογοτεχνικές και καλλιτεχνικές αβάν-γκαρντ την ίδια περίοδο (Λαλιώτη 2022), στα πρώτα πειράματα ηλεκτρονικής λογοτεχνίας (Δημητρούλια 2006) και στις παραστατικές τέχνες. Η χρήση των ψηφιακών μέσων στο θέατρο μάλιστα ήταν τόσο σημαντική, ώστε ο Hans Thies Lehmann να τα θεωρεί πρωταρχικό μοχλό για την εμφάνιση του μεταδραματικού θεάτρου (2006), όπου και συνεχίζουν ποικιλοτρόπως να κυριαρχούν (Chapple και Kattenbelt 2006· Marranca 2009 και 2010). Είναι προφανές, ως εκ τούτου, ότι το θεωρούμενο ως χρονικό ορόσημο της γέννησης των ψηφιακών ανθρωπιστικών επιστημών στις αρχές της νέας χιλιετίας δεν σχετίζεται εντέλει με την ανάδυση της ψηφιακότητας, που αναπτύσσεται ποικιλοτρόπως σε προγενέστερο χρόνο και για κάποιους ανάγεται έως και στην αρχαιότητα (Δημητρούλια 2006), αλλά με την απλοποίηση και ευρεία διάχυση των υπολογιστών και των δικτύων και, κυρίως, με τον «πυρετό ψηφιοποίησης» στη δεκαετία του '90: «το μεγαλύτερο μέρος της παγκόσμιας πολιτισμικής κληρονομιάς είχε πλέον υποστεί ψηφιακό αναδιπλασιασμό

¹ Ειδικά η «γλωσσολογική στατιστική», που θα οδηγήσει στη λεξικομετρία και τη σημερινή, διευρυμένη λογομετρία και κειμενομετρία, έχει γαλλική αφετηρία, βλ. Guiraud και Whatmouth 1954. Μια δεκαετία αργότερα, θα κάνει την εμφάνισή της η αγγλική «ποσοτική γλωσσολογία» (Herdan 1964).

και, χάρη στην ταυτόχρονη ανάπτυξη του Παγκόσμιου Ιστού, προσφερόταν για μελέτη, μετασχηματισμό και αξιοποίηση από ειδικούς και μη» (Πατηνιώτης 2020, 7).

Αυτή η συνθήκη δεν υλοποιείται ομοιοτρόπως για όλους τους γλωσσικούς και πολιτιστικούς χώρους, αφού οι πολιτικο-οικονομικές και πολιτιστικές ιεραρχίες αναπαράγονται και στο ψηφιακό περιβάλλον – η αλλιώς, «αυτό που μεταβαίνει στην ψηφιακή διάσταση είναι η κοινωνία με τους ανταγωνισμούς της» (Πατηνιώτης 2020, 14). Από την άλλη, είναι εξίσου σαφές ότι η εμφάνιση του εν λόγω πεδίου εξέφρασε τον υπαρκτό ψηφιακό μετασχηματισμό της γνώσης και απάντησε στις νέες ανάγκες που δημιούργησε στην έρευνα η ψηφιοποίηση, ειδικά σε μια εποχή υποχώρησης των ανθρωπιστικών σπουδών. Αλλά στο επίκεντρο των διαφορετικών στάσεων που διαμορφώνονται απέναντι στο πεδίο βρίσκεται η (φιλοσοφική) αντίληψη περί της ψηφιακότητας, όπως εκφράζεται στον συγκεκριμένο χώρο, ο οποίος, παρά τις όποιες αμφισβητήσεις, εξαπλώνεται σήμερα σε όλο και περισσότερους γλωσσικούς και πολιτισμικούς χώρους και συγκροτείται θεσμικά με έδρες, επιστημονικές ενώσεις, συνέδρια και περιοδικά.

Ανάμεσα στους ερευνητές που δραστηριοποιούνται στις ψηφιακές ανθρωπιστικές επιστήμες, προερχόμενοι από πολλά διαφορετικά πεδία, ορισμένοι τον ορίζουν αρκετά απλά, με αναφορά στην «αυξημένη χρήση των ψηφιακών τεχνολογιών στους κόλπους των ανθρωπιστικών επιστημών και, ευρύτερα, των ανθρωπιστικών και κοινωνικών επιστημών (ΑΚΕ), η οποία παρακολουθεί την τάση των ίδιων των αντικειμένων των εν λόγω κλάδων, που διαρκώς και αναπόδραστα ψηφιοποιούνται (Burnard, 2012)» (Granjon 2016). Κάποιοι μιλούν, σε πολύ πιο δομικό επίπεδο, για αλλαγή παραδείγματος και κάποιοι άλλοι, τέλος, για το «next big thing», μια νέα μεγάλη αφήγηση που σχετίζεται με τη φιλελευθεροποίηση της τριτοβάθμιας εκπαίδευσης (Grusin, 2013)» (Granjon, *ibid.*).

Διαφωνίες υπάρχουν επίσης ως προς τις επιστήμες που συγκαταλέγονται στο πεδίο, οι οποίες εκκινούν προφανώς από τις διαφορετικές επιστημολογικές προκείμενες στους γλωσσικούς και πολιτισμικούς χώρους· αλλά και προβληματισμοί για την ίδια την έννοια των ψηφιακών ανθρωπιστικών σπουδών σε σχέση με τη θέση της τεχνολογίας στην έρευνα –με δεδομένο ότι τα ψηφιακά μέσα, ως τεχνολογίες, είναι τα ίδια πολιτισμικές κατασκευές, με προσίδια λογική, την οποία εγγράφουν στον πολιτισμό, αναδιαμορφώνοντάς τον. Η εστίαση στην (τεχνολογική) διάσταση της έρευνας και όχι στην ερευνητική παράδοση των επιστημών, μια επιλογή που καθορίζει και το εύρος των προσδοκιών από το πεδίο, βρίσκεται στην καρδιά του αντιρρητικού λόγου που αρθρώνεται για τις ψηφιακές ανθρωπιστικές επιστήμες. Ασχέτως της (συχνά έντονης) αυτής συζήτησης, παραμένει, ωστόσο, το γεγονός ότι η ψηφιακότητα έχει μετασχηματίσει, όπως και όλες οι τεχνολογίες του παρελθόντος άλλωστε, τις σύγχρονες πολιτισμικές πρακτικές, της έρευνας συμπεριλαμβανομένης. Από αυτή τη σκοπιά, είναι αναπόδραστη και η παρούσα «ψηφιακή στροφή» και στις ΑΚΕ και τις τέχνες και το μείζον διακύβευμα συνίσταται σ' αυτή ακριβώς τη συζήτηση για τους όρους της.

Σ' αυτή τη στροφή, το κείμενο βρέθηκε από την πρώτη στιγμή στο επίκεντρο. Ιστορικά επικαθορισμένο ως προς την υλικότητα και τη λειτουργία του, έχει επαναπροσδιοριστεί στην ψηφιακή συνθήκη μέσα από τις πολιτισμικές πρακτικές παραγωγής και πρόσληψής του, παραμένοντας ωστόσο πανταχού παρόν (Sinclair και Rockwell 2016, 275). Δεν είναι τυχαίο δε ότι οι ψηφιακές ανθρωπιστικές επι-

στήμες στον αγγλόφωνο χώρο ξεκίνησαν, όπως φαίνεται, από τμήματα φιλολογίας, για λόγους που σχετίζονται με το κείμενο καθαυτό, καθώς η επεξεργασία του με υπολογιστές έχει ήδη μακρά ιστορία, αλλά και με τον ίδιο τον τύπο των κειμενικών δεδομένων, τα οποία είναι τα πιο πρόσφορα για ηλεκτρονική ανάλυση μετά τα αριθμητικά δεδομένα (Kirschenbaum 2010, 60).

Η ψηφιακότητα, όμως, είναι μια κοινωνικοπολιτισμική συνθήκη με ιστορικό πρόσημο, που αναπαράγει, όπως προαναφέραμε, υφιστάμενες πολιτισμικές ιεραρχίες και δημιουργεί νέες. Ως εκ τούτου, ο ρυθμός ενσωμάτωσης των νέων εργαλείων και, κυρίως, των νέων μεθόδων στην έρευνα είναι διαφορετικός τόσο στις διάφορες επιστήμες όσο και στα διάφορα γλωσσικά και πολιτισμικά πεδία: στην Ελλάδα, η συζήτηση για την ηλεκτρονική ανάλυση κειμένου γενικά, σε διάφορους κλάδους των ανθρωπιστικών και κοινωνικών επιστημών, και στις λογοτεχνικές σπουδές ειδικότερα, έχει πολύ πρόσφατα ξεκινήσει, όταν στο εξωτερικό οι εφαρμογές πηγαίνουν πίσω αρκετές δεκαετίες και αφορούν τόσο μακρο- όσο και μικρο-αναλύσεις της λογοτεχνίας, με ειδικά λογισμικά, και σε ποικίλες κατευθύνσεις.²

Στην καθυστέρηση αυτή συμβάλλουν προφανώς οι πολλαπλές δυσκολίες που σχετίζονται με την ψηφιοποίηση των κειμένων και την ανάπτυξη εργαλείων γλωσσικής τεχνολογίας λόγω της ιδιαιτερότητας της ελληνικής· αλλά και η παραδοσιακή επιφυλακτικότητα των ερευνητών της λογοτεχνίας απέναντι στα ηλεκτρονικά εργαλεία –μια επιφυλακτικότητα που δεν αποτελεί πάντως ίδιον γνώρισμα των ελλήνων ερευνητών, αλλά απαντά και σε πολλά άλλα πολιτισμικά περιβάλλοντα. Παρά ταύτα, η ψηφιοποίηση των λογοτεχνικών κειμένων και των κειμένων αναφοράς, των εργαλείων δηλαδή στα οποία στηρίζεται η φιλολογική έρευνα, προχωρά, έστω και με αργούς ρυθμούς και στη χώρα μας, αλλά και το ενδιαφέρον για τις δυνατότητες της ηλεκτρονικά υποστηριζόμενης ανάλυσης κειμένου αυξάνει, μαζί με το γενικότερο ενδιαφέρον για τις Ψηφιακές Σπουδές με τις πολλαπλές συνιστώσες τους, και η σχετική ελληνόφωνη βιβλιογραφία εμπλουτίζεται (Λαλιώτη 2022· Πατηνιώτης 2020· Δημητρούλια και Τικτοπούλου 2015).

Στόχος του παρόντος άρθρου δεν είναι να αποτελέσει μια γενική επισκόπηση του πεδίου των ψηφιακών λογοτεχνικών σπουδών, που καλύπτουν ένα ευρύ φάσμα ερευνών και μεθοδολογιών (Eve 2022· Earhart 2015· Δημητρούλια και Τικτοπούλου 2015· Siemens and Schreibman 2008). Αφορά την ηλεκτρονική ανάλυση κειμένου, όπως κι αν αυτή ονομάζεται στις διάφορες προσεγγίσεις –λεξικομετρία, λογομετρία, κειμενομετρία, ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου (βλ. για μια γενική επισκόπηση, Δημητρούλια, υπό δημ.)– στο ειδικό πεδίο της λογοτεχνίας. Εστιάζει στα ψηφιακά ή ηλεκτρονικά σώματα λογοτεχνικών κειμένων και στην ανάλυσή τους, εν είδει πρακτικού οδηγού. Λαμβάνει υπόψη το επίπεδο ψηφιοποίησης όπως διαμορφώνεται σήμερα στον ελληνικό χώρο, την υφιστάμενη ανάπτυξη

² Για μια επισκόπηση του γαλλικού χώρου, βλ. Etienne Brunet 2009, 2011 και 2016. Στον Brunet οφείλουμε χάριτες για τις πολυάριθμες βάσεις με έργα γάλλων συγγραφέων (περισσότερες από πενήντα), για τα εργαλεία που κατασκεύασε, όπως το Hyperbase, αλλά και τη συμβολή του στο μεγάλο γαλλικό σώμα αναφοράς Frantext. Για μια επισκόπηση των δημοσιεύσεών του, βλ. <http://ancilla.unice.fr/~brunet/PUB/articles.html>. Για τη μέθοδο της πρωτοπόρας ομάδας λογοτεχνικής έρευνας Hubert de Phalèse, βλ. Béhar 1995, όπως και τις συλλογικές δημοσιεύσεις της ομάδας στις εκδόσεις Klincksieck.

της γλωσσικής τεχνολογίας που αφορά την ελληνική γλώσσα, τις διεπιστημονικές συνέργειες που ήδη αναπτύσσονται και την αυξανόμενη ευχρηστία των λογισμικών ανάλυσης. Και ξεκινά από τη βασική παραδοχή ότι η χρήση των ηλεκτρονικών εργαλείων και των νέων ερμηνευτικών μεθόδων δημιουργεί, όπως η χρήση όλων των τεχνολογιών ιστορικά, νέα δεδομένα, αλλά επ' ουδενί θέτει υπό αίρεση την παραδοχή ότι η έρευνα καθορίζεται πάντοτε από τον ερευνητή και τα ερευνητικά του ζητούμενα.

Στο συγκεκριμένο πλαίσιο της διάδρασης ανθρώπου-υπολογιστή (Human-Computer Interaction, HCI), και αναλόγως με την ερευνητική οπτική και μεθοδολογία, τα κείμενα μπορούν, δυνάμει, να φωτιστούν διαφορετικά και κρυμμένες πτυχές τους να αναδειχθούν, τόσο στο επίπεδο της εκ του σύνεγγυς όσο και της εξ αποστάσεως ανάγνωσης. Εκ πρώτης όψεως, οι δυνατότητες συλλογής και επεξεργασίας δεδομένων του υπολογιστή δείχνουν να είναι ιδιαιτέρως χρήσιμες στην εξ αποστάσεως ανάγνωση και ερμηνεία. Η αλήθεια όμως είναι ότι ακριβώς επειδή η ερμηνεία βρίσκεται στο επίκεντρο της ερευνητικής πρακτικής στη λογοτεχνία, είναι πολλά τα μεθοδολογικά προαπαιτούμενα για την αποτελεσματική, ερμηνευτική ηλεκτρονική ανάλυση μαζικών λογοτεχνικών δεδομένων. Από την άλλη, η εκ του σύνεγγυς ερμηνευτική ανάλυση με ηλεκτρονική υποστήριξη, ως πιο οικεία πρακτική σε επίπεδο μεγεθών, ανοίγει, πάντα υπό συγκεκριμένες προϋποθέσεις, νέους δρόμους προσέγγισης των κειμένων, αναδιαμορφώνει τις ερευνητικές υποθέσεις και τα ζητούμενα και δημιουργεί ανάγκες για νέες προσεγγίσεις και νέες μεθόδους (Rockwell και Sinclair 2016).

Αφορμή για το παρόν άρθρο αποτέλεσε κατά πρώτον η ενασχόλησή μου με τα ψηφιακά εργαλεία από τα μέσα της δεκαετίας του '90, όταν, πέραν του χώρου της μεταφραστικής τεχνολογίας που με ενδιέφερε και με απασχολούσε ερευνητικά και πρακτικά,³ στο άλλο πεδίο ενδιαφέροντός μου, τη λογοτεχνία, είχα έρθει σε επαφή με το γαλλικό, κατά κύριο λόγο, μεθοδολογικό παράδειγμα ανάλυσης κειμένου, όντας σε επικοινωνία με την ομάδα Hubert de Phalèse. Και κατά δεύτερον, η συμμετοχή μου στις εργασίες της ελληνικής ψηφιακής υποδομής για τη γλωσσική τεχνολογία CLARIN:EL και της εθνικής υποδομής για τις Ψηφιακές Ανθρωπιστικές Επιστήμες και Τέχνες και για τη Γλωσσική Έρευνα και Καινοτομία 'Απολλωνίς', σε ένα περιβάλλον πολύ διαφορετικό από αυτό της δεκαετίας του '90, με πολύ περισσότερες δυνατότητες σε όλα τα επίπεδα. Στόχος των ευρωπαϊκών και συνακόλουθα των ελληνικών ψηφιακών υποδομών είναι η δημιουργία πόρων, εργαλείων και υπηρεσιών για την προαγωγή της έρευνας στις ανθρωπιστικές, κοινωνικές επιστήμες και τις τέχνες, η εκπαίδευση στη χρήση τους, η δημιουργία καλών πρακτικών και η συγκρότηση μιας διεπιστημονικής ερευνητικής κοινότητας, της οποίας τα μέλη θα συνομιλούν, θα συνεργάζονται και θα καινοτομούν. Οι στόχοι αυτοί αφορούν και την ερευνητική κοινότητα της λογοτεχνίας, με τα ιδιαίτερα χαρακτηριστικά της, και εξειδικεύονται σε διάφορες μεθοδολογίες, μία από τις οποίες είναι και η ηλε-

³ Πρακτικό αποτέλεσμα αυτής της ενασχόλησης υπήρξε το μάθημα «Εισαγωγή στην πληροφορική», που διδάχτηκε από το 2000 στο Διαπανεπιστημιακό, διατμηματικό μεταπτυχιακό πρόγραμμα μετάφρασης του ΕΚΠΑ, και αφορούσε τη μεταφραστική τεχνολογία, με πρώτο εργαλείο αναφοράς το ελληνικό και πολύ φιλικό Traid του Ινστιτούτου Επεξεργασίας του λόγου. Στη συνέχεια, διάφορα σχετικά μαθήματα δημιουργήθηκαν στο ΑΠΘ, διαθέσιμα στον κατάλογο μαθημάτων σχετικών με τις ψηφιακές ανθρωπιστικές επιστήμες <https://campus.dariah.eu/course-registry>.

κτρονική ανάλυση κειμένου (στο εξής ΗΑΚ) στις διάφορες μορφές της. Επιδίωξη του παρόντος σύντομου άρθρου είναι να δείξει συνοπτικά τι μπορεί να κάνει συγκεκριμένα ο υπολογιστής για τους ερευνητές στον χώρο της λογοτεχνίας στο επίπεδο της ανάλυσης κειμένου, με κάποια σχετικά παραδείγματα, και να παρακινήσει στην περαιτέρω εξερεύνηση αυτού του αναδυόμενου, στα καθ' ημάς, μεθοδολογικού παραδείγματος.

2. Τα ηλεκτρονικά σώματα κειμένων στην έρευνα της νεοελληνικής λογοτεχνίας

Μπορεί η ΗΑΚ να χρησιμοποιείται εδώ και πολλά χρόνια στην έρευνα της λογοτεχνίας σε άλλους γλωσσικούς και πολιτισμικούς χώρους, αλλά αυτό δεν σημαίνει ότι έχει καθιερωθεί ως μέθοδος στις λογοτεχνικές σπουδές, πολλώ δε μάλλον ότι δεν δημιουργεί αντιπαραθέσεις και έριδες, με αιχμή του δόρατος την ποσοτικοποίηση, που αδίκως εντέλει της καταλογίζεται, αφού τα όποια δεδομένα της ανάλυσης μόνο ποιοτικά μπορούν να αξιολογηθούν – στη σχέση ποσοτικών και ποιοτικών δεδομένων θα επανέλθω στη συνέχεια. Το μόνο βέβαιο πάντως είναι ότι όλοι οι ερευνητές πλέον, ασχέτως του βαθμού στον οποίο γνωρίζουν και αποδέχονται την ΗΑΚ ως μέθοδο, επηρεάζονται στην έρευνά τους από τα ψηφιοποιημένα σώματα κειμένων ελεύθερης πρόσβασης και από τις ποικίλες βάσεις, βιβλιογραφικές και/ή πλήρους κειμένου. Πραγματοποιούν αναζητήσεις, υιοθετώντας εν τοις πράγμασι πρακτικές αδιανόητες εκτός της ψηφιακής συνθήκης – ο ίδιος ο Παγκόσμιος Ιστός αποτελεί, άλλωστε, υπό μία έννοια, ένα τεράστιο σώμα κειμένων, παρότι ο ορισμός αυτός πρέπει να νοείται με πολλούς και σημαντικούς περιορισμούς. Έτσι, ο αναπόδραστος –όπως ελπίζουμε, τουλάχιστον– πολλαπλασιασμός των (ελεύθερα προσβάσιμων) ψηφιοποιημένων ή ψηφιακών, εξ αρχής συγκροτημένων ψηφιακά δηλαδή, σωμάτων κειμένων, σε συνδυασμό με την απλοποίηση των εργαλείων ανάλυσης, αναμένεται λογικά να υποστηρίξουν την ευρύτερη υιοθέτησή της, σε συγκεκριμένα ερευνητικά έργα.⁴ Η αποτελεσματικότητα και η καινοτομία της όμως στο πεδίο της έρευνας της λογοτεχνίας συναρτάται, όπως άλλωστε και στα άλλα πεδία, προς τη σαφή αντίληψη της λογικής των εργαλείων ως μέσων και τη στάθμιση των ποικίλων παραμέτρων που καθορίζουν τη χρήση τους.

2.1. Ψηφιοποιημένα σώματα κειμένων: τα αρχεία του τύπου, περιορισμοί και επικαθορισμοί

Στην περίπτωση ενός έτοιμου ΗΣΚ, ενός σώματος αναφοράς δηλαδή ελεύθερα διαθέσιμου στο διαδίκτυο, το οποίο επιπροσθέτως δεν έχει κωδικοποιηθεί για να επιτρέπει προηγμένες μορφές αναζήτησης, όπως τα αρχεία του τύπου που χρησιμοποιούνται ευρύτατα στην έρευνα της νεοελληνικής λογοτεχνίας, ο χρήστης δεν έχει πρόσβαση σε κειμενικά δεδομένα παρά μόνο σε μορφή εικόνας. Έχει τη δυνατότητα εξερεύνησής τους, η αξία της οποίας είναι φυσικά αδιαμφισβήτητη, με συγκεκριμέ-

⁴ Η διάκριση μεταξύ ψηφιοποιημένων και ψηφιακών σωμάτων κειμένων θέτει ποικίλα ζητήματα, εξού και προτιμούμε τον όρο Ηλεκτρονικά Σώματα Κειμένων (ΗΣΚ).

νους όρους, τους οποίους θέτει η κατασκευή και οργάνωση του ΗΣΚ. Μπορεί με άλλα λόγια να εντοπίσει συγκεκριμένα στοιχεία, μέσα στο πλαίσιο που ορίζουν οι περιορισμοί της αναζήτησης και οι παράμετροι που αυτοί εισάγουν στην έρευνα. Θα αντιλέξει κανείς ότι αποθηκεύοντας κάποια φύλλα ή σελίδες φύλλων, μπορεί να τα μετατρέψει σε κειμενικά δεδομένα στη συνέχεια. Αλλά η διαδικασία αυτή όχι μόνο είναι επίπονη, με τα προβλήματα που παρουσιάζει η οπτική αναγνώριση χαρακτήρων, ειδικά για προγενέστερες μορφές της ελληνικής, αλλά και ελάχιστα χρήσιμη, στον βαθμό που αφορά μερικά δεδομένα και μόνο, ενώ η αξία χρήσης των μεγάλων αρχείων έγκειται στην αναζήτηση σε μαζικά δεδομένα.

Η χρήση ανάλογων ΗΣΚ αναφοράς, η οποία είναι όλο και περισσότερο διαδεδομένη, παρότι λειτουργεί σαφέστατα υποστηρικτικά στην ποσοτική έρευνα σε μια ποιοτική προοπτική, μας δίνει επίσης ένα μέτρο της λογικής του μέσου, ειδικά ως προς την αξιοπιστία και τη χρησιμότητα των ποσοτικών δεδομένων που αντλούνται από αυτά. Στη Γλωσσολογία Σωμάτων Κειμένων, γίνεται μια σημαντική διάκριση ανάμεσα στην προσέγγιση με βάση τα ΗΣΚ (corpus-based analysis) και στην προσέγγιση με γνώμονα τα ΗΣΚ (corpus-driven analysis). Μεταφέροντας κανείς τη διάκριση αυτή στην έρευνα της ιστορίας της λογοτεχνίας, λόγου χάρη, με στοιχεία από τα αρχεία στον τύπο, αντιλαμβάνεται αμέσως ότι η έρευνα αυτή στα έντυπα, ασχέτως του ειδικού ερωτήματος και του οριζόμενου χρονικού εύρους της αναζήτησης, ήταν πάντοτε, σε έναν βαθμό τουλάχιστον, μια έρευνα με γνώμονα τα σώματα κειμένων. Ο ερευνητής που διάβαζε τις εφημερίδες ή τα περιοδικά σε έντυπη μορφή, όσο κι αν ξεκινούσε από ένα ερώτημα ή μια υπόθεση και αναζητούσε στοιχεία σε ένα συγκεκριμένο corpus, εξετίθετο διαρκώς σε γνώση που μπορούσε να τον ανακατευθύνει, που δημιουργούσε τις προϋποθέσεις για διαφορετική πλαισίωση των υποθέσεων, αλλά και τη διατύπωση νέων, συχνά πιο σύνθετων ερωτημάτων.

Όταν ο ερευνητής προσεγγίζει τα μαζικά δεδομένα μιας ψηφιοποιημένης εφημερίδας, λόγου χάρη, το πλαίσιο της έρευνας αφενός διευρύνεται και αφετέρου, όσο κι αν φαίνεται παράξενο, περιορίζεται, για πολλούς λόγους. Κατά πρώτον, τα ψηφιοποιημένα αρχεία επιτρέπουν τις αναζητήσεις με βάση μια ‘ανάγνωση’ του κειμενικού υλικού, η οποία γίνεται αυτόματα και, στην περίπτωση των μαζικών δεδομένων, δεν είναι τις περισσότερες φορές, για λόγους εντελώς πρακτικών, επαρκώς ελεγμένη και επεξεργασμένη. Έτσι, ακόμη και για απλές αναζητήσεις ονομάτων, δεν είναι προφανές ότι κανείς θα λάβει το σύνολο των αναφορών, είτε επειδή μπορεί να υπάρχουν περιορισμοί στη δυνατότητα εμφάνισης αποτελεσμάτων, είτε απλώς επειδή δεν ορίζονται επαρκείς ή ακριβείς περιορισμοί για την ακριβή ταυτοποίηση των οντοτήτων. Αρκεί κανείς να δει πόσα άσχετα αποτελέσματα εμφανίζονται σε κάθε αναζήτηση σε αρχεία εφημερίδων, επειδή ακριβώς το σύστημα εκλαμβάνει συγκεκριμένους τύπους ως συμβατούς προς αυτήν, ενώ δεν έχουν καμία απολύτως σχέση με το ζητούμενο. Πολύ πιο σημαντική είναι όμως η –γνωστή– αδυναμία του εργαλείου να διαχειριστεί ακόμη και στοιχειωδώς τη σημασία. Είναι δεδομένο ότι πάντοτε ο υπολογιστής μάς δίνει, σε επίπεδο αναζήτησης, ό,τι ακριβώς του ζητάμε και δεν έχει δυνατότητα πλαισίωσης του ερωτήματος. Εάν λοιπόν αναζητεί κανείς τις αναφορές στον Σεφέρη σε μια χρονική περίοδο, θα λάβει πίσω μόνο τα κείμενα που αναφέρουν το όνομά του – ενδεχομένως μαζί με κάμποσα που δεν θα τον αφορούν, όπως όλοι όσοι κάνουμε έρευνα σε ψηφιακά αρχεία τύπου γνωρίζου-

με πολύ καλά, είτε γιατί οι Σεφέρηδες είναι πολλοί, είτε ακόμη διότι ανακαλούνται λιγότερο ή περισσότερο συναφείς φωνητικά τύποι. Το κυριότερο όμως είναι ότι ο Σεφέρης μπορεί να αναφέρεται σε ένα κείμενο χωρίς να ονομάζεται, ως ο διπλωμάτης ποιητής, ο νομπελίστας, ο εισηγητής του αγγλόφωνου μοντερνισμού ή ο ρυθμιστής του λογοτεχνικού πεδίου του ελληνικού ποιητικού μοντερνισμού και ούτω καθεξής. Είναι προφανές ότι αυτές οι αναφορές, τις οποίες ο ερευνητής θα εντόπιζε ακόμη και διατρέχοντας απλώς τα κείμενα της εφημερίδας, θα απουσιάζουν από μια αναζήτηση με λέξη-κλειδί το όνομα του Σεφέρη, κάτι που θα έχει αντίκτυπο όχι ποσοτικό αλλά ποιοτικό στη χρήση των αποτελεσμάτων.

Σε έναν δεύτερο χρόνο, οι αναφορές στον Σεφέρη καθαυτές δεν είναι εξ ορισμού σημαντικές. Επομένως, ο ποσοτικός δείκτης είναι μικρού –αν και καθόλου αμελητέου– ενδιαφέροντος, εάν δεν συναρτάται προς μια ερμηνευτική, ερευνητική υπόθεση. Μπορεί η ποσοτική αναφορά να δηλώνει την παρουσία ενός προσώπου, εν προκειμένω του Σεφέρη, στη δημόσια συζήτηση, αλλά μόνο η ποιοτική ανάλυση των αποτελεσμάτων μπορεί να οδηγήσει στην αξιοποίησή τους για την έρευνα. Για παράδειγμα, ενδιαφέρει αν οι αναφορές αυτές περιλαμβάνονται σε κριτικές, αν αφορούν το έργο ή το πρόσωπο και ποιες πτυχές του, με ποιο τρόπο το πλαισιώνουν και πώς τα στοιχεία που προκύπτουν εντάσσονται στην υφιστάμενη έρευνα και την εμπλουτίζουν – αναλόγως με το ερευνητικό ερώτημα, στο πεδίο της ιστορίας της λογοτεχνίας και της πρόσληψης σε συγκεκριμένα πλαίσια. Εκατό δημοσιογραφικές αναφορές στο βραβείο Νόμπελ κατά πάσα πιθανότητα απλώς επιβεβαιώνουν τη σημασία που είχε για την εθνική υπερηφάνεια, αλλά μόνο η προσεκτική ανάγνωση των αναφορών αυτών μπορεί να εκμαιεύσει κάποια, πολύ λιγότερα, ποιοτικά στοιχεία σχετικά με τη συζήτηση για το Νόμπελ στη δημόσια σφαίρα και/ή στο λογοτεχνικό πεδίο. Πενήντα συναναφορές στον Σεφέρη και τον Έλιοτ μπορεί να μην προσθέτουν τίποτα στα όσα ήδη γνωρίζουμε για τη σχέση τους. Αντίθετα, μία και μόνη κριτική μπορεί να παρουσιάζει μεγαλύτερο ενδιαφέρον από ό,τι όλα τα προηγούμενα κείμενα μαζί, στον βαθμό που θα ρίξει νέο φως σε λογοτεχνικά δίκτυα και/ή ανταγωνισμούς του λογοτεχνικού πεδίου σε μια συγκεκριμένη περίοδο. Η ανάσυρση των αποτελεσμάτων όμως κατευθύνει εξ ορισμού τον ερευνητή σε μια προσέγγιση με βάση τα ΗΣΚ: επιδιώκει να επιβεβαιώσει ή να καταρρίψει κάποια υπόθεση με βάση το ΗΣΚ και με δεδομένους τους περιορισμούς του. Άρα η λογική το εργαλείου, ειδικά καθώς δεν υπάρχει άμεση πρόσβαση στα κειμενικά δεδομένα ως σύνολο, προσανατολίζει τη μεθοδολογία της έρευνας. Όπως επισημαίνει ο Lev Manovich, η λογική του υπολογιστή επηρεάζει προφανώς «την παραδοσιακή πολιτισμική λογική του μέσου» (2001, 46), δηλαδή εν προκειμένω των αρχείων.

2.2. Ψηφιακά σώματα κειμένων με προηγμένες λειτουργίες αναζήτησης και εμφάνισης των δεδομένων: το ψηφιακό εργαλείο της Ανεμόσκαλας

Ένα παράδειγμα ΗΣΚ, που δεν παρέχει πρόσβαση στα δεδομένα, υπό την έννοια ότι δεν μπορεί κανείς να τα πάρει όλα ή κάποια από αυτά με εύκολο τρόπο και να τα χρησιμοποιήσει σε άλλα έργα, είναι το γνωστό και εξαιρετικό αρχείο [TLG](#), του πανεπιστημίου Irvine της Καλιφόρνιας, που συγκεντρώνει τα ελληνικά κείμενα από την εποχή του Ομήρου ως την πτώση του Βυζαντίου το 1453. Στον χώρο της νεοελληνικής φιλολογίας, η [Ανεμόσκαλα](#) είναι ένα πολύ ενδιαφέρον συλλογικό και διεπι-

στημονικό έργο του Κέντρου Ελληνικής Γλώσσας, που παρουσιάζεται στην ιστοσελίδα της Πύλης για την Ελληνική Γλώσσα ως «ένα ψηφιακό εργαλείο [που] σχεδιάστηκε για να διευκολύνει τη μελέτη της νεοελληνικής ποίησης και να ενθαρρύνει την άμεση και αδιαμεσολάβητη ανάγνωση του ποιητικού λόγου. Για τον σκοπό αυτό το δυναμικό περιβάλλον στο διαδίκτυο επιτρέπει την *ad hoc* δημιουργία Συμφραστικών πινάκων λέξεων των ποιητικών κειμένων, τις αναζητήσεις λέξεων στο σύνολο των ποιητικών έργων και για κάθε ποιητή χωριστά, τη μετάβαση από τις λέξεις στα ποιήματα και αντιστρόφως κ.ά. [...] Τα κείμενα έχουν εκδοθεί ψηφιακά σύμφωνα με το διεθνές πρότυπο TEI, που προσφέρει σύνθετες επιλογές κωδικοποίησης του κειμένου τόσο για την ψηφιακή αναπαράστασή του όσο και για την επισημείωσή του. Η βασική διαφορά του ψηφιακού corpus σε σχέση με μια συμβατική έκδοση είναι οι σύνθετες δυνατότητες αναζήτησης που δίνονται στον χρήστη, μέσω της ένταξης του κωδικοποιημένου κειμένου σε σχεσιακή βάση δεδομένων, η οποία λειτουργεί ως εργαλείο δημιουργίας καταλόγων λεκτικών τύπων ενταγμένων στα συμφραζόμενά τους, ό,τι καταχρηστικά στη συνέχεια αποκαλείται Συμφραστικός Πίνακας Λέξεων [=λεκτικών τύπων] (ορολογία που παραπέμπει σε έντυπες εκδόσεις, διαφορετικής μεθοδολογίας)».

Το εργαλείο υποστηρίζει την ανάγνωση και μελέτη του σχολικού ποιητικού κανόνα, με βιβλιογραφική πλαισίωση και στηρίζεται στις βασικές λειτουργίες ανάλυσης κειμένου, όπως η αναζήτηση λεξικών τύπων στο συγκεκριμένο τους, με ταυτόχρονη πρόσβαση στο πλήρες κείμενο, όπου αυτό είναι εφικτό, με βάση τα πνευματικά δικαιώματα. Για τη δημιουργία του, τα έργα των ποιητών κωδικοποιήθηκαν με το σύστημα TEI, που παρουσιάζει ιδιαίτερο ενδιαφέρον για την έκδοση λογοτεχνικών έργων σε ψηφιακή μορφή (Τικτοπούλου 2015). Η σήμανση και επισημείωση των κειμένων και η οργάνωσή τους σε βάση δεδομένων επιτρέπει ποικίλες αναζητήσεις σε υποσώματα κειμένων, που μπορούν να χρησιμεύσουν είτε σε ερευνητικό είτε σε εκπαιδευτικό πλαίσιο, ενώ θέτουν και περαιτέρω ερωτήματα σχετικά με τις δυνατότητες (*affordances*) των ψηφιακών εργαλείων και τα όριά τους όσο και για τις ερευνητικές προοπτικές της χρήσης τους.

Χωρίς να μπορούμε εδώ στη συζήτηση περί του αν τα κείμενα του ΗΣΚ της Ανεμόσκαλας είναι ψηφιοποιημένα ή εξαρχής ψηφιακά συγκροτημένα όπως διατείνεται η Ακριτίδου (2014, 3), μια θέση κατά τη γνώμη μου συζητήσιμη στον βαθμό που η ψηφιοποίηση, με οποιονδήποτε τρόπο, των δεδομένων αποτελεί τη βάση για την έκδοση του ψηφιακού κειμένου αλλά και τη (μερική εν προκειμένω) επεξεργασία του σε επίπεδο ανάλυσης, σημειώνω τη σημασία του έργου και σε μεθοδολογικό επίπεδο έκδοσης και οργάνωσης σωμάτων λογοτεχνικών κειμένων.

The screenshot shows the website greek-language.gr with a search for the word 'φωτιάς'. The left sidebar lists authors like Μανώλης Αναγνωστάκης and Αριστοτέλης Βαλαωρίτης. The main area displays the word 'φωτιάς' and its occurrences in various texts, such as 'ΠΡΟΜΗΘΕΑΣ' and 'ΕΙΣΑΓΩΓΗ'.

Εικόνα 1. Ανεμόσκαλα, Σώματα κειμένων και συμφραστικοί πίνακες νεοελλήνων ποιητών, <https://bit.ly/3WkHrHB> (λέξη-κλειδί: φωτιά, αναζήτηση στο έργο των Αναγνωστάκη, Βάρναλη, Παλαμά, Ρίτσου)

Στο εξωτερικό, οι ανάλογες προσπάθειες ενσωματώνουν πολύ περισσότερες λειτουργίες της ΗΑΚ, όπως στο ακόλουθο παράδειγμα Συμφραστικού Πίνακα Λέξεων (ΣΠΛ) του Balzac:

The screenshot shows the eBalzac website with a search for the word 'argent'. The left sidebar lists authors like Honoré de Balzac. The main area displays the word 'argent' and its occurrences in various texts, such as 'LA MAISON DU CHAT-QUI-PELOTE'.

Εικόνα 2. Συμφραστικός πίνακας απάντων του Balzac, <https://marat.uchicago.edu/philologic/eBalzac/navigate/19/table-of-contents> (λέξη-κλειδί: argent).

Ας ελπίσουμε ότι σύντομα θα έχουμε ανάλογες εξελίξεις και στην καθ' ημάς φιλολογία.

2.3. Ηλεκτρονικά Σώματα Κειμένων και ηλεκτρονική ανάλυση

Η ΗΑΚ βασίζεται στα ΗΣΚ και σε συγκεκριμένα εργαλεία γλωσσικής τεχνολογίας, αλλά αυτό δεν σημαίνει ότι ταυτίζεται με τη Γλωσσολογία Σωμάτων Κειμένων (ΓΣΚ) ή με την υπολογιστική γλωσσολογία. Προφανώς η ΓΣΚ, οι μέθοδοι και οι εφαρμογές της κατέχουν σημαντική θέση στην ΗΑΚ και η υπολογιστική γλωσσολογία αποτελεί προνομιακό πεδίο αναφοράς της. Όπως όμως επισημαίνουν οι Rockwell και Sinclair, σημασία έχει η οπτική της ανάλυσης, η οποία στην περίπτωση των λογοτεχνικών σπουδών δεν μπορεί παρά να είναι ερμηνευτική και να εντάσσεται στην υφιστάμενη ερμηνευτική παράδοση, σε διάφορες χώρες και γλώσσες, με στόχο την ενίσχυση και διεύρυνσή της προς όφελος της μελέτης των κειμένων. Αποτελεί δηλαδή κύριο ζητούμενο η σαφής εικόνα του πώς μπορούν οι υπολογιστές να υποστηρίξουν τον ερευνητή στην ερμηνευτική του πρακτική και δεδομένο ότι το νέο στοιχείο που η ΗΑΚ κομίζει στη μεθοδολογία της έρευνας της λογοτεχνίας είναι μεθοδολογικό. Πρόκειται για έναν συλλογικό τρόπο δουλειάς, μια διεπιστημονική, διαλογική ερευνητική πρακτική που προκύπτει τόσο από τη συνεργασία ερευνητών από διαφορετικές περιοχές όσο και από τη διάδραση των ερευνητών με τα εργαλεία. Λόγω του ότι τα ίδια τα εργαλεία έχουν τη δική τους λογική, διαμορφωμένη από τη χρήση τους στις θετικές επιστήμες, οι Rockwell και Sinclair επέλεξαν, λόγου χάρη, να δημιουργήσουν τα δικά τους εργαλεία για την ανάγνωση και μελέτη των κειμένων, τα [Voyant Tools](#), των οποίων η ανάπτυξη κινείται στην κατεύθυνση της ερμηνευτικής τους οπτικής και της, κατά το δυνατόν, υπέρβασης των περιορισμών που προκύπτουν από την πληροφορική προέλευσή τους (2016, 6-23). Έτσι, η ΗΑΚ μπορεί να συνδυάσει ποικίλα εργαλεία, διαφορετικής προέλευσης και χρήσης, με άξονα την ερμηνευτική οπτική, με την προϋπόθεση του ελέγχου των εργαλείων αυτών ως προς τα προϋποτιθέμενά τους και της συνείδησης της – αναπόδραστης – επίδρασής τους στη διαδικασία της έρευνας.

Προϋπόθεση για την ΗΑΚ αποτελούν τα ΗΣΚ, τα οποία δημιουργούνται με συγκεκριμένους όρους και διαφοροποιούνται από τις συλλογές κειμένων και τα απλά ψηφιοποιημένα κείμενα, που μπορεί κανείς να βρει στο διαδίκτυο – και τα οποία μπορούν, ωστόσο, να αποτελέσουν, με την κατάλληλη επεξεργασία, την πρώτη ύλη για ένα ΗΣΚ, ειδικά διαμορφωμένο για ανάλυση κειμένου. Ένα κλασικό παράδειγμα συλλογής αποτελεί το [gutenberg.org](#), ένα επίσης κλασικό παράδειγμα ψηφιοποιημένου κειμένου τα έργα του Αλέξανδρου Παπαδιαμάντη, που είναι αναρτημένα στη σελίδα της [Εταιρείας Παπαδιαμαντικών σπουδών](#), σε μορφή κειμένου εντούτοις και όχι πλέον εικόνας – ένα σημαντικό βήμα για περαιτέρω επεξεργασίες. Μια πρώτη και κύρια διαφορά τους σχετίζεται με τη δυνατότητα εξερεύνησής τους με ηλεκτρονικά εργαλεία, γεγονός που καθορίζει τη μορφή των δεδομένων τους – παρότι όλο και περισσότερα λογισμικά ανάλυσης κειμένου, των Voyant tools συμπεριλαμβανομένων, δέχονται πλέον κείμενα σε πολλά και διαφορετικά μορφότυπα. Το μορφότυπο που επιτρέπει εντούτοις τις πιο προηγμένες επεξεργασίες κειμένων δεδομένων παραμένει το απλό κείμενο (txt), και μάλιστα με συγκεκριμένη κωδικοποίηση (UTF-8). Εάν για κάποιον νεοεισερχόμενο στον χώρο της ΗΑΚ αυτό ακούγεται ήδη πολύπλοκο, να σημειωθεί απλώς ότι η δυνατότητα μετατροπής ενός κειμένου γραμμένου στο Word σε αυτή τη μορφή παρέχεται ως δυνατότητα κατά την αποθήκευσή του.

Το πιο σημαντικό στοιχείο όμως όσον αφορά ένα ειδικής χρήσης, Do-It-Yourself (DIY) ΗΣΚ, που θα χρησιμοποιηθεί σε ΗΑΚ αφορά το ίδιο το ΗΣΚ, τον τύπο του, τον τρόπο της συγκρότησής του, με αναφορά στο κατά περίπτωση ερευνητικό ερώτημα. Είναι σημαντικοί οι όροι της κατασκευής του, η επιλογή των κειμένων καθαυτή, η έκδοση των κειμένων που χρησιμοποιείται, η ποιότητα της ηλεκτρονικής αναπαραγωγής της, δηλαδή των δεδομένων της έρευνας, η αντιπροσωπευτικότητα και η στάθμιση των κειμένων με αναφορά στο ερευνητικό ερώτημα, η αναπαραστάση και η ενδεχόμενη σήμανση των κειμένων. Το ζητούμενο είναι η αξιοπιστία των αποτελεσμάτων της ανάλυσης, ώστε να μπορούν να υποστηρίξουν την ποιοτική, ερμηνευτική μελέτη των κειμένων, που αποτελεί το ένα και μόνο ζητούμενο στην έρευνα της λογοτεχνίας.

Στο επίπεδο της επιλογής, οι όροι της έρευνας δεν διαφοροποιούνται σε σχέση με την παραδοσιακή, φιλολογική έρευνα. Ο ερευνητής επιλέγει συγκεκριμένα κείμενα και τα μελετά για να απαντήσει σε ένα ή περισσότερα ερωτήματα. Είτε συγκροτεί ένα ΗΣΚ για να ελέγξει μια υπόθεση ή μια θεωρητική παραδοχή (προσέγγιση με βάση τα σώματα κειμένων)· ή προκειμένου να εντοπίσει στοιχεία που θα τον καθοδηγήσουν στη διατύπωση ερευνητικών υποθέσεων (προσέγγιση με γνώμονα τα σώματα κειμένων). Στην πράξη, οι δύο πρακτικές συχνά συνδυάζονται, αφού από τη μελέτη ενός ΗΣΚ μπορεί να προκύψει η ανάγκη για προσφυγή σε ευρύτερο λογοτεχνικό κειμενικό υλικό και/ή σε ποικίλες συγκριτικές προσεγγίσεις. Η ιδιαιτερότητα των λογοτεχνικών έργων, ως παγιωμένων, κλειστών κειμένων με ιδεολογική και αισθητική λειτουργία, καθορίζει και τη χρήση της μεθόδου. Εξού και η σημασία της επιλογής της έκδοσης, κι ένα κατεξοχήν παράδειγμα για τις νεοελληνικές σπουδές αποτελεί η σολωμική ποίηση.

Πριν εξετάσουμε τις βασικές λειτουργίες της ΗΑΚ, με ειδικότερη αναφορά στα εργαλεία Voyant Tools, των οποίων η διεπαφή αναμένεται να είναι άμεσα διαθέσιμη και στα ελληνικά, να επισημάνουμε το ζήτημα της διαθεσιμότητας ηλεκτρονικών κειμένων, η οποία συναρτάται παντού προς τα πνευματικά δικαιώματα, αλλά ειδικά στα ελληνικά έχει επίσης να κάνει και με άλλους παράγοντες, όπως οι δυσκολίες που δημιουργεί η γλώσσα των παλαιότερων κειμένων, οι διάφορες μορφές καθαρεύουσας. Παρότι έχει υπάρξει μεγάλη πρόοδος στην ψηφιοποίηση κειμένων της παλαιότερης νεοελληνικής γραμματείας, παραμένει το γεγονός ότι τα ψηφιοποιημένα αυτά κείμενα δεν μπορούν, τις περισσότερες φορές τουλάχιστον, να χρησιμοποιηθούν, αποτελεσματικά, στην ΗΑΚ, λόγω περιορισμών των εργαλείων.

Η επιλογή των κειμένων προϋποθέτει τον έλεγχο των στοιχείων και της ποιότητας των διαθέσιμων ηλεκτρονικών εκδόσεων και στον βαθμό που αυτές δεν είναι διαθέσιμες την ψηφιοποίηση των αναγκαίων εκδόσεων με ένα λογισμικό οπτικής αναγνώρισης χαρακτήρων (OCR). Ορισμένα από αυτά λειτουργούν πλέον πολύ ικανοποιητικά για τα νέα ελληνικά, ενώ τα προβλήματα παραμένουν για τις παλαιότερες μορφές της ελληνικής –παρότι και στον χώρο αυτό σημειώνονται σημαντικές εξελίξεις, με την εκπαίδευση των λογισμικών. Σε όλες τις περιπτώσεις, η τελική επιμέλεια του ψηφιοποιημένου κειμένου είναι απαραίτητη και ενδεχομένως ιδιαιτέρως βαρύνουσα για το τελικό αποτέλεσμα της έρευνας, ενώ η ποιότητα της ψηφιοποίησης θεωρείται, σε ένα επίπεδο τουλάχιστον, απαραίτητη. Για να γίνει κατανοητό αυτό το «σε ένα επίπεδο», ας σημειωθεί ενδεικτικά ότι οι λανθασμένοι χαρακτήρες παρεμποδίζουν την αναζήτηση λεξικών τύπων –κι αυτό είναι ένα σημαντικό

πρόβλημα στο μη επιμελημένο OCR που επιτρέπει την αναζήτηση σε συλλογές αρχείων εικόνων-, ενώ η λανθασμένη στίξη υπονομεύει, σε μέγιστο βαθμό, την υφολογική μελέτη – τη μελέτη του μήκους και του ρυθμού των προτάσεων, λόγου χάρη. Τα προβλήματα τονισμού μπορούν να επηρεάσουν, υπό συγκεκριμένες συνθήκες, τα αποτελέσματα. Ειδικά λογισμικά μπορούν να βοηθήσουν στον καθαρισμό των ψηφιοποιημένων κειμένων, ειδικά όταν έχουμε μαζικά δεδομένα που δεν επιτρέπουν τη χειρωνακτική επιμέλεια, πάντα με ένα περιθώριο αστοχίας.

Στα ελληνικά, τα ελεύθερα διαθέσιμα, αξιόπιστα ηλεκτρονικά λογοτεχνικά κείμενα, όπως τα έργα του Παπαδιαμάντη ή το έργο ορισμένων ποιητών της Ανεμόσκαλας, είναι πολύ περιορισμένα, όπως άλλωστε και το ψηφιακό περιεχόμενο εν γένει, σε σύγκριση με το περιεχόμενο στις κεντρικές γλώσσες και την υπερκεντρική αγγλική. Ακόμη πιο περιορισμένα είναι τα κείμενα που μπορούν να χρησιμοποιηθούν άμεσα σε ΗΑΚ. Το ίδιο ισχύει και για άλλες μικρές και λιγότερο διαδεδομένες γλώσσες, γεγονός που δυσκολεύει την έρευνα και στο πεδίο της συγκριτικής γραμματολογίας. Η εν λόγω συνθήκη υπογραμμίζει την ακόμη μεγαλύτερη αναγκαιότητα για τον ελληνικό ερευνητικό χώρο του διαμοιρασμού των πόρων, είτε ελεύθερα είτε υπό όρους, εφόσον υπάρχουν πνευματικά δικαιώματα. Τα αποθετήρια των ερευνητικών υποδομών φιλοδοξούν να προαγάγουν αυτόν τον διαμοιρασμό, που θα δώσει τεράστια ώθηση στην έρευνα. Ακόμη και αν ένα ΗΣΚ δεν μπορεί να είναι ελεύθερα προσβάσιμο, η ανάρτηση των μεταδεδομένων του και η δυνατότητα επικοινωνίας με τον ερευνητή για την απόκτησή του αποτελεί ήδη ένα σημαντικό βήμα για τη συγκρότηση μιας δεξαμενής πολύτιμων ψηφιακών κειμενικών πόρων. Ακόμη πιο χρήσιμη είναι η ανάρτηση στα αποθετήρια ελεύθερων ΗΣΚ σε συγκεκριμένη μορφή που επιτρέπει την ανάλυσή τους με ποικίλα εργαλεία. Με τη λογική αυτή, στο αποθετήριο ΑΠΘ του Clarin-Απολλωνίς έχουν αναρτηθεί τα έργα του Παπαδιαμάντη, συγκροτημένα σε σώμα κειμένων αδόμετων δεδομένων, χωρίς σήμανση και επισημείωση δηλαδή, και μορφότυπο UTF-8. Η δημιουργία μιας δεξαμενής νεοελληνικών λογοτεχνικών κειμένων, η οποία θα διευρύνεται διαρκώς και, ιδανικά, θα απόκειται σε ένα σημείο στο διαδίκτυο είναι εξίσου αναγκαία με την περαιτέρω ανάπτυξη των εργαλείων που αφορούν την ελληνική γλώσσα και την εκπαίδευση των ερευνητών στην (αποτελεσματική) χρήση τους. Τα ΗΣΚ αυτά μπορούν να χρησιμοποιηθούν άλλωστε για σκοπούς διαφορετικούς από αυτούς για τους οποίους συγκροτήθηκαν. Ένα παράδειγμα αποτελεί το ΗΣΚ των πρωτότυπων έργων του Αλέξανδρου Παπαδιαμάντη στο αποθετήριο Clarin-ΑΠΘ, μαζί με το ΗΣΚ ορισμένων βέβαιων μεταφράσεών του. Παρότι τα εν λόγω ΗΣΚ είχαν χρησιμοποιηθεί στο πλαίσιο υφομετρικής ανάλυσης για την απόδοση πατρότητας σε ανώνυμες μεταφράσεις που εικάζονταν ότι του ανήκαν, μπορούν να υποστηρίξουν ποικίλα άλλα ερευνητικά ερωτήματα.

Οι βασικές λειτουργίες των εργαλείων ΗΑΚ, τα οποία έχουν διαβαθμισμένες δυνατότητες, είναι οι λειτουργίες του Συμφραστικού Πίνακα Λέξεων (ΣΠΛ, concordance), ως καταλόγου όλων των λέξεων (και όχι των λημμάτων, αλλά των λεξικών τύπων) του κειμένου και των εμφανίσεών τους στο ΗΣΚ (κατάλογος συχνότητας, wordlist), είτε αλφαβητικά είτε με βάση τον αριθμό των εμφανίσεων αυτών. Είναι αυτονόητο ότι οι περισσότερες, αριθμητικές, εμφανίσεις αφορούν τις γραμματικές, λειτουργικές λέξεις. Ενώ στην ΗΑΚ άλλων ειδών λόγου οι λειτουργικές αυτές λέξεις μπορούν να παραλειφθούν, με τη δημιουργία καταλόγου αποκλειόμενων, ε-

ξαιρούμενων λέξεων (stopwords) που ενσωματώνεται στην εφαρμογή, στην έρευνα της λογοτεχνίας οι συγκεκριμένες λέξεις μπορεί να έχουν σημαντικό ευρετικό ενδιαφέρον (Culpeper 2002 και 2009) – ειδικά σε μια έρευνα με γνώμονα τα ΗΣΚ. Από την άλλη, σε έρευνες που αφορούν πραγματολογικά στοιχεία στο ΗΣΚ ενός συγγραφέα ή σε ένα ευρύτερο ΗΣΚ μιας περιόδου, μπορεί οι λέξεις αυτές να μπορούν να εξαιρεθούν χωρίς πρόβλημα και η εξαίρεσή τους να υποβοηθήσει μάλιστα την ανάλυση.

Το ζήτημα είναι ότι διάφοροι έτοιμοι κατάλογοι αποκλειόμενων λέξεων είναι διαθέσιμοι στις μεγάλες γλώσσες και ειδικά στα αγγλικά, αλλά όχι στα ελληνικά. Αυτό σημαίνει ότι ο ερευνητής τούς δημιουργεί μόνος του, τους αποθηκεύει και τους χρησιμοποιεί ή τους αναδιαμορφώνει κατά περίπτωση. Ένας τρόπος δημιουργίας σχετικών καταλόγων για τα νεοελληνικά κείμενα είναι η παρακολούθηση της μικρής γραμματικής του Τριανταφυλλίδη: οριστικό άρθρο, αόριστο άρθρο, αριθμητικά, αντωνυμίες, επιρρήματα, προθέσεις, σύνδεσμοι, επιφωνήματα, σε συνδυασμό με τη γραμματική. Για παλαιότερα κείμενα, μια αντίστοιχη δουλειά μπορεί να γίνει με βάση μια γραμματική της καθαρεύουσας, όπως του Τζάρτζανου. Ο κατάλογος αυτός, εφόσον δημιουργηθεί και είναι χωρισμένος σε δημοτική και καθαρεύουσα και σε μέρη του λόγου, μπορεί να παραμετροποιηθεί και να χρησιμοποιηθεί σε πολλές, διαφορετικές περιστάσεις, με διαφορετικές ανάγκες αποκλεισμού και συμπερίληψης όρων. Η δημιουργία του ωστόσο συναρτάται και προς άλλα προβλήματα που παρουσιάζονται στην ΗΑΚ στα ελληνικά, όπως η λημματοποίηση. Εάν δεν υπάρχει η δυνατότητα λημματοποίησης στα κείμενα, ο εν λόγω κατάλογος θα πρέπει να περιλαμβάνει όλους τους τύπους των κλιτών μερών του λόγου. Η λημματοποίηση, όμως, δεν υποστηρίζεται για παλαιότερες μορφές της ελληνικής, ενώ ως δυνατότητα εν γένει παρέχεται από περιορισμένο αριθμό εργαλείων. Αυτό καθιστά πιο σημαντική την ύπαρξη ενός ευρέος, γενικού καταλόγου, αλλά και όλων των διαθέσιμων ειδικών καταλόγων που μπορούν να εμπλουτιστούν ή να χρησιμοποιηθούν σε διαφορετικά συμφραζόμενα. Αν ο ερευνητής μοιραστεί έναν τέτοιο κατάλογο, αναρτώντας τον στον διαδίκτυο, θα βοηθήσει πολλούς άλλους και άλλες συναδέλφους του να αποφύγουν μια εργώδη διαδικασία – στην περίπτωση, βέβαια, των ειδικών καταλόγων, καλό είναι η ανάρτηση να συνοδεύεται από έναν σύντομο σχολιασμό για την περίσταση στην οποία χρησιμοποιήθηκαν, ώστε ο ενδιαφερόμενος ερευνητής να έχει μια εικόνα των περιορισμών τους. Υπογραμμίζεται έτσι και πάλι η ανάγκη της συλλογικότητας στην έρευνα της λογοτεχνίας, που θέτει υπό αίρεση, όπως επεσήμαναν και οι Rockwell και Sinclair το επί μακρόν κυρίαρχο μοντέλο του μοναχικού ερευνητή (2016). Ως προς τους καταλόγους λέξεων, στο άμεσο μέλλον θα αναρτηθεί στο αποθετήριο Clarin-Απολλωνίς ΑΠΘ ένας κατάλογος μη λημματοποιημένων τύπων της δημοτικής και της καθαρεύουσας. Ως προς την τόσο σημαντική λημματοποίηση, αναμένουμε πολύ σύντομα τη δυνατότητα εξαγωγής των αποτελεσμάτων του εργαλείου λημματοποίησης της καθομιλουμένης, που είναι ελεύθερα διαθέσιμο σε επίπεδο υπηρεσίας στο αποθετήριο του CLARIN:EL, σε μορφή που θα επιτρέπει τη χρησιμοποίησή τους σε εργαλεία ανάλυσης κειμένου. Παρότι είναι προφανές ότι θα χρειαστούν έλεγχοι διαλειτουργικότητας, είναι ακόμη περισσότερο προφανές ότι αυτή η εξέλιξη θέτει τις βάσεις για την ανάπτυξη της ΗΑΚ στα ελληνικά.

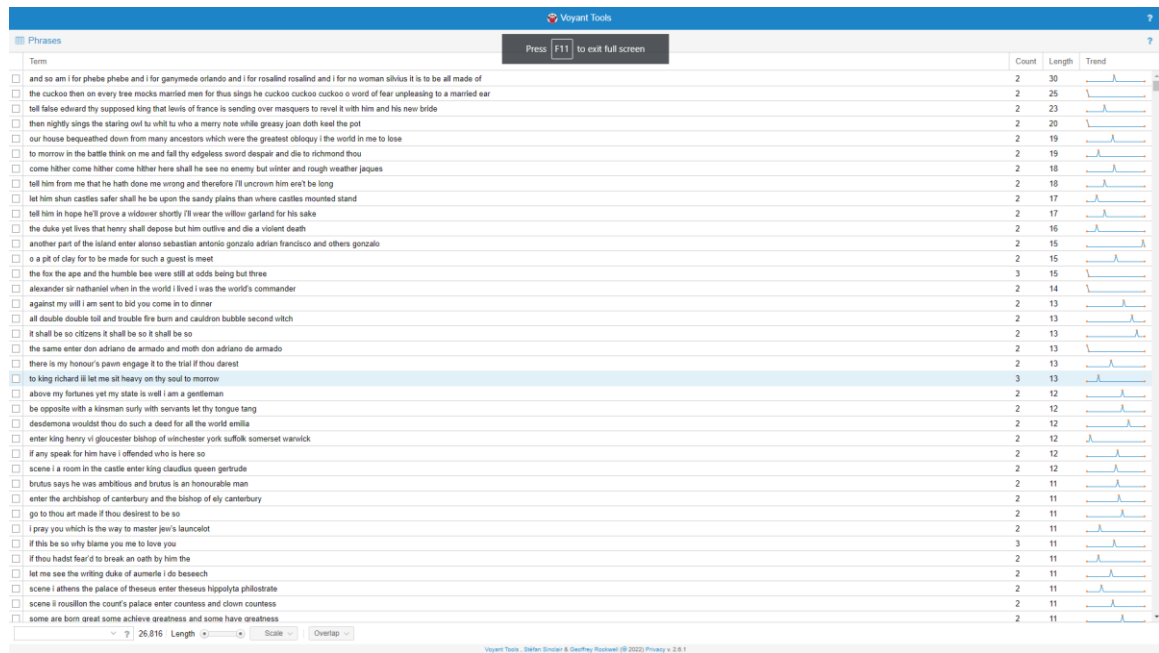
Πέραν των καταλόγων συχνότητας, που μπορούν επίσης να είναι αντίστροφοι, να εμφανίζουν δηλαδή τις λέξεις με αλφαβητική σειρά με βάση την κατάληξή τους, κάτι που μπορεί να παρουσιάζει ενδιαφέρον στη μελέτη της ποίησης, σημαντική λειτουργία στην ΗΑΚ αποτελεί η εμφάνιση των λέξεων μέσα στο συγκεκριμένο τους (ο καθαυτό συμφραστικός πίνακας) –με την έκταση του συγκεκριμένου να ορίζεται από τον ερευνητή και με διάθεσιμη τη δυνατότητα αναγωγής στο σύνολο του κειμένου.⁵ Άλλη μια ιδιαίτερα σημαντική λειτουργία των λογισμικών ανάλυσης κειμένου, ιδιαίτερης χρησιμότητας για τη λογοτεχνική έρευνα, είναι η δυνατότητα μελέτης των συνάψεων σε λεξικό και συντακτικό επίπεδο (collocations, colligations), η οποία χειρωνακτικά παρουσιάζει μεγάλες δυσκολίες. Επίσης, η έρευνα για τη σημαντικότητα (keyness) συγκεκριμένων όρων, με σύγκριση του ΗΣΚ με κάποιο άλλο ΗΣΚ. Τέλος, σημαντική είναι η δυνατότητα σύγκρισης καταλόγων συχνότητας διαφορετικών ΗΣΚ, η οποία είναι διαθέσιμη ωστόσο σε περιορισμένο αριθμό εφαρμογών, όπως π.χ τα Wordsmith tools. Στη λογοτεχνία, η σύγκριση αυτή μπορεί να οδηγήσει σε πολύ ενδιαφέροντα συμπεράσματα, σε μικρο-και μακρο-επίπεδο, από τη διερεύνηση της εξέλιξης του ύφους σε έναν συγγραφέα ως τη μελέτη της δηλωτικής και συνδηλωτικής σημασίας συγκεκριμένων όρων σε ευρέα ΗΣΚ. Είναι δε προφανώς ιδιαίτερως χρήσιμη σε αναλύσεις με γνώμονα τα ΗΣΚ.

Παρότι σε όλα τα λογισμικά παρέχονται δυνατότητες οπτικής απεικόνισης των δεδομένων που διευκολύνουν την αντίληψη τάσεων στα κείμενα (οπτικοποίηση), οι δυνατότητες αυτές διαφοροποιούνται σημαντικά. Στα Voyant Tools, η οπτικοποίηση είναι ιδιαίτερως εξελιγμένη και παρέχει στοιχεία σε πολλά διαφορετικά επίπεδα και με διαφορετικές μορφές απεικόνισης, των χαρτών συμπεριλαμβανομένων, αν και σε πειραματικό επίπεδο. Η ίδια η οπτικοποίηση αποτελεί αντικείμενο προβληματισμού, ωστόσο, σε σχέση με την προέλευσή της από τις θετικές επιστήμες και τους περιορισμούς που αυτό μπορεί να συνεπάγεται (Drucker Rockwell και Sinclair).

Τα εργαλεία Voyant Tools δίνουν τη δυνατότητα ταυτόχρονης προβολής των διαφόρων στοιχείων ανάλυσης σε διακριτά παράθυρα, όπως τα συννεφώλεξα (Cirrus), ο κατάλογος συχνότητας (Terms), το κείμενο καθαυτό (Reader), ο συμφραστικός πίνακας λεξικών τύπων (Contexts), οι συνεμφανίσεις (Termberry και Correlations) αλλά και η (ποικιλοτρόπως οπτικοποιημένη) κατανομή τους στο σώμα και τα υποσώματα κειμένων (Trends), μαζί με τις πληροφορίες για το ΗΣΚ (Summary, Documents, Phrases).

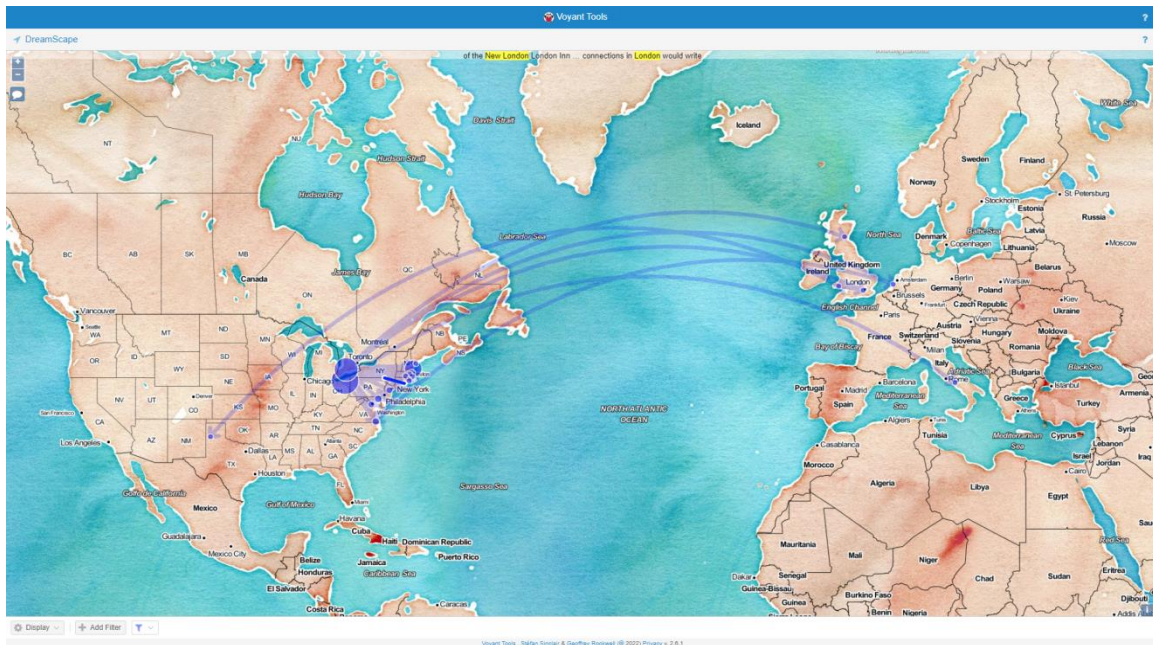
⁵ Να θυμίσουμε εδώ ότι οι συμφραστικοί πίνακες υπάρχουν από τον 13ο αιώνα ήδη σε χειρόγραφοι και στη συνέχεια σε έντυπη μορφή (σχετικά με τον πρώτο συμφραστικό πίνακα της Βουλγάτας, βλ. Albaric 2004). Προφανώς, η ρύθμιση της έκτασης του συγκεκριμένου δεν αποτελεί χαρακτηριστικό τους.

Documents					
	Title	Words	Types	Ratio	Words/Sentence
1	1590 Love's Labour's Lost	23.1	3.767	16%	14.3
2	1591 Comedy of Errors	16.3	2.534	16%	16.2
3	1591 King Henry VI (Part I)	23.2	3.886	17%	17.7
4	1591 Romeo and Juliet	26.1	3.713	14%	14.9
5	1591 Two Gentlemen of Verona	18.4	2.737	15%	13.9
6	1592 Titus Andronicus	21.8	3.426	16%	19.1
7	1593 King Henry VI (Part II)	27.1	4.106	15%	17.4
8	1593 King Richard III	31.8	4.101	13%	16.9
9	1594 Merchant of Venice	22.4	3.281	15%	18.0
10	1594 Midsummer Night's Dream	17.3	3.065	17%	16.5
11	1594 Taming of the Shrew	32.3	3.273	15%	15.8
12	1595 King Henry VI (Part III)	26.2	3.607	14%	17.4
13	1596 King John	21.9	3.560	16%	20.4
14	1597 King Henry IV (Part I)	26.3	3.866	15%	17.5
15	1597 King Henry IV (Part II)	28.2	4.125	15%	16.6
16	1597 Kings Richard II	24.1	3.676	15%	21.5
17	1598 All's Well That Ends Well	24.5	3.527	14%	17.3
18	1598 As You Like It	23.0	3.272	14%	16.9
19	1599 Hamlet	32.3	4.728	15%	16.2
20	1599 Julius Caesar	21.0	2.889	14%	14.7
21	1599 King Henry V	27.8	4.544	16%	20.3
22	1599 Merry Wives of Windsor	24.0	3.255	14%	13.6
23	1599 Twelfth Night or What You Will	21.5	3.116	14%	14.7
24	1600 Much Ado about Nothing	22.7	2.979	13%	15.4
25	1602 Troilus and Cressida	27.7	4.263	15%	14.3
26	1604 Measure for Measure	23.2	3.356	14%	15.0
27	1604 Othello	28.0	3.801	14%	13.6
28	1605 King Lear	27.9	4.176	15%	14.2
29	1605 Macbeth	18.4	3.358	18%	15.0
30	1607 Antony and Cleopatra	27.2	3.980	15%	14.3
31	1607 Timon of Athens	19.8	3.334	17%	14.5
32	1608 Coriolanus	29.4	4.067	14%	15.9
33	1608 Pericles Prince of Tyre	19.8	3.306	17%	16.2
34	1610 Cymbeline	29.2	4.289	15%	16.6
35	1610 Winter's Tale	26.1	3.911	15%	18.5
36	1611 Tempest	17.5	3.268	18%	15.1
37	1613 King Henry VIII	28.2	3.879	14%	19.4



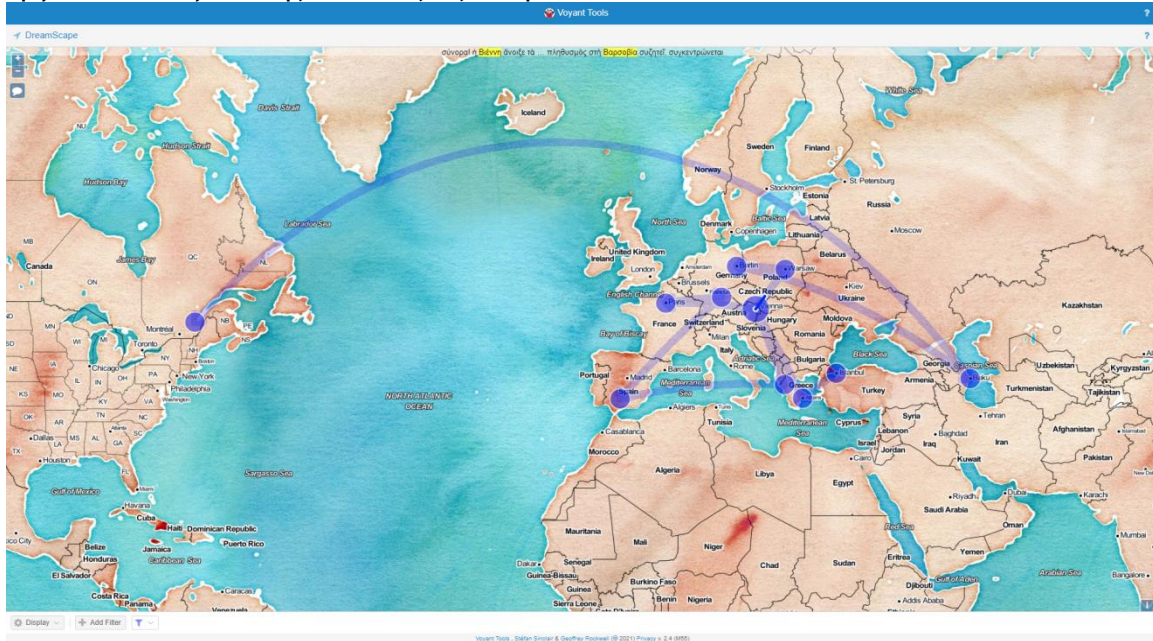
Εικόνα 5. Phrases (προτάσεις): οι προτάσεις που επαναλαμβάνονται στο ΗΣΚ έργων του Σαίξπηρ, εμφανιζόμενες με βάση τον αριθμό των επαναλήψεων και τον αριθμό των λέξεών τους.

Είναι προφανές, ότι τα γενικά στοιχεία και μόνο, πριν περάσει κανείς στην ανάλυση, είναι πολύτιμα για την έρευνα. Όσο για την οπτικοποίηση, είναι κοινή η παραδοχή σε πολλούς επιστήμονες ότι εισάγει το δικό της επιχείρημα στην ανάλυση, καθώς αναπαριστά μεμονωμένα στοιχεία των κειμένων (Drucker 2020· Sinclair και Rockwell 2016). Ωστόσο, οι δυνατότητες οπτικοποίησης διαρκώς διευρύνονται με βάση και τις ανάγκες των ανθρωπιστικών επιστημών. Ένα παράδειγμα αποτελεί η χρήση γεωχωρικών δεδομένων και η απεικόνισή τους, στατική ή δυναμική, σε χάρτη. Στα Voyant tools, η πιλοτική λειτουργία dreamscape διερευνά δυναμικά γεωχωρικές όψεις των κειμένων, όπως στην εικόνα που ακολουθεί και απεικονίζει τα σχετικά δεδομένα στα έργα της Jane Austen, που δίνονται ως δείγμα στα εργαλεία (η δυναμική απεικόνιση προφανώς δεν αποτυπώνεται στην εικόνα):



Εικόνα 6. Dreamscape, ΗΣΚ έργων της Jane Austen, διαθέσιμο ως δείγμα στα Voyant Tools.

Ακολουθεί μια εικόνα από την οπτικοποίηση γεωχωρικών δεδομένων στο έργο *Το σπίτι μου*, της Μέλπως Αξιώτη:



Εικόνα 7. Μέλπω Αξιώτη, *Το σπίτι μου*.

Οι χάρτες αποτελούν πανίσχυρα εργαλεία σε ρητορικό επίπεδο και η εν λόγω εφαρμογή είναι πειραματική. Ωστόσο, ανοίγουν νέες προοπτικές στην απεικόνιση ποικίλων δικτύων, που ενδιαφέρει πολλαπλά τη λογοτεχνία.

3. Ενδεικτικά πεδία εφαρμογής της ΗΑΚ

Ως προς τις υφιστάμενες μελέτες με χρήση ΗΑΚ, τα πεδία εφαρμογής είναι πολλά και διαφορετικά και η σχετική βιβλιογραφία σε άλλες γλώσσες πλούσια. Συχνά ακούγεται ότι η ΗΑΚ αποτελεί μια ποσοτική μέθοδο περιορισμένης χρησιμότητας, με κύριο πεδίο αναφοράς την υφολογία. Στη συνέχεια θα δώσω ορισμένα ενδεικτικά παραδείγματα μελετών, θεματικών, υφολογικών, ιστορικών, κοινωνιολογικών και άλλων που μπορούν να υποστηριχθούν από την ΗΑΚ, με αναφορές σε ζητήματα της νεοελληνικής λογοτεχνίας.

- Θεματικές μελέτες σε ποικίλα πεδία, για θεματικές έργων, για συγκεκριμένα πολιτισμικά στοιχεία, για έννοιες και ομάδες εννοιών στο έργο ενός ή περισσοτέρων συγγραφέων ή σε συγκεκριμένες χρονικές περιόδους, με όρους ερμηνευτικούς και διεπιστημονική στόχευση (βλ. λόγου χάρη την έρευνα για την εξέλιξη της έννοιας του χρόνου, Brunet 1994). Μια θεματική που μπορεί να μελετηθεί σε νεοελληνικά αφηγηματικά έργα είναι η εξέλιξη της έννοιας της εργασίας στον 20ό αιώνα, η οποία, μετά από κάποιες σχετικές δοκιμές που έχω πραγματοποιήσει, φαίνεται ότι δίνει πολύ σημαντικά στοιχεία κοινωνιολογικού και ιδεολογικού ενδιαφέροντος. Ειδικά η σύγκριση ανάμεσα σε έργα του μεσοπολέμου, της μεταπολεμικής και της μεταπολιτευτικής περιόδου αναδεικνύει την κοινωνικοπολιτική συνθήκη και ανοίγει πολλαπλά πεδία διεπιστημονικής συνέργειας σε επίπεδο ερμηνείας.
- Έλεγχος υποθέσεων σχετικά με συγκεκριμένους συγγραφείς, έργα, είδη και περιόδους. Στα νεοελληνικά συμφραζόμενα, ανάλογα ερωτήματα μπορούν να διατυπωθούν για πολλούς συγγραφείς και ποιητές, αλλά και διάφορα είδη, όπως λόγου χάρη τα παλαιότερα απόκρυφα ή οι σύγχρονες δυστοπίες, σε επίπεδο αφηγηματολογικό, υφολογικό και πραγματολογικό. Ένα ερώτημα σε σχέση με την αστυνομική λογοτεχνία και τη θέση του Todorov «περί λογοτεχνίας με συνταγή», θα μπορούσε να αφορά τη συγκριτική υφολογική και αφηγηματολογική ανάλυση α. του έργου διαφορετικών ελλήνων συγγραφέων αστυνομικής λογοτεχνίας σε μια ή σε περισσότερες περιόδους· β. μυθιστορημάτων ή διηγημάτων αστυνομικής λογοτεχνίας και άλλων που δεν ανήκουν στο είδος, με στόχο να διερευνηθεί αν και κατά πόσο το εύρος των διαφοροποιήσεων επιβεβαιώνει τον Todorov ή τη θέση που υποστηρίζει την καθ' έκαστον εξέταση και των αστυνομικών έργων.
- Έλεγχος ή υποστήριξη μιας κριτικής θέσης, για παράδειγμα σχετικά με τη θέση των μερών του λόγου και τη σημασία τους σε ένα έργο, το σύνολο των έργων ενός συγγραφέα ή μιας περιόδου. Ανάλογες μελέτες θα μπορούσαν να γίνουν στα ελληνικά με αναφορά στη χρήση των επιθέτων στον Παπαδιαμάντη λόγου χάρη, ή στο έργο συγκεκριμένων ποιητών και μάλιστα στην εξέλιξή του, ή σε ειδικά στοιχεία της σύνταξης ή της στίξης κ.ο.κ.
- Μελέτη των όρων διαφοροποίησης του ιδιολέκτου των προσώπων σε ένα ή περισσότερα είδη, π.χ. στα μυθιστορήματα και στα θεατρικά έργα, αλλά

και στην αλληλογραφία συγγραφέων, ενδεχομένως και σε συγκριτική βάση.

- Μελέτη της εξέλιξης του ύφους συγγραφέων και της εξέλιξης των ειδών, αλλά και της επίδρασης των συμβάσεών τους στον λόγο των κειμένων, τη διάρθρωση ή τη χαρακτηρισποίηση. Ως προς την εξέλιξη του ύφους των συγγραφέων, κάναμε ένα πείραμα σύγκρισης καταλόγων λέξεων τριών σωμάτων κειμένων της Αξιώτη, για να επιβεβαιώσουμε την υπόθεση ότι η τελευταία φάση του έργου της επανασυνδέεται με την πρώτη. Το πρώτο σώμα αφορούσε τα λογοτεχνικά της έργα ως το 1940 (*Δύσκολες νύχτες, Σύμπτωση, Θέλετε να χορέψουμε, Μαρία;*). Το δεύτερο τα έργα της από το 1941 ως το 1958 (*Εικοστός αιώνας, Σύντροφοι, καλημέρα*). Και το τρίτο, τα έργα από το 1959 ως το 1972 (*Κοντραμπάντο, Θαλασσινά, Το σπίτι μου, Η Κάδμω*). Η υπόθεση επιβεβαιώθηκε σε θεματικό επίπεδο, όπου καταγράφονται ποικίλες επαναλήψεις, αλλά και σε υφολογικό, με αναφορά στην αναλογία διαλεκτικής και κοινής γλώσσας, μήκους και δομής της πρότασης, στίξης κ.ά. Ο κατάλογος λέξεων της Αξιώτη υπήρξε βοηθητικός στη δημιουργία ενός ύφους που να μην την προδίδει στη μετάφραση του γαλλικού μυθιστορηματός της *Ρεπυμπλικ-Βαστίλλη*. Όταν η μετάφραση συμπεριλήφθηκε πιλοτικά στο σώμα κειμένων 1940-1958, η απόκλιση από τα αντίστοιχα έργα της περιόδου ήταν σημαντική και επιβεβαίωνε τη μεταβατική θέση του κειμένου, που συνδέεται τόσο με τα έργα της αρχής όσο και της ύστερης περιόδου παραγωγής της. Ως προς την εξέλιξη του ύφους του Παπαδιαμάντη, πάλι, ενδιαφέροντα στοιχεία προκύπτουν από τη σύγκριση των μυθιστορημάτων και των διηγημάτων του.
- Συσχέτιση της γραφής με την καταγωγή και την παιδεία του συγγραφέα, δηλαδή με το κατά Bourdieu *habitus*, με ανάλυση του έργου ενός ή περισσότερων συγγραφέων. Συναφείς είναι οι μελέτες που αφορούν όλα τα στοιχεία που σχετίζονται με το *habitus*, πέραν της ταξικής καταγωγής, η φυλή, το χρώμα, το φύλο, η ηλικία κ.ά. Η Μαρία Ιορδανίδου, για παράδειγμα, θα μπορούσε να αποτελέσει μια ενδιαφέρουσα μελέτη περίπτωσης με αναφορά σε περισσότερα του ενός στοιχεία και ειδικά στην ηλικία.
- Μελέτη της παρωδίας, ένα από τα πιο ενδιαφέροντα πεδία έρευνας της ΗΑΚ. Η περίπτωση των καβαφικών ποιημάτων είναι μια περίπτωση που μπορεί να δώσει πολύ σημαντικά αποτελέσματα, όσον αφορά τους τρόπους της μίμησης σε όλα τα επίπεδα.
- Υφολογικές αναλύσεις σε ποικίλα επίπεδα, συγγραφέων, ποιητών, ειδών, περιόδων, συγκριτικές και μη και υφομετρικές αναλύσεις και ειδικά αναλύσεις απόδοσης πατρότητας, όπως αυτές που εφαρμόστηκαν σε ανώνυμες μεταφράσεις για τη συσχέτισή τους με τον Παπαδιαμάντη (Πολίτου-Μαρμαρινού, Μικρός και Δημητρούλια 2013). Στην περίπτωση των υφομετρικών αναλύσεων ωστόσο, η κλασική μεθοδολογία που ακολουθήθηκε για τον Παπαδιαμάντη είναι εντελώς διαφορετική σε σχέση με την ερμηνευτική ανάλυση κειμένου και στηρίζεται σε μοντέλα και μεταβλητές, δηλαδή σε μια εντελώς διαφορετική προσέγγιση της ανάλυσης.

Οι μελέτες που έχουν ήδη γίνει σε διάφορα επίπεδα για άλλες λογοτεχνίες (βλ. ενδεικτικά Hoover 2013) δεν ενδιαφέρουν μόνο ως πηγή έμπνευσης αλλά και ως προς τα συγκεκριμένα μεθοδολογικά και ερμηνευτικά τους πλαίσια, τα οποία μπορούν να προκαλέσουν γόνιμες συζητήσεις και αντιπαραθέσεις. Θα δώσω ένα παράδειγμα. Προσφάτως, έγινε μια έρευνα, χωρίς χρήση ΗΑΚ, για το λεξιλογικό πεδίο της τροφής και του ποτού στον Παπαδιαμάντη, από μια ομάδα ερευνητών υπό τον συντονισμό της Γ. Φαρίνου Μαλαματάρη.⁶ Η αποδελτίωση αυτή, που παρουσιάζεται και σε πίνακα, θα μπορούσε να έχει υποστηριχθεί από ΗΑΚ του ΗΣΚ των έργων του που είναι διαθέσιμο στο αποθετήριο Clarin-Απολλωνίς ΑΠΘ, γλιτώνοντας τους ερευνητές από πολύ χρόνο και κόπο. Ο πρώτος τρόπος, ο οποίος είναι διερευνητικός, έχει να κάνει με τη λειτουργία του Καταλόγου λέξεων που παρέχουν όλα τα λογισμικά ανάλυσης κειμένου. Αντί για ανάγνωση των κειμένων, ο ερευνητής ή η ερευνήτρια θα μπορούσε να διατρέξει τον κατάλογο των λέξεων του κειμένου, κάτι πολύ πιο εύκολο από ό,τι μπορεί να φαίνεται αν το δοκιμάσει κανείς, και οπωσδήποτε λιγότερο χρονοβόρο από την προσεκτική ανάγνωση των διηγημάτων για τον εντοπισμό των σχετικών δεδομένων, ειδικά αν αφαιρεθούν οι λειτουργικοί τύποι, αλλά και όποιες άλλες λέξεις θεωρούνται ότι δεν επηρεάζουν την ποιότητα της ανάλυσης, π.χ. ορισμένα ρήματα. Η λημματοποίηση υπάρχει ως υπηρεσία σε ορισμένα προγράμματα ανάλυσης κειμένου (π.χ. AntConc, Sketch Engine) και ως ανεξάρτητη υπηρεσία στο αποθετήριο του CLARIN:EL, και θα μπορούσε να είναι πολύ βοηθητική. Αλλά σημαντικό περιορισμό στη χρήση της θέτει το πολυτονικό κείμενο, το οποίο τα περισσότερα λογισμικά δυσκολεύονται να επεξεργαστούν. Στην περίπτωση αυτή, είναι προτιμότερη η εκ των προτέρων μετατροπή του πολυτονικού κειμένου σε μονοτονικό, με κάποια μακροεντολή, ώστε να είναι τελικά εφικτή η εμφάνιση καταλόγου λημμάτων.

Ένας άλλος τρόπος έρευνας στο ΗΣΚ των έργων του Παπαδιαμάντη είναι η αναζήτηση ενός συνόλου λέξεων που έχουν οριστεί εκ των προτέρων, ενός λεξιλογικού πεδίου δηλαδή εν προκειμένω του φαγητού. Στην παρακάτω εικόνα, εμφανίζεται η συχνότητα της λέξης «ψωμί» στα διηγήματα του Π., η κατανομή της με οπτικοποίηση και οι εμφανίσεις της στα συμφραζόμενά τους. Σ' αυτή την περίπτωση, ωστόσο, θα πρέπει να βρεθούν κριτήρια για τη συγκρότηση του λεξιλογικού πεδίου, ενώ πάντα η έρευνα θα χρειαζόταν να συμπληρωθεί με την εξέταση του καταλόγου λέξεων, καθώς το σύνολο δεν μπορεί σε καμία περίπτωση να είναι πλήρες. Η δυνατότητα εξέτασης των όρων στο συγκεκριμένο τους, με ταυτόχρονη αναγωγή στο ίδιο το κείμενο, μπορεί να προσφέρει ενδιαφέροντα στοιχεία σχετικά με το ζήτημα, πέραν των μεμονωμένων στοιχείων του υλικού πολιτισμού, όπως στοιχεία κοινωνικής και γεωγραφικής διαστρωμάτωσης, που μπορούν να εξεταστούν συγκριτικά με την εμφάνισή τους σε άλλα έργα της εποχής, αναλυμένα με τον ίδιο τρόπο.

Ένα δεύτερο παράδειγμα, αφορά τη μελέτη των χρωμάτων στο έργο του Παπαδιαμάντη. Σ' αυτή την περίπτωση, αφού ακολουθηθεί η αρχική διαδικασία, συγκροτείται ένα λεξιλογικό πεδίο χρωμάτων με βάση ένα λεξικό ή με συνδυαστική χρήση διαφόρων πηγών. Προφανώς, το ενδιαφέρον στον Παπαδιαμάντη έγκειται στα άλλα, πλην των συνήθων, χρώματα που χρησιμοποιεί, συχνά εκφρασμένα με σύνθετα επίθετα και η συνδυαστική αναζήτηση είναι επιβεβλημένη. Αλλά η σχέση

⁶ <https://www.papadiamantis.net/aleksandros-papadiamantis/meletes/trexousa-erevna>.

μεταξύ των δύο μπορεί επίσης να παρουσιάζει ενδιαφέρον, με βάση την κατανομή και τις συνδηλώσεις τους, όπως προκύπτουν από τις λεξικές και γραμματικές συνάψεις τους. Ενώ η μελέτη της σημαντικότητας των χρωμάτων σε συγκεκριμένα έργα και/ή σε περιόδους του έργου του μπορεί επίσης να δώσει ή να επιβεβαιώσει ερμηνευτικές κατευθύνσεις. Επομένως, η ποσοτική ανάλυση είναι πάντα ποιοτική, ερμηνευτική ανάλυση, όπως φαίνεται ακόμη και από τα απλά αυτά παραδείγματα.

4. Οπτικές-προοπτικές

Στη σύντομη αυτή αναφορά, δεν παρουσιάσαμε παρά κάποια μόνο στοιχεία της ηλεκτρονικής ανάλυσης με συγκεκριμένα εργαλεία και από τη σκοπιά της ερμηνευτικής, που θεωρούμε ότι ορίζει τη μελέτη της λογοτεχνίας. Όπως αναφέρθηκε εξ αρχής, στόχος της είναι να κινήσει το ενδιαφέρον για την ηλεκτρονική ανάλυση κειμένου, η οποία σήμερα εμπλουτίζεται με ποικίλα εργαλεία, πολλά από τα οποία ενσωματώνουν τεχνητή νοημοσύνη. Δεν αναφέρθηκα καθόλου στις συναρπαστικές πτυχές των νέων αυτών δυνατοτήτων. Επιφυλάσσομαι να το κάνω σε έναν δεύτερο χρόνο, με ένα άρθρο που θα αποτελεί εφαρμογή μεθοδολογίας σε συγκεκριμένη μελέτη περίπτωσης και θα ανοίγει τη συζήτηση και για τη χρήση της τεχνητής νοημοσύνης στην ερμηνευτική ανάλυση. Σε κάθε περίπτωση, η συζήτηση είναι ανοιχτή και οι νέοι συνομιλητές καλοδεχούμενοι, όπως καλοδεχούμενοι είναι και όσοι και όσες θα ήθελαν να εκπαιδευτούν στην ηλεκτρονική ανάλυση κειμένου. Η ομάδα Απολλωνίς του ΑΠΘ ξαναρχίζει σύντομα τα σεμινάρια της.

Βιβλιογραφία

- Albaric, Michel. 2004. Hugues de Saint-Cher et les concordances bibliques latines (xiii^e-xviii^e siècles)». In Louis-Jacques Bataillon, Gilbert Dahan και Pierre-Marie Gy (dir.), *Hugues de Saint-Cher († 1263), bibliste et théologien: Études réunies*, 467-479. Turnhout: Brepols.
- Brunet, Etienne. 1994. Quand le temps change avec le temps. *Textes*, 60-86.
- Brunet, Etienne. 2009. *Comptes d'auteurs. Études statistiques de Rabelais à Gracq*. Textes édités par Damon Mayaffre, préface d'Henri Béhar. Paris: Champion.
- Brunet, Etienne. 2011. *Ce qui compte. Méthodes statistiques*. Textes édités par Céline Poudat, préface de Ludovic Lebart. Paris: Champion.
- Busa, Roberto. 1980. The Annals of Humanities Computing: The Index Thomisticus. *Computers and the Humanities* 14, 2, 83-90, <https://www.alice.id.tue.nl/references/bus-1980.pdf>.
- Chapple, Freda and Chiel Kattenbelt (eds). 2006. *Intermediality in theatre and performance*. New York: Rodopi.
- Culpeper, Jonathan. 2002. Computers, Language and Characterisation: An Analysis of Six Characters in *Romeo and Juliet*. In, Ulla Melander-Marttala, Carin Ostman και Merja Kytö (eds.), *Conversation in Life and in Literature: Papers from the ASLA Symposium, Association Suedoise de Linguistique Appliquée (ASLA)*, 15, 11-30. Universitetsstryckeriet: Uppsala, https://lexically.net/wordsmith/corpus_linguistics_links/Keywords-Culpeper.pdf.
- Culpeper, Jonathan. 2009. Keyness: Words, Parts-of-Speech and Semantic Categories in the Character talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics* 14, 1, 29-59. https://www.researchgate.net/publication/233493422_Keyness_Words_parts-of-speech_and_semantic_categories_in_the_character-talk_of_Shakespeare%27s_Romeo_and_Juliet.
- Drucker, Johanna. 2020. *Visualization and Interpretation. Humanistic Approaches to Display*. Cambridge, Mass.: MIT Press.
- Eve, Martin Paul. 2022. *The Digital Humanities and Literary Studies*. Oxford: Oxford Academic, <https://doi.org/10.1093/oso/9780198850489.001.0001>, accessed 11 Nov. 2022.
- Granjon, Fabien. 2016. Présentation du dossier. Les sciences humaines et sociales au prisme du *digital turn*. *Variations*, 19. DOI: <https://doi.org/10.4000/vari-ations.726>.
- Guiraud, P. και J. Whatmouth. 1954. *Bibliographie de statistique linguistique*. Utrecht/Anvers: Mouton.
- Béhard, Henri. 1995. Hubert de Phalèse's Method. *Literary and Linguistic Computing*, 10, 2, 1995, 129-134, <https://doi.org/10.1093/lc/10.2.129-b>
- Herdan, Gustav. 1964. *Quantitative Linguistics*. London : Butterwoths.
- Hockey, Susan. 2004. The History of Humanities Computing. In Susan Schreibman, Ray Siemens and John Unsworth (eds), *A Companion to Digital Humanities*. Oxford: Blackwell Publishing, 3-19.

- Hoover, David L. 2013. Textual Analysis. In Kenneth M. Price και Ray Siemens (eds), *MLA Commons. Literary Studies in the Digital Age: An Evolving Anthology*, <https://dlsanthology.mla.hcommons.org/textual-analysis>.
- Kirschenbaum, Matthew. 2010. What is Digital Humanities and What's It Doing in English Departments. *ADE Bulletin* 150, 55-61, <https://www.uvic.ca/humanities/english/assets/docs/kirschenbaum.pdf>
- Lehmann, Hans Thies. 2006. *Postdramatic theatre*. Transl., intr. by Karen Jürs-Munby. London: Routledge [1999. *Postdramatisches Theater*. Frankfurt am Main: Verlag der Autoren].
- Manovich, Lev. 2001. *The language of new media*. Cambridge, Mass.: MIT Press.
- Marranca, Bonnie. 2009. Mediaturgy: A conversation with Marianne Weems. *International Journal of Arts and Technology*, 2, 3, 173-86.
- Marranca, Bonnie. 2010. Performance as design: The mediaturgy of John Jesurun's Firefall. *PAJ*, 96, 16-24.
- Rockwell, Geoffrey και Stéfán Sinclair. 2016. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Cambridge, Mass: MIT Press.
- Siemens, Ray and Susan Schreibman (eds). 2008. *A Companion to Digital Literary Studies*. Oxford: Blackwell, 2008. <https://companions.digitalhumanities.org/DLS>.
- Sinclair, Stéfán και Rockwell Geoffrey. 2016. Text Analysis and Visualization: Making Meaning Count». In Susan Schreibman, Ray Siemens και John Unsworth (eds), *A New Companion to Digital Humanities*, 274-290. Oxford: Blackwell, 2016.
- Susan Schreibman, Ray Siemens and John Unsworth (eds). 2004. *A Companion to Digital Humanities*. Oxford: Blackwell Publishing.
- Brunet, Etienne. 2016. *Tous comptes faits. Questions linguistiques*. Textes édités par Bénédicte Pincemin, préface de François Rastier. Paris : Champion.
- Winter, Thomas Nelson. 1999. Roberto Busa, S.J., and the Invention of the Machine-Generated Concordance. Faculty Publications, Classics and Religious Studies Department 70, <http://digitalcommons.unl.edu/classicsfacpub/70>.
- Ανεμόσκαλα: σώματα κειμένων και συμφραστικοί πίνακες για μείζονες Νεοέλληνες ποιητές. Κέντρο Ελληνικής Γλώσσας, 2011-2017. <http://www.greek-language.gr/digitalResources/literature/tools/concordance/index.html>
- Δημητρούλια, Ξανθίππη και Αικατερίνη Τικτοπούλου. 2015. *Ψηφιακές λογοτεχνικές σπουδές*. Αθήνα: ΣΕΑΒ, <https://hdl.handle.net/11419/5827>.
- Δημητρούλια, Τιτίκα. 2005. *Μεταφραστική. Ο λόγος περί της μετάφρασης στην εποχή της ψηφιοποίησης*. Ανέκδοτη διδακτορική διατριβή. Αθήνα: Πάντειο Πανεπιστήμιο.
- Δημητρούλια, Τιτίκα. 2006. Κυβερνολογοτεχνία: μια πρόκληση του μέλλοντος. *Σύγκριση*, 17, 94-113. <https://doi.org/10.12681/comparison.10218>.
- Δημητρούλια, Τιτίκα. Υπό δημ. Ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου: νέες ερμηνευτικές προοπτικές. Στο Δέσποινα Βαλατσού, Άντα Διάλλα, Ιουλία Πεντάζου, Πέτρος Πετρίδης, Ηλίας Στουραϊτης, *Ψηφιακές Ανθρωπιστικές Επιστήμες στην Ελλάδα: προβληματισμοί και προκλήσεις*. Πρακτικά ομότιτλου

συνεδρίου, 16.2.2018, Αθήνα, ΑΣΚΤ. Αθήνα: Κέντρο έρευνας για τις ανθρωπιστικές επιστήμες.
Λαλιώτη, Βασιλική (επιμ.). 2022. *Ψηφιακές τεχνολογίες και τέχνες*. Θεσσαλονίκη: Ροπή.
Πατηνιώτης, Μανώλης (επιμ.). 2020. *Εισαγωγή στις ψηφιακές σπουδές*. Θεσσαλονίκη: Ροπή.

Résumé

Titika Dimitroulia

Les corpus numériques et leur analyse. Perspectives interprétatives dans l'étude de la littérature

Cet article se propose de donner un aperçu de l'utilisation des corpus numériques dans les études littéraires et de l'apport de leur analyse à l'aide des outils numériques à la recherche. Il vise à présenter brièvement, dans une perspective herméneutique, la textométrie, comme cette méthode s'appelle dans le domaine francophone où elle a une longue histoire, de discuter les problèmes relatifs aux outils de traitement de la langue grecque, et de donner quelques exemples de son application à l'étude de la littérature néohellénique. Son objectif serait d'encourager l'exploration de la méthode par les chercheurs grecs, dans le cadre la transformation numérique de la recherche des sciences humaines et sociales.