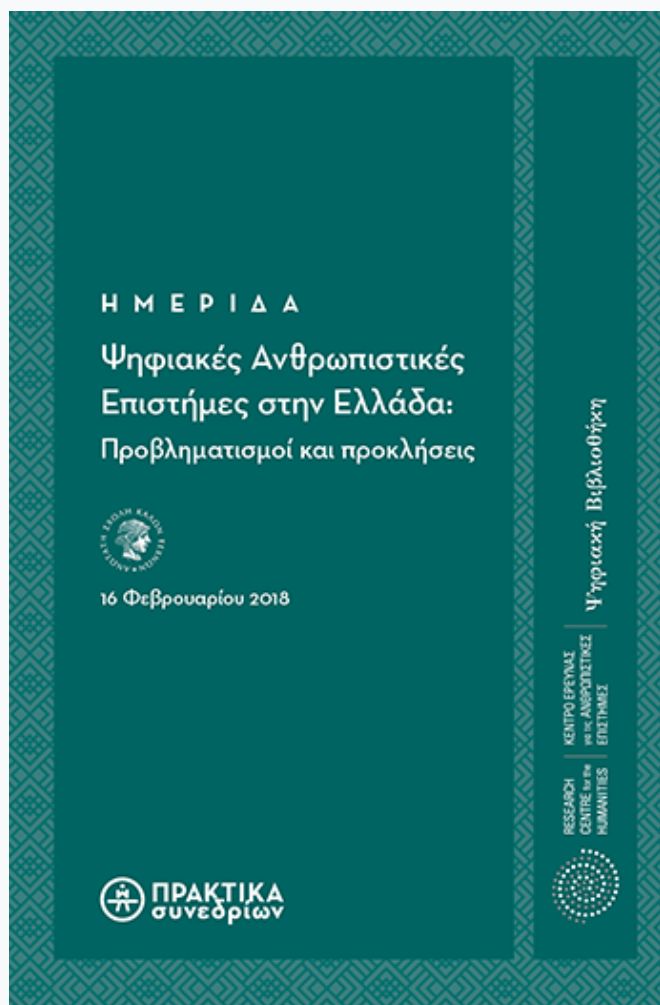


Πρακτικά Συνεδρίων

(2023)

Ψηφιακές Ανθρωπιστικές Επιστήμες στην Ελλάδα: προβληματισμοί και προκλήσεις (16 Φεβρουαρίου 2018)



Ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου

Τιτίκα Δημητρούλια

doi: [10.12681/praktika.5210](https://doi.org/10.12681/praktika.5210)

Η Μ Ε Ρ Ι Δ Α

Ψηφιακές Ανθρωπιστικές Επιστήμες στην Ελλάδα: Προβληματισμοί και προκλήσεις



16 Φεβρουαρίου 2018

 ΠΡΑΚΤΙΚΑ
συνεδρίων

Ψηφιακή Βιβλιοθήκη

ΚΕΝΤΡΟ ΕΡΕΥΝΑΣ
για τις ΑΝΘΡΩΠΙΣΤΙΚΕΣ
ΕΠΙΣΤΗΜΕΣ

RESEARCH
CENTRE for the
HUMANITIES



Ψηφιακές ανθρωπιστικές επιστήμες στην Ελλάδα: Προβληματισμοί και προκλήσεις (Αθήνα, 16 Φεβρουαρίου 2018)

Επιστημονική επιμέλεια: Πέτρος Πετρίδης

Τιτίκα Δημητρούλια

Ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου: Νέες ερμηνευτικές προοπτικές

<https://doi.org/10.12681/praktika.5210>

© Ψηφιακή Βιβλιοθήκη ΚΕΑΕ 2023

ΨΗΦΙΑΚΟΣ ΣΥΝΕΚΔΟΤΗΣ

Εθνικό Κέντρο Τεκμηρίωσης

(<https://eproceedings.epublishing.ekt.gr/index.php/praktika>)

ΕΠΙΜΕΛΕΙΑ ΚΕΙΜΕΝΩΝ

Πελαγία Μαρκέτου

ΓΡΑΦΙΣΤΙΚΗ ΕΠΙΜΕΛΕΙΑ ΣΕΙΡΑΣ

ΣΧΕΔΙΑΣΜΟΣ ΕΞΩΦΥΛΛΟΥ

Εύη Καλογεροπούλου

ΨΗΦΙΑΚΗ ΕΚΔΟΤΙΚΗ ΥΠΟΔΟΜΗ

Γιώργος Ρεγκούκος

ΣΥΝΤΟΝΙΣΜΟΣ

Αθηνά Μποζίκια

Επιστημονική Επιτροπή

Άντα Διάλλα Καθηγήτρια Ευρωπαϊκής Ιστορίας,

Τμήμα Θεωρίας και Ιστορίας της Τέχνης, Ανώτατη Σχολή Καλών Τεχνών

Μάρω Γερμανού Καθηγήτρια Αγγλικής Λογοτεχνίας,

Τμήμα Αγγλικής Γλώσσας και Φιλολογίας, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Ελένη Γούστη-Σταμπόγλη Ιστορικός, Εκδότρια

Καλλιρρόη Λινάρδου Επίκουρη Καθηγήτρια Ιστορίας της Τέχνης του Βυζαντίου και του Δυτικού

Μεσαίωνα, Τμήμα Θεωρίας και Ιστορίας της Τέχνης, Ανώτατη Σχολή Καλών Τεχνών

Μυρτώ Μαλούτα Επίκουρη Καθηγήτρια Ελληνικής Παπυρολογίας,

Τμήμα Αρχαιονομίας, Βιβλιοθηκονομίας και Μουσειολογίας, Ιόνιο Πανεπιστήμιο

ISBN 978-618-85268-8-4

www.rchumanities.gr



ΚΕΝΤΡΟ ΕΡΕΥΝΑΣ
ΓΙΑ ΤΙΣ ΑΝΘΡΩΠΙΣΤΙΚΕΣ
ΕΠΙΣΤΗΜΕΣ

Ψηφιακή Βιβλιοθήκη

Ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου: Νέες ερμηνευτικές προοπτικές

Τιτίκα Δημητρούλια

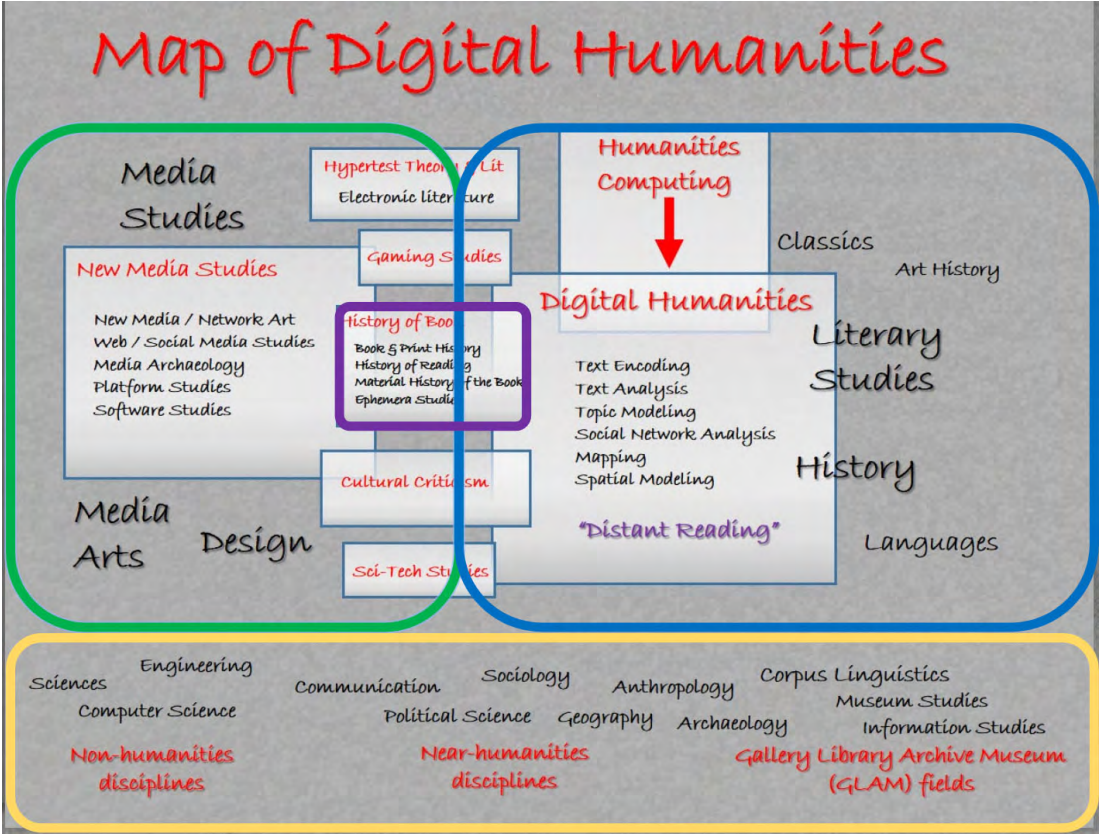
Το πεδίο της έρευνας: Το κείμενο και η ηλεκτρονική ανάλυσή του

Παρά τις όποιες απόψεις για την κυριαρχία της εικόνας στη σύγχρονη ψηφιακή πραγματικότητα, το κείμενο διατηρεί την ηγεμονική του θέση. Ο Jerome McGann περιέγραφε την «πανταχού παρουσία και δύναμη του λόγου –της λέξης (word)– στον πολιτισμό» παραφράζοντας το ποίημα του J. R. R. Tolkien για τα δαχτυλίδια στον Άρχοντα των δαχτυλιδιών και αντικαθιστώντας το ένα παντοδύναμο δαχτυλίδι με το κείμενο: το κείμενο είναι «η μία και μόνη λέξη που κυβερνά τους πάντες και στα σκοτεινά τους ρίχνει τα δεσμά» (McGann 2004, 198).¹ Οι Stéfan Sinclair και Geoffrey Rockwell (2016, 275), από την άλλη, σχολιάζουν την πολιτισμική πρωτοκαθεδρία του κειμένου κάνοντας λόγο για «πανταχού παρόν κείμενο» (ubiquitous text).

Το κείμενο βρίσκεται στην καρδιά της έρευνας στις ανθρωπιστικές επιστήμες και συνακόλουθα στις ψηφιακές ανθρωπιστικές επιστήμες (ΨΑΕ, Digital Humanities, DH). Τη θέση της ηλεκτρονικής ανάλυσης του κειμένου (textual analysis / text analysis) στις ψηφιακές ανθρωπιστικές επιστήμες την αντιλαμβάνεται κανείς εύκολα διατρέχοντας τη σχετική βιβλιογραφία² και τους καταλόγους των σχετικών έργων,³ μελετώντας τα αντίστοιχα προγράμματα σπουδών, πολλά από τα οποία προσφέρονται από τμήματα φιλολογίας (ως επί το πλείστον της αγγλικής, αλλά όχι

-
1. Οι μεταφράσεις όλων των παραθεμάτων είναι της συγγραφέα. Το ακριβές απόσπασμα από το ποίημα για τα δαχτυλίδια στον Άρχοντα των δαχτυλιδιών αφορά τη δύναμη του ενός και πανίσχυρου δαχτυλιδιού σε σχέση με τα υπόλοιπα δεκαεννέα· έχει ως εξής: «One Ring to rule them all, One Ring to find them, / One Ring to bring them all, and in the darkness bind them». Το ποίημα τίθεται ως μόντο στο πρώτο βιβλίο, το *Η συντροφιά του δαχτυλιδιού*.
 2. Βλ. λόγου χάρη τα σχετικά άρθρα σε περιοδικά της κοινότητας των ψηφιακών ανθρωπιστικών επιστημών, όπως είναι, μεταξύ άλλων, τα εξής: *Digital Humanities Quarterly*, <http://www.digitalhumanities.org/dhq>; *Digital Scholarship in the Humanities*, <https://academic.oup.com/dsh>; *Digital Studies / Le champ numérique*, <https://www.digitalstudies.org>.
 3. Για έργα των ψηφιακών ανθρωπιστικών επιστημών, βλ. τη βιβλιογραφία της υποδομής DARIAH-EU, https://www.zotero.org/groups/113737/doing_digital_humanities_-_a_dariah_bibliography. Για έργα υποδομής και περιοδικά του χώρου, βλ. τον ιστότοπο της European Association of Digital Humanities (EADH), <https://eadh.org/projects>.





Εικόνα 1. Ο χάρτης των ψηφιακών ανθρωπιστικών επιστημών σύμφωνα με τον Alan Liu, με σημειωμένες τις υποπεριοχές του.

μόνο πλέον),⁴ ή και διαβάζοντας απλώς τον γνωστό χάρτη των ψηφιακών ανθρωπιστικών επιστημών που έχει σχεδιάσει ο Alan Liu (2015) για να απεικονίσει τις μεταξύ τους διασυνδέσεις. Εδώ τον παραθέτουμε σε μια δεύτερη εκδοχή του, με υποπεριοχές σημειωμένες με μπλε περίγραμμα που δηλώνει τις κύριες ψηφιακές ανθρωπιστικές επιστήμες κατά τον Liu (Εικόνα 1).

Σύμφωνα με τον Matthew Kirschenbaum, το γεγονός ότι οι ψηφιακές ανθρωπιστικές επιστήμες θεραπεύονται σε μεγάλο βαθμό σε τμήματα φιλολογίας οφείλεται μεταξύ άλλων και στο κείμενο καθαυτό. Διότι «πρώτον, μετά τα αριθμητικά δεδομένα, το κείμενο είναι ο πιο πρόσφορος τύπος δεδομένων για επεξεργασία από τους υπολογιστές»· και, δεύτερον, «οι υπολογιστές και η σύνταξη κειμένου συνδέονται στενά εδώ και πολλά χρόνια» (Kirschenbaum 2010, 60). Η κυριαρχία του κειμένου και της ερμηνείας του στις ψηφιακές ανθρωπιστικές επιστήμες

4. Βλ. το μητρώο μαθημάτων Digital Humanities Course Registry των υποδομών CLARIN-ERIC και DARIAH-EU, επίσκεψη 30 Ιανουαρίου 2023, <https://dhcr.clarin-dariah.eu/>.

αποτυπώνεται επίσης στον χάρτη του Liu, με τις γνωστικές περιοχές που έχουν το κείμενο στο επίκεντρό τους, όπως και με τις ποικίλες προσεγγίσεις και μεθόδους επεξεργασίας του: φιλολογία, κλασική φιλολογία, ιστορία, δημοσιογραφία, γλωσσολογία σωμάτων κειμένων (corpus linguistics)· κωδικοποίηση / χαρακτηρισμός κειμένου (text encoding), ανάλυση κειμένου (text analysis), εξ αποστάσεως ανάγνωση (distant reading), θεματική μοντελοποίηση (topic modeling)...

Ο ίδιος χάρτης, πάντως, θέτει επίσης μια σειρά επιστημολογικά ζητήματα σχετικά με την ταυτότητα των ψηφιακών ανθρωπιστικών επιστημών, στον βαθμό που συμπεριλαμβάνει σε αυτές τη δημοσιογραφία και τα ΜΜΕ –δύο τομείς που συγκαταλέγονται σαφώς στις κοινωνικές επιστήμες– αλλά όχι την πολιτική επιστήμη, την κοινωνιολογία ή την ανθρωπολογία, οι οποίες χαρακτηρίζονται ως «κλάδοι όμοροι προς τις ανθρωπιστικές επιστήμες» (near-humanities disciplines). Λαμβάνοντας υπόψη μας το μανιφέστο του πρώτου ευρωπαϊκού ThatCamp στο Παρίσι (2010), που διακηρύσσει ότι «οι ψηφιακές ανθρωπιστικές επιστήμες αφορούν το σύνολο των ανθρωπιστικών, κοινωνικών επιστημών, των τεχνών και των γραμμάτων»,⁵ τη σημαντική θέση που καταλαμβάνουν το κείμενο και η ανάλυσή του στις κοινωνικές επιστήμες, στις οποίες συζητείται και χρησιμοποιείται πλέον συστηματικά η ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου σε όλες τις μορφές της (Pollach 2012· Wiedemann 2013, 2016) και την προσωπική ερευνητική μας εμπειρία, συμπεριλαμβάνουμε στις ψηφιακές ανθρωπιστικές επιστήμες τις κοινωνικές επιστήμες και τις τέχνες. Δεν ξεχνάμε, φυσικά, ότι η ανάλυση κειμένου καθαυτή προσδιορίζεται διαφορετικά σε κάθε επιστήμη και προσαρμόζεται στο ειδικό της αντικείμενο (Castelfranchi 2017· Schreibman, Siemens και Unsworth 2016· Hoover 2013· Brier και Hopp 2011), ούτε ότι οι κλασικές μέθοδοι ανάλυσης συνεχίζουν να χρησιμοποιούνται συστηματικά, καθώς οι νέες μέθοδοι απαιτούν νέες δεξιότητες, τις οποίες πολλοί ερευνητές δεν διαθέτουν (Jockers και Underwood 2015, 291-292).⁶ Ειδικά δε όσον αφορά τις κοινωνικές επιστήμες, δεν παραγνωρίζουμε ότι οι μέθοδοί τους διαφέρουν συχνά από εκείνες των ανθρωπιστικών επιστημών και προέρχονται από διαφορετικές ερευνητικές παραδόσεις (Rockwell και Sinclair 2016β, 19).

Σε μια προσπάθεια συστηματοποίησης των προσεγγίσεων και των μεθόδων της *ηλεκτρονικά υποστηριζόμενης ανάλυσης κειμένου* (Computer-Assisted ή Aided Text Analysis, CATA), έχουν προταθεί διαχρονικά διαφορετικές τυπολογίες της, οι οποίες συναρτώνται προς την ερευνητική οπτική και πρακτική που επιλέγεται,

5. Βλ. και «Manifeste des Digital humanities», <https://tcp.hypotheses.org/318>.

6. Για τη μεθοδολογία των κοινωνικών επιστημών στις ψηφιακές ανθρωπιστικές επιστήμες, βλ. Clement (2016β).

αλλά και προς το γνωστικό πεδίο. Ορισμένοι ερευνητές διαχωρίζουν την ερμηνευτική ανάλυση κειμένου (text analysis, analyse des données textuelles, ADT / textométrie / logométrie)⁷ από την εξόρυξη δεδομένων από κείμενα (text mining), ένα υποσύνολο της εξόρυξης δεδομένων από ποικίλα μέσα-υλικά (data mining), η οποία χρησιμοποιεί προηγμένα εργαλεία και μεθόδους επεξεργασίας της φυσικής γλώσσας, καθώς και τη μοντελοποίηση των δεδομένων για την ανάλυση μαζικών δεδομένων (Pincemin 2020, 2· Sinclair και Rockwell 2016, 273· Jockers και Underwood 2015, 291). Υποστηρίζουν ότι η ανάλυση κειμένου, σε αντίθεση με την εξόρυξη δεδομένων από κείμενα, αποτελεί μια καινοτόμα ανάγνωση των κειμένων με στόχο την κατανόηση και την ερμηνεία τους, στο πλαίσιο της οποίας τα εργαλεία αναλαμβάνουν «το μηχανικό μέρος της ανάλυσης και μόνο, τον ρόλο του εκτελεστικού οργάνου, ενώ στον ερευνητή εναπόκεινται πάντα η ερμηνεία, όλες οι επιλογές ανάλυσης, ο έλεγχος των διεργασιών και η ευφυΐα των παρατηρήσεων» (Pincemin 2020, 2). Διακρίνουν έτσι ανάμεσα στον *ιδιογραφικό* προσανατολισμό των ανθρωπιστικών επιστημών, «που παράγει γνώση μέσα από τη μελέτη ειδικών περιπτώσεων», και «τη *νομοθετική* τάση των επιστημών που προβαίνουν σε γενικεύσεις βασισμένες σε ευρέα δεδομένα αλλά με υψηλή τυποποίηση» (Siemens κ.ά. 2009). Άλλοι πάλι ερευνητές και άλλες ερευνήτριες τις συνεξετάζουν ως όλον, επισημαίνοντας ωστόσο τις διαφορές σε επίπεδο μεθόδου (Drucker 2021· Hoover 2013· Argamon 2009). Μια ενδιαφέρουσα τυπολογία, η οποία διακρίνει και μαζί συνεξετάζει τις διαφορετικές προσεγγίσεις της *ηλεκτρονικά υποστηριζόμενης ανάλυσης κειμένου* στις ψηφιακές ανθρωπιστικές επιστήμες, προτείνει ο Gregor Wiedemann (2013, 2015), εκκινώντας από την ανάλυση κειμένου στις κοινωνικές επιστήμες. Η τυπολογία αυτή συνοψίζεται ως εξής:

1. Χρήση λογισμικών για χειρωνακτικό χαρακτηρισμό και επισημείωση (Computer-Assisted Qualitative Data Analysis, CAQDAS), η οποία λαμβάνει υπόψη τα γλωσσικά συμφραζόμενα και την κατάσταση επικοινωνίας.
2. Ηλεκτρονική ανάλυση περιεχομένου (Computational Content Analysis, CCA), με παραγωγή αυτόματα επισημειωμένων κειμένων, σύμφωνα με τη συχνότητα εμφάνισης όρων, χωρίς μέριμνα για τα συμφραζόμενα.

7. Η κειμενομετρία-λογομετρία (textométrie-logométrie) στη γαλλόφωνη παράδοση είναι η συνέχεια της λεξικομετρίας (lexicométrie). Έτσι ονομάζεται η ερμηνευτική αναλυτική πρακτική που εξετάζει ποικίλες διαστάσεις του κειμένου και όχι μόνο το λεξιλόγιό του (Lejeune 2010· Brunet 2011· Lejeune και Bénel 2012· Rastier 2011· Lebart, Pincemin και Poudat 2019). Η αγγλόφωνη βιβλιογραφία συνεχίζει εντούτοις να χρησιμοποιεί ως επί το πλείστον τον όρο λεξικομετρία (lexicometry).

3. Λεξικομετρικές (κειμενομετρικές) μέθοδοι / μέθοδοι της γλωσσολογίας σωμάτων κειμένων με γνώμονα τα δεδομένα, οι οποίες επιτρέπουν την επαγωγική διερεύνηση γλωσσικών σχημάτων, με έμφαση στα συμφραζόμενα.
4. Μέθοδοι εξόρυξης μαζικών δεδομένων από κείμενα, με την εφαρμογή πολύπλοκων στατιστικών μοντέλων, τα οποία ερευνούν τα ρητά και τα υπόρρητα συμφραζόμενα.

Ασχέτως τυπολογιών και επιμέρους ορισμών, πάντως, που διαφέρουν όχι μόνο στις διαφορετικές επιστήμες, αλλά συχνά και στις ερευνητικές εργασίες, η ηλεκτρονική ανάλυση κειμένου είναι κατ' αρχάς πάντα ποιοτική, παρότι «οι αλγόριθμοι που τρέχουν τα εργαλεία χρησιμοποιούν ποσοτικές μεθόδους, όπως και διεργασίες αναζήτησης και ταύτισης, για να εντοπίσουν τα στοιχεία και τα χαρακτηριστικά σε κάθε κείμενο» (Drucker κ.ά. 2014, 43). Επίσης, στις διαφορετικές προσεγγίσεις και μεθόδους των ψηφιακών ανθρωπιστικών επιστημών χρησιμοποιούνται ευλόγως διαφορετικά εργαλεία, εφόσον «η σύλληψη και ο σχεδιασμός ενός έργου καθορίζουν την επιλογή των εργαλείων» (Drucker 2021, 131), όπως και στην παραδοσιακή έρευνα. Έτσι, στην ερμηνευτική ανάλυση κειμένου χρησιμοποιούνται ως επί το πλείστον εργαλεία που δεν απαιτούν γνώσεις προγραμματισμού και προϋποθέτουν μόνον την απλή εξοικείωση με τη λειτουργία τους (π.χ. AntConc, Sketch Engine Tools, Wordsmith Tools, Voyant Tools, Hyperbase, Lexico 5, Le Trameur, IRaMuTeQ) για την ανάλυση κειμένων και (ηλεκτρονικών) σωμάτων κειμένων με μεθόδους της γλωσσολογίας σωμάτων κειμένων (Γούτσος και Φραγκάκη 2015, 61-79) – αλλά όχι μόνον, όπως θα δούμε στη συνέχεια. Η γλωσσολογία σωμάτων κειμένων διανύει σήμερα την τρίτη φάση της, μετά την εμφάνιση και την καθιέρωσή της ως υποκλάδου της γλωσσολογίας (McEnery και Hardie 2012), και τείνει να ενσωματωθεί σε πολλές άλλες επιστήμες. Πάντως, τα προαναφερθέντα εργαλεία δεν προσφέρουν όλα τις ίδιες δυνατότητες ανάλυσης: άλλα είναι πιο απλά και άλλα πολύ σύνθετα· επομένως, είναι στο χέρι του ερευνητή να εκμεταλλευτεί το σύνολο των λειτουργιών τους, όπως π.χ. στα Voyant Tools, Hyperbase, Lexico 5, Le Trameur, IRaMuTeQ. Άλλα εργαλεία, εξίσου απλά στη χρήση τους, αλλά προερχόμενα από διαφορετική ερευνητική παράδοση (Rockwell και Sinclair 2016β, 19), όπως το Jstor Labs Text Analyzer, χρησιμοποιούν τη θεματική μοντελοποίηση στην ανάλυση κειμενικού περιεχομένου· ή υποστηρίζουν την ποιοτική ανάλυση ποικίλης φύσης δεδομένων, κυρίως στις κοινωνικές επιστήμες (ATLAS.ti, NVivo, MAXQDA, αλλά και IRaMuTeQ) – συχνά ενσωματώνοντας σήμερα και κειμενομετρικά εργαλεία, όπως είναι λόγου χάρη το MaxDictio στο MAXQDA Plus. Άλλα εργαλεία απαιτούν αυξημένες ψηφιακές δεξιότητες, αλλά όχι γνώσεις προγραμματισμού, όπως το «MALLET natural language processing toolkit», που επιτρέ-

πει μια σειρά από προηγμένες αναλύσεις κειμένου.⁸ Τέλος, υπάρχουν και εκείνα που απαιτούν γνώση γλωσσών προγραμματισμού, κυρίως της Python –η δεύτερη γλώσσα που χρησιμοποιείται συστηματικά στη γλωσσική τεχνολογία (Language Technology ή Human Language Technology) είναι η R–, όπως μεταξύ άλλων το πακέτο βιβλιοθηκών και εργαλείων Natural Language Toolkit-NLTK, το Gensim και το TextBlob, δίνοντας μεγάλες δυνατότητες ανάλυσης μαζικών δεδομένων.

Ηλεκτρονικά σώματα κειμένων, «σακιά με λέξεις» και ερμηνευτική

Όλα τα εργαλεία προϋποθέτουν την εφαρμογή τους σε ψηφιακά κείμενα και σώματα κειμένων. Μπορούν προφανώς να χρησιμοποιηθούν και σε ευσύνοπτα κείμενα, αλλά η χρήση τους παρουσιάζει αυξημένο ενδιαφέρον σε μεγαλύτερα ή πολύ μεγάλα ηλεκτρονικά σώματα κειμένων (ΗΣΚ), τα οποία ο ανθρώπινος νους δεν μπορεί να προσπελάσει χωρίς δυσκολία. Τα ηλεκτρονικά σώματα κειμένων βρίσκονται στη βάση των ψηφιακών ανθρωπιστικών επιστημών και αναλύονται σήμερα με ποικίλες μεθόδους αναλόγως με το υλικό τους, που μπορεί πέραν των γραπτών κειμένων να είναι ηχητικά αρχεία, εικόνες, βίντεο (Barber 2016· Budzise-Weaver 2016· Clement 2016α· Manovich 2015· O'Halloran 2015).

Ως ηλεκτρονικό σώμα κειμένων ορίζεται ένα σύνολο ψηφιακών, αυθεντικών γραπτών, προφορικών ή πολυτροπικών κειμένων, ποικίλης έκτασης, συγκροτημένο και μορφοποιημένο με βάση συγκεκριμένα κριτήρια και σε σχέση με συγκεκριμένους ερευνητικούς στόχους, το οποίο μπορεί να αποτελέσει αντικείμενο επεξεργασίας με ηλεκτρονικά εργαλεία. Οι δύο κύριες αναλυτικές προσεγγίσεις των ηλεκτρονικών σωμάτων κειμένων στη γλωσσολογία σωμάτων κειμένων είναι η προσέγγιση *με βάση* τα σώματα κειμένων (Corpus-Based Approach, CBA) και *με γνώμονα* τα σώματα κειμένων (Corpus-Driven Approach, CDA). Στην πρώτη περίπτωση, οι ερευνητές χρησιμοποιούν τα ηλεκτρονικά σώματα κειμένων για επαλήθευση ερευνητικών υποθέσεων, ενώ στη δεύτερη η προσέγγιση διαμορφώνεται επαγωγικά από τα δεδομένα που προκύπτουν από την ανάλυση των ηλεκτρονικών σωμάτων κειμένων (Tognini-Bognelli 2001· Biber 2010). Οι προσεγγίσεις αυτές έχουν εφαρμογή και στην ανάλυση κειμένου, αναλόγως με τα έργα. Η ιδιαιτερότητα του γνωστικού πεδίου καθορίζει επίσης τη συγκρότηση του ηλεκτρονικού

8. Βλ. McCallum 2002, <https://mimno.github.io/Mallet/index>.

σώματος κειμένων σε επίπεδο επιλογής κειμένων, στάθμισης και αντιπροσωπευτικότητας (σε ένα ηλεκτρονικό σώμα κειμένων νεοελληνικής λογοτεχνίας, για παράδειγμα, μπορεί να έχουμε κείμενα του ρεαλισμού, του μοντερνισμού, σε γλώσσα δημοτική ή καθαρεύουσα κ.λπ., αλλά με βάση το ερευνητικό ερώτημα το δείγμα μας θα πρέπει να πληροί συγκεκριμένες προϋποθέσεις).

Η κειμενομετρία αφορά προφανώς τα σώματα γραπτών κειμένων ή κειμένων που έχουν μεταγραφεί με παραδοσιακές ή νέες μεθόδους (π.χ. με εργαλεία μετατροπής της ομιλίας σε κείμενο). Προϋπόθεσή της είναι η γνώση του σώματος κειμένων, πράγμα που περιορίζει εξ ορισμού το μέγεθός του. Πέραν τούτου, όπως σημειώνει ο Rockwell (2003, 209) στο κείμενό του «What is Text Analysis, Really?», ειδικά στις ανθρωπιστικές σπουδές, η ερμηνευτική –ιδιογραφική– ανάλυση κειμένου πάντα σχετίζεται τόσο με τη χρήση και την ψηφιακή αναπαράσταση των κειμένων όσο και με τις θεωρίες της κειμενικότητας. Προς επίρρωσιν της θέσης για τη σχέση της ανάλυσης κειμένου και της ερμηνείας, ως συνδυασμού ανάλυσης και σύνθεσης σε όλα τα επιστημονικά πεδία των ψηφιακών ανθρωπιστικών επιστημών, παραθέτουμε το απόσπασμα από τον Κοντιγιάκ που προτάσσει ο Rockwell στο ίδιο κείμενό του:⁹

Κατά τη γνώμη μου η ανάλυση είναι μια διεργασία που προκύπτει από τη σύγκλιση προηγούμενων διεργασιών. Συνίσταται στη σύνθεση και την ανασύνθεση των ιδεών μας, έτσι ώστε να τις συγκρίνουμε και να ανακαλύψουμε με αυτό τον τρόπο τις σχέσεις που έχουν μεταξύ τους, όπως και τις νέες ιδέες που αυτές μπορούν να παραγάγουν.¹⁰

Οι Sinclair και Rockwell (2016, 285) θέτουν έτσι στο επίκεντρο της ανάλυσης κειμένου την ερμηνευτική διαδικασία, «η οποία μπορεί να μάθει από και να οικοδομήσει πάνω στη γλωσσολογία σωμάτων κειμένων, στην υφολογία και στις μεθόδους ανάλυσης μαζικών δεδομένων» (Rockwell και Sinclair 2016β, 19). Χρησιμοποιούν δε τον όρο *hermeneutica* «για τα διαδραστικά και ερμηνευτικά εργαλεία ανάλυσης, που διευκολύνουν την αναδιάταξη και τον χειρισμό των κειμένων για την καλύτερη μελέτη και την κατανόησή τους» (Sinclair και Rockwell 2016, 285).

Ο Rockwell (2003), όπως και άλλοι ερευνητές, θεωρεί τα εργαλεία της ηλεκτρονικά υποστηριζόμενης ανάλυσης κειμένου εργαλεία «επαύξησης» (*augmentation*) των ανθρώπινων νοητικών ικανοτήτων. Συνομιλεί έτσι με τη θέση του John Unsworth (2001) ότι η «υπολογιστική αναπαράσταση μεγάλων όγκων υλικού, που

9. Το Rockwell και Sinclair (2016β) έχει τον χαρακτηριστικό τίτλο *Hermeneutica* και ξεκινά με τον Λόγο περί της μεθόδου του Καρτέσιου και τη σχέση του με τον στοχασμό στις ανθρωπιστικές επιστήμες.

10. Ο Rockwell παραπέμπει στην αγγλική μετάφραση, Condillac, *Essay on Understanding*, 48. Η μετάφρασή μας έλαβε υπόψη και το πρωτότυπο κείμενο, Condillac, *Essai sur l'origine*, 65.

έχει κωδικοποιηθεί και έχει υποστεί επεξεργασία σύμφωνα με μια αυστηρή, καλά σχεδιασμένη αναπαράσταση της γνώσης, προσφέρει ευκαιρίες για να αντιληφθούμε και να αναλύσουμε μοτίβα, συνδυασμούς, διασυνδέσεις και απουσίες στοιχείων, τα οποία ο άνθρωπος χωρίς τη βοήθεια του υπολογιστή δεν θα ήταν σε θέση να εντοπίσει».

Η ανάλυση κειμένου σε όλες τις μορφές της προϋποθέτει την ύπαρξη ψηφιακών ή ψηφιοποιημένων κειμένων, σε μορφή επεξεργάσιμη από εργαλεία. Το ψηφιοποιημένο, σε μορφή εικόνας, τυπωμένο κείμενο δεν μπορεί να αποτελέσει αντικείμενο κανενός είδους επεξεργασίας, όπως στην περίπτωση της σαρωμένης πρώτης σελίδας από την ποιητική συλλογή *Ατθίδες αύραι* του Γεώργιου Βιζυηνού, στην οποία το ποίημα «Οι γάμοι» ξεκινά για να συνεχιστεί στην επόμενη σελίδα της έκδοσης.¹¹

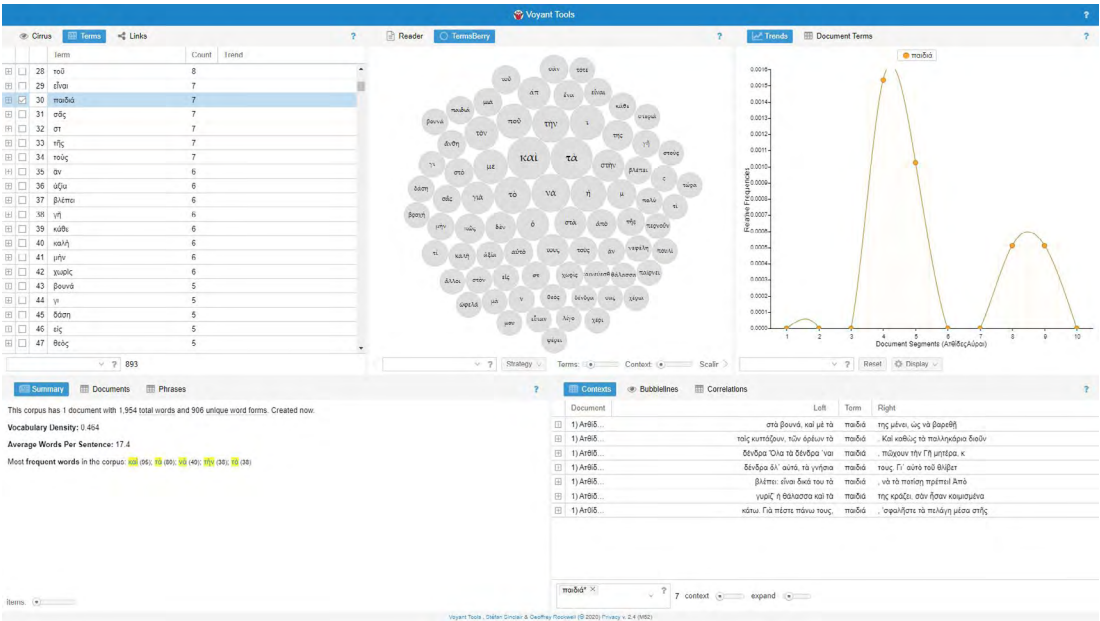
Για να μπορέσει να χρησιμοποιήσει αυτή τη σελίδα ένα εργαλείο, πρέπει να μετατραπεί σε κείμενο. Το ίδιο ποίημα, όπως παρουσιάζεται στη Βικιθήκη, έχει ήδη σημαντική διαφορά, αφού στη συγκεκριμένη μορφή του κειμένου μπορεί κανείς να παρέμβει, να το αντιγράψει εν μέρει ή εν όλω και να αλλάξει εύκολα τον μορφότυπό του, ώστε να μπορεί να το επεξεργαστεί με εργαλεία – τα οποία συχνά προϋποθέτουν συγκεκριμένο μορφότυπο (ο πιο ασφαλής είναι το απλό κείμενο, το txt, με κωδικοποίηση UTF-8), παρότι ο κατάλογος των μορφότυπων που επεξεργάζονται πολλά εργαλεία διαρκώς διευρύνεται.¹²

Όπως και να έχει, συγκεντρώνοντας τα ποιήματα σε ένα αρχείο –ή και σε περισσότερα, ώστε να έχει τις πληροφορίες των τίτλων και χωριστά αποτελέσματα για το καθένα–, χρησιμοποιώντας τον κατάλληλο μορφότυπο, με ελάχιστο έλεγχο και επιμέλεια μπορεί κανείς να τα επεξεργαστεί σε διάφορα επίπεδα, όπως φαίνεται στην εικόνα από τα Voyant Tools (**Εικόνα 2**), που παρουσιάζει μέρος του καταλόγου λέξεων (wordlist) δέκα ποιημάτων της συλλογής, τις υψηλής συχνότητας λέξεις με τις συνάψεις τους οπτικοποιημένες στη λειτουργία TermBerry, την οπτικοποιημένη κατανομή της λέξης «παιδιά», η οποία έχει επτά εμφανίσεις στα ποιήματα, και το περικείμενό της στον συμφραστικό πίνακα λέξεων (concordance).

Το συγκεκριμένο παράδειγμα μοναδικό στόχο έχει να δείξει τη διεπαφή ενός κειμενομετρικού εργαλείου και δεν παρουσιάζει ευρετικό ενδιαφέρον στο πλαίσιο μιας ερευνητικής υπόθεσης. Από την τελευταία εικόνα, όμως, μπορεί κανείς

11. Βλ. Γεώργιος Μ. Βιζυηνός, «Οι γάμοι», *Ατθίδες αύραι: Συλλογή ποιημάτων*, 2η έκδοση (1884), 1, Ανέμη, επίσκεψη 19 Νοεμβρίου 2022, Ψηφιακή Βιβλιοθήκη Πανεπιστημίου Κρήτης, <https://anemi.lib.uoc.gr/metadata/8/2/b/metadata-22-0000207.tkl>.

12. Βλ. Γεώργιος Μ. Βιζυηνός, «Οι γάμοι», *Ατθίδες αύραι*, Βικιθήκη, επίσκεψη 19 Νοεμβρίου 2022, https://el.wikisource.org/wiki/%CE%9F%CE%B9_%CE%93%CE%AC%CE%BC%CE%BF%CE%B9.



Εικόνα 2. Αποτελέσματα ανάλυσης των δέκα πρώτων ποιημάτων της συλλογής του Γ. Βιζυηνού *Ατθίδες αὖραι* (Voyant Tools).

να καταλάβει γιατί το κείμενο ορίζεται ως «σάκος λέξεων»,¹³ «bag of words» (BoW), στην επεξεργασία φυσικής γλώσσας (natural language processing, NLP). Ή γιατί ο Rockwell (2003, 213) ορίζει το κείμενο που προκύπτει από την ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου, όπως τα παραπάνω δέκα ποιήματα του Βιζυηνού, ως συνδυασμό του κειμένου του συγγραφέα, των επιλογών του αναγνώστη και των διαδικασιών επεξεργασίας του, και μιλά για κείμενο «υβρίδιο» ή και «τέρας», ενώ μαζί με τον Sinclair χαρακτηρίζουν τη σύνθεσή του ανάλογη με τη δημιουργία του Φρανκενστάιν (Sinclair και Rockwell 2016, 284).

Οι συμφραστικοί πίνακες λέξεων (ΣΠΛ), που είναι ένα από τα παλαιότερα εργαλεία κειμενικής ανάλυσης, κατέχουν κεντρική θέση στα κειμενομετρικά εργαλεία. Ο πρώτος συμφραστικός πίνακας λέξεων ήταν αυτός της *Βουλγάτας* και καταρτίστηκε χειρωνακτικά τον 13ο αιώνα από τον καρδινάλιο Hugues de Saint-Chef¹⁴ και μια ομάδα πεντακοσίων δομινικανών μοναχών (Albaric 2004).

Αντίστοιχη θέση, όμως, κατέχουν οι ηλεκτρονικοί συμφραστικοί πίνακες λέξεων στο πεδίο της πληροφορικής στις ανθρωπιστικές επιστήμες (humanities computing) κατά την πρώτη φάση της ανάπτυξης των ψηφιακών ανθρωπιστικών

13. Μια άλλη απόδοση του όρου είναι «σύνολο όρων ή λέξεων».

14. Άλλες γραφές του ονόματος του καρδινάλιου αυτού είναι Hugo et de Saint Cher, Hugo de Sancto C(h)aro-βλ. το ψηφιοποιημένο έργο στο Hugo a Sancto Caro: Concordantiae Bibliae (Concordantiae Sancti Jacobi minores), Bayersiche StaatBibliothek, <https://bildsuche.digitale-sammlungen.de/index.html?c=viewer&b=andnumber=bsb00034253&pimage=00004&einzelsegment=&v=100&l=es>.

επιστημών (Fraser 1996· Hockey 2004· Burton 1981α, β· Rockwell 2003· Sinclair και Rockwell 2018). Ο πρώτος τέτοιος πίνακας, ο *Index Thomisticus* (1951), που υπήρξε δημιούργημα του καθολικού ιερέα Roberto Busa, του πατέρα των ψηφιακών ανθρωπιστικών επιστημών και ειδικότερα της ανάλυσης κειμένου, πραγματοποιήθηκε σε συνεργασία με την IBM και αφορούσε το έργο του Θωμά Ακινάτη (Busa 1980· Winter 1999).¹⁵ Πολύ νωρίς, εξάλλου, οι ηλεκτρονικοί συμφραστικοί πίνακες λέξεων αναγνωρίστηκαν ως ιδιαιτέρως χρήσιμοι για πολλές επιστήμες (Tasman 1957, 256).

Ως συμφραστικός πίνακας λέξεων (ΣΠΛ) νοείται ο κατάλογος όλων των λέξεων τις οποίες περιλαμβάνει ένα μεμονωμένο κείμενο ή ένα ηλεκτρονικό σώμα κειμένων, με αναφορά στη θέση τους σε αυτό και με εμφάνιση του συγκεκριμένου τους.¹⁶ Στα κειμενομετρικά λογισμικά, ο κατάλογος λέξεων καταγράφει τη συχνότητα και τη θέση εμφάνισης της κάθε λέξης (με ή χωρίς λημματοποίηση), οδηγεί συνήθως στο κείμενο καθαυτό ή σε απλούστερες ή πιο πολύπλοκες απεικονίσεις της ανάλυσής του, και δίνει οπωσδήποτε τη δυνατότητα εμφάνισης του συγκεκριμένου (keyword in context), το οποίο είναι στενότερο ή ευρύτερο με βάση τη ρύθμιση του ερευνητή. Επιπλέον, δίνει τη δυνατότητα εύρεσης των λεξικών και γραμματικών συνάψεων (collocations και colligations αντίστοιχα) ή των λεξικών συμπλεγμάτων (lexical bundles ή clusters), που είναι «ακολουθίες περισσότερων από δύο λέξεις με στατιστικά σημαντική συνεμφάνιση σε ένα κείμενο ή σε ένα σώμα κειμένων, αλλά και η ευρύτερη έννοια των ν-γράφων (n-grams), που αφορά τις ακολουθίες οποιωνδήποτε στοιχείων της γλώσσας με στατιστικά σημαντική συνεμφάνιση» (Γούτσος και Φραγκάκη 2015, 70).

Μπορεί κανείς εύκολα να φανταστεί την ευρετική σημασία της κειμενομετρικής επεξεργασίας σε έργα συγκεκριμένων ποιητών ή συγγραφέων, ή σε ηλεκτρονικά σώματα κειμένων που συγκεντρώνουν, λόγου χάρη, έργα από διάφορα λογοτεχνικά είδη ή ρεύματα με συγκεκριμένα κριτήρια. Μπορεί επίσης να φανταστεί πόσο χρόνο θα χρειαζόταν μια τέτοια ανάλυση, αλλά και τους περιορισμούς της, χωρίς την «επαυξητική» συνδρομή της τεχνολογίας.

Στην ερμηνευτική ανάλυση των κειμένων, όλες οι λέξεις μπορεί να έχουν σημασία, αλλά αναλόγως με το ερευνητικό ερώτημα, μια σειρά από λέξεις μπορεί επίσης να αποκλειστεί από την επεξεργασία – συνήθως λέξεις λειτουργικές ή εκείνες που ο ερευνητής αποφασίζει ότι πρέπει να εξαιρεθούν για συγκεκριμένους λόγους (stop words ή stopwords). Υπάρχουν έτοιμοι κατάλογοι εξαιρούμενων λέξεων, οι οποίοι εφαρμόζονται ή εισάγονται με απλό τρόπο στα εργαλεία, αλλά

15. Βλ. και Roberto Busa κ.ά., *Corpus Thomisticum*, <http://www.corpusthomisticum.org/it/index.age>.

16. Σημαντικό έργο όσον αφορά την κατάρτιση συμφραστικών πινάκων λέξεων για κείμενα αρχαίων συγγραφέων, αλλά και νεοελλήνων ποιητών, έχει να επιδείξει το Κέντρο Ελληνικής Γλώσσας.

μπορεί κανείς να δημιουργήσει και προσαρμοσμένους καταλόγους.¹⁷ Εναπόκειται πάντως στον ερευνητή να σταθμίσει κάθε φορά αν τον καλύπτουν οι έτοιμοι κατάλογοι εξαιρούμενων λέξεων, αν θα δημιουργήσει ο ίδιος τον δικό του ή αν δεν θα χρησιμοποιήσει καν εξαιρούμενες λέξεις.

Ο Jonathan Culpeper (2002, 2009), λόγου χάρη, χρησιμοποιώντας εργαλεία Wordsmith, κατέληξε σε πολύ ενδιαφέροντα συμπεράσματα για το ύφος των έξι προσώπων που μιλούν περισσότερο στο έργο *Ρωμαίος και Ιουλιέτα* μέσα από την εξέταση και των λειτουργικών λέξεων. Ο Culpeper συνέκρινε τον λόγο του κάθε προσώπου με το σύνολο του λόγου των υπόλοιπων εξεταζόμενων προσώπων, επιστρατεύοντας μια συγκεκριμένη ερευνητική συλλογιστική, σχετική με τη σημασία της αλληλεπίδρασης μεταξύ προσώπων για τη συγκρότησή τους σε συγκεκριμένο πλαίσιο, το οποίο ορίζεται από το έργο καθαυτό. Εντόπισε τις πιο σημαντικές λέξεις (keywords) στον λόγο τους. Μεταξύ άλλων, παρατήρησε ότι η Ιουλιέτα χρησιμοποιεί πολύ συχνά τον σύνδεσμο «if» (εάν), ο οποίος είναι δηλωτικός της ανησυχίας της. Έτσι, μέσα από μια ερμηνευτική ανάλυση σε ηλεκτρονικά σώματα κειμένων περιορισμένης έκτασης, μια λειτουργική λέξη ανέδειξε μια πτυχή της ψυχολογίας της Ιουλιέτας.

Η Drucker (2011) ονομάζει τα δεδομένα που χρησιμοποιούνται σε ένα έργο όχι πια «data», αφού αυτά είναι τα δεδομένα που υπάρχουν γενικώς προς χρήση, αλλά «carpa», δεδομένα που έχουν επιλεγεί και η επιλογή τους αποτελεί ήδη ερμηνευτική κίνηση – όπως ο λόγος των έξι προσώπων στον Culpeper.¹⁸ Θα δώσουμε ένα δεύτερο παράδειγμα ανάλυσης ειδικά επιλεγμένων δεδομένων, δηλαδή carpa, αυτή τη φορά στο πεδίο της πολιτικής επιστήμης. Στο Παρατηρητήριο Λαϊκιστικού Λόγου, που δημιουργήθηκε στο πλαίσιο ενός ερευνητικού προγράμματος για τη μελέτη του λαϊκιστικού φαινομένου,¹⁹ οι χρήστες είχαν τη δυνατότητα να χρησιμοποιούν ήδη υπάρχοντα ηλεκτρονικά σώματα κειμένων, συγκροτημένα με βάση συγκεκριμένες λέξεις-κλειδιά, επιλεγμένες σύμφωνα με το θεωρητικό

17. Στην πλατφόρμα GitHub (<https://github.com/>) όπου βρίσκει κανείς έργα ελεύθερης πρόσβασης, τον κώδικά τους και βοηθητικό υλικό για τη χρήση τους, υπάρχουν επίσης κατάλογοι εξαιρούμενων λέξεων για πολλές γλώσσες, στις οποίες συμπεριλαμβάνονται τα ελληνικά.

18. Αν τα δεδομένα είναι αναλογικά, όπως συμβαίνει πολύ συχνά στην ελληνική πραγματικότητα, τότε πρέπει να ψηφιοποιηθούν. Αυτό μπορεί να γίνει είτε με λογισμικά οπτικής αναγνώρισης χαρακτήρων (Optical Character Recognition, OCR), οπότε χρειάζονται μικρή ή μεγαλύτερη επιμέλεια, αφού στα ελληνικά τα σχετικά λογισμικά παρουσιάζουν ιδιαίτερα προβλήματα, ειδικά στα πολυτονικά κείμενα (Γούτσος και Φραγκάκη 2014)· είτε με άλλο πρόσφορο τρόπο (π.χ. πληκτρολόγηση), η οποία ωστόσο είναι κοπιώδης και δυσχερής όταν αφορά εκτενή ηλεκτρονικά σώματα κειμένων. Με δεδομένες αυτές δυσκολίες, αναδεικνύεται πιο έντονα η αναγκαιότητα της συλλογικής δουλειάς και των ανοιχτών, ελεύθερα προσβάσιμων γλωσσικών πόρων στις ψηφιακές ανθρωπιστικές επιστήμες.

19. Επιστημονικός υπεύθυνος ήταν ο Γιάννης Σταυρακάκης, καθηγητής πολιτικής επιστήμης στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης. Για το έργο, βλ. <https://goo.gl/WdyrCd>· βλ. επίσης Νίκος Νικήσιανης κ.ά. (2016) και Nikos Nikisianis κ.ά. (2018).

μοντέλο της σχολής του Έσεξ, ή να συγκροτούν αντίστοιχα νέα ηλεκτρονικά σώματα κειμένων από εφημερίδες ή tweets, με βάση τις συγκεκριμένες λέξεις και πάλι, και να τα αναλύουν κειμενομετρικά. Οι λέξεις-κλειδιά όριζαν εξαρχής ερμηνευτικά τη συγκρότηση των ηλεκτρονικών σωμάτων κειμένων, ασχέτως της περαιτέρω επεξεργασίας τους.

Ενώ όμως οι ερμηνευτικές προοπτικές που ανοίγει σήμερα η ανάλυση κειμένου διευρύνονται διαρκώς, με δεδομένη την αύξηση των διαθέσιμων ψηφιακών κειμένων, οι ανισότητες μεταξύ των γλωσσών (de Swaan 2001· Casanova 1999) παραμένουν και στο ψηφιακό περιβάλλον: η αγγλική ως υπερκεντρική γλώσσα κυριαρχεί και η ελληνική κινδυνεύει με εξαφάνιση λόγω της ανεπαρκούς υποστήριξης στο επίπεδο της γλωσσικής τεχνολογίας και των γλωσσικών πόρων, όπως σημείωνε στην έκθεσή της η ομάδα META-NET (Gavriliadou κ.ά. 2012). Οι ηλεκτρονικοί πόροι και η ανάλυσή τους ήταν δύο από τους δείκτες της μελέτης, ως προς τους οποίους η ελληνική κατατάσσεται στις γλώσσες με αποσπασματική υποστήριξη.

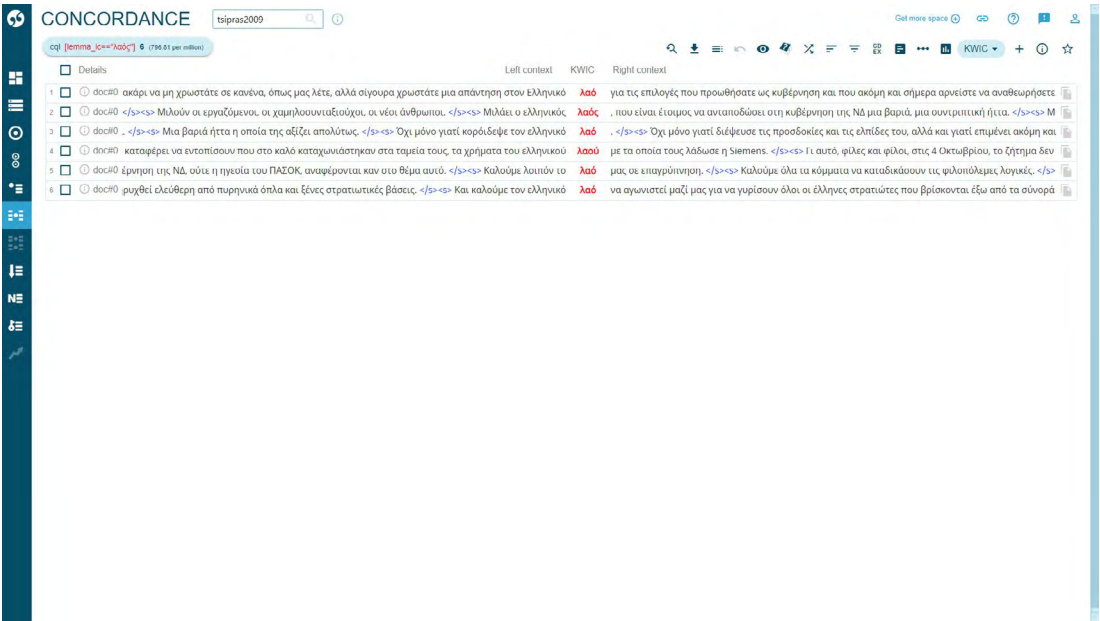
Οπτικοποίηση δεδομένων, χαρακτηρισμός και επισημείωση ηλεκτρονικών σωμάτων κειμένων

Οι Sinclair και Rockwell (2016, 276, 278) προσεγγίζουν στην ανάλυση κειμένου τα κείμενα σφαιρικά, στο πλαίσιο ενός «ευέλικτου ερμηνευτικού κύκλου»: «Διαβάζουμε κείμενα που μας αρέσουν, στη συνέχεια τα εξερευνούμε και τα μελετούμε με εργαλεία ανάλυσης και διεπαφές οπτικοποίησης, μέσα από τα οποία επανερχόμαστε πάλι στα κείμενα και τα ξαναδιαβάζουμε με διαφορετικό τρόπο». Δηλαδή, υιοθετούν μια συλλογική, διαλογική, συνεργατική πρακτική την οποία ονομάζουν «ευέλικτη ερμηνευτική» (agile hermeneutics, AH, Rockwell και Sinclair 2016β, 7). Ο κύκλος αυτός, όπως διαπιστώνουμε, περιλαμβάνει ποικίλα εργαλεία ανάλυσης, μαζί και την οπτικοποίηση, δηλαδή την οπτική απεικόνιση της ανάλυσης των δεδομένων (Sinclair και Rockwell 2016, 278). Η οπτικοποίηση βρίσκεται στο επίκεντρο της πρακτικής, αλλά και της συζήτησης που διεξάγεται στις ψηφιακές ανθρωπιστικές επιστήμες, και αφορά όλους τους τύπους ηλεκτρονικής ανάλυσης κειμένου (Drucker 2011, 2016, 2017, 2020· Kucher και Kerren 2015· Clement 2013· Sinclair, Ruecker και Radzikowska 2013).

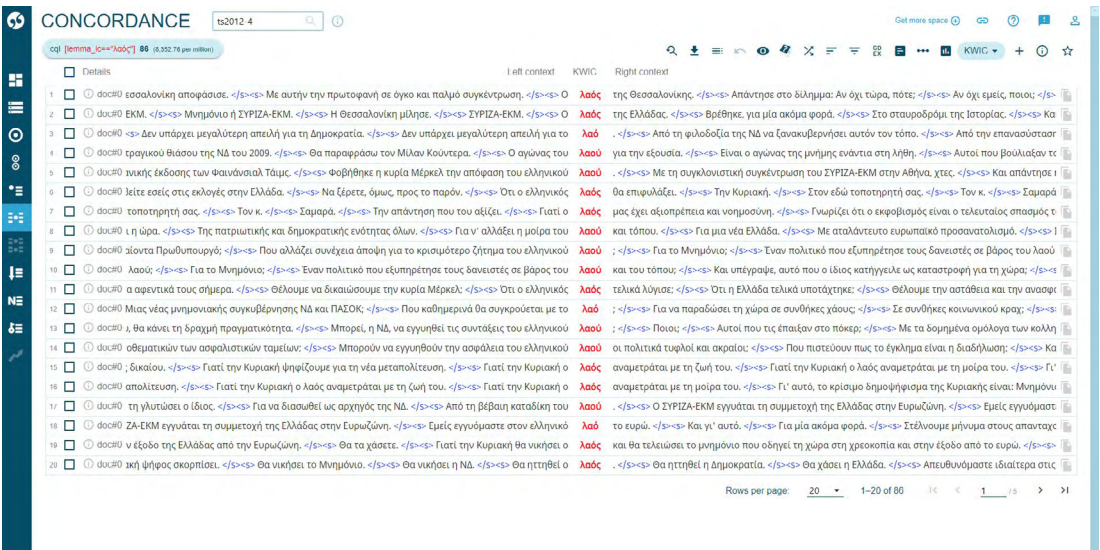
Δεν είναι τυχαίο ότι σε ένα σχήμα εμπνευσμένο από τον χάρτη των ψηφιακών ανθρωπιστικών επιστημών του Liu, που αποτυπώνει τους κύριους τρόπους με τους οποίους μπορεί ο υπολογιστής να υποστηρίξει τον ερευνητή στην κατανόηση και την ερμηνεία των κειμένων, ο Underwood (2015) θέτει στο επίκεντρο την οπτικοποίηση μεμονωμένων κειμένων (visualize single texts).

Η ποικιλία των μεθόδων οπτικοποίησης γίνεται άμεσα αντιληπτή στην πλατφόρμα Text Visualization Browser, όπου το 2021 καταγράφονται 440 τεχνικές, με παραπομπές στις εργασίες από τις οποίες προέρχονται. Σε σχέση με την αμεσότητα της οπτικοποίησης, θα αναφέρουμε ένα ακόμα παράδειγμα από την πολιτική επιστήμη: πρόκειται για τη συγκριτική ανάλυση των πολιτικών λόγων του Αλέξη Τσίπρα το 2009 (δύο ομιλίες, από τις εθνικές εκλογές και τις ευρωεκλογές), το 2012 (τρεις ομιλίες από τις εθνικές εκλογές, στην Αθήνα και τη Θεσσαλονίκη) και το 2014 (μία ομιλία από τις ευρωεκλογές, στην Αθήνα), η οποία παρουσιάστηκε από τους Γούτσο και Δημητρούλια σε μεθοδολογικό εργαστήριο στο πλαίσιο του προαναφερθέντος ερευνητικού προγράμματος για τον λαϊκισμό. Στόχος της ανάλυσης ήταν να αναζητήσει στοιχεία που θα επιβεβαίωναν ή θα κατέρριπταν τη θέση ότι οι λόγοι του Τσίπρα εμφάνιζαν αυξανόμενα στοιχεία λαϊκισμού, όπως αυτός ορίζεται στην πολιτική θεωρία από τη σχολή του Έσεξ, εφαρμόζοντας μια προσέγγιση με βάση τα σώματα κειμένων.

Στο ηλεκτρονικό σώμα κειμένων με τους λόγους του Αλέξη Τσίπρα το 2009, στο Sketch Engine, το λήμμα «λαός», ως πυρηνικός όρος του λαϊκισμού, αριθμούσε έξι εμφανίσεις, από τις οποίες οι δύο σε συνεμφάνιση με το ρήμα «καλώ», στοιχείο που συναρτάται προς την περίπτωση επικοινωνίας των εκλογών και ως εκ τούτου αποδυναμώνεται ερμηνευτικά (Εικόνα 3). Αντίθετα, στους λόγους του 2012 και του 2014, το λήμμα «λαός» στο ίδιο λογισμικό είχε 86 εμφανίσεις (Εικόνα 4).



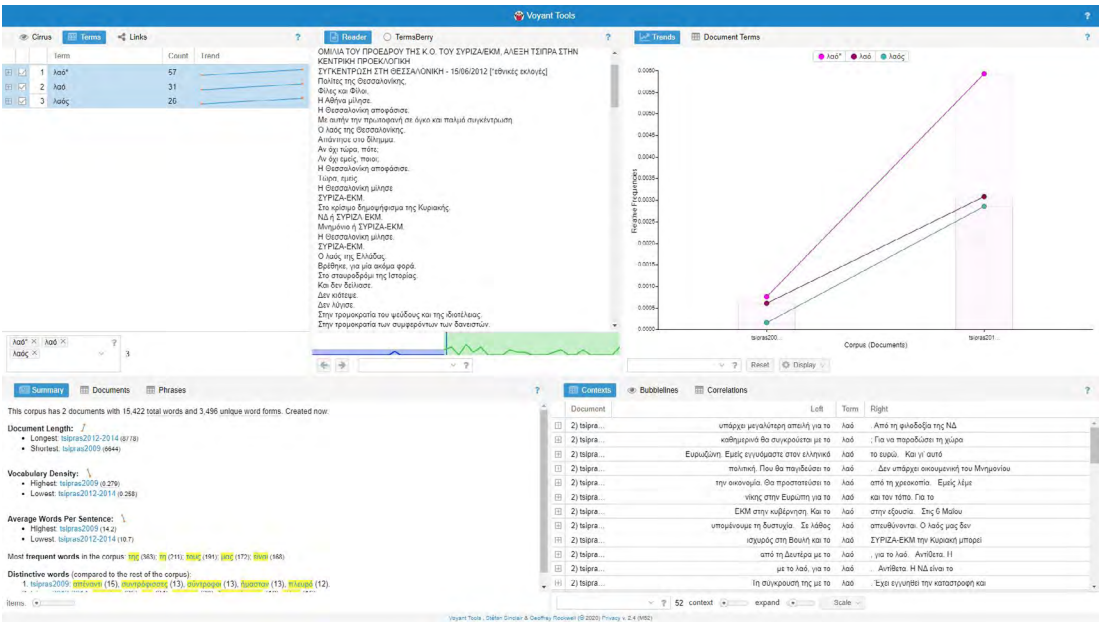
Εικόνα 3. Συμφραστικός πίνακας λέξεων του λήμματος «λαός» στους πολιτικούς λόγους του Αλέξη Τσίπρα, 2009 (Sketch Engine).



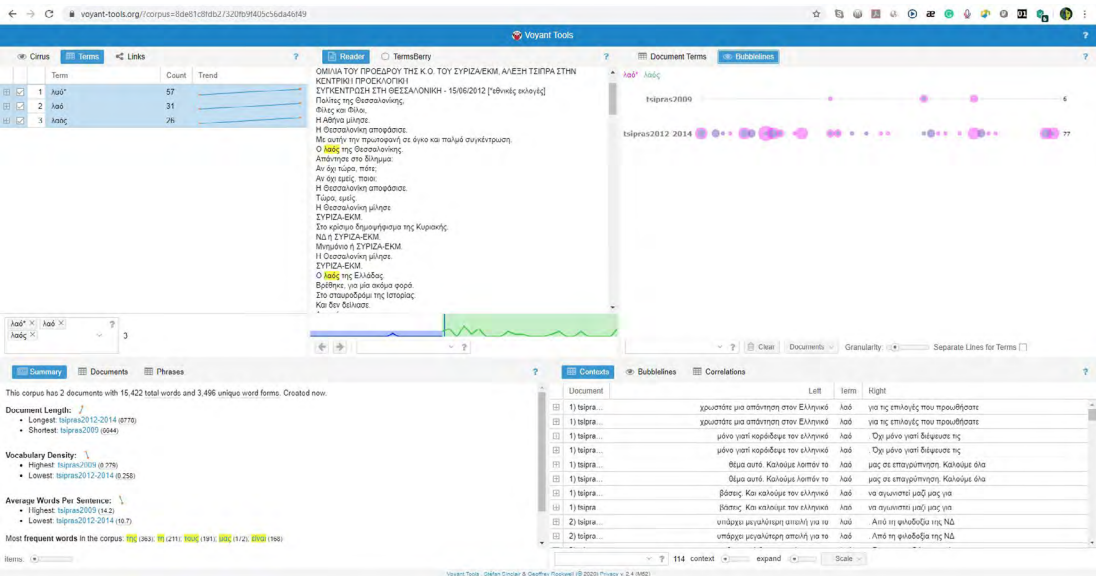
Εικόνα 4. Συμφραστικός πίνακας λέξεων του λήμματος «λαός» στους πολιτικούς λόγους του Αλέξη Τσίπρα, 2012-2014 (Sketch Engine).

Η συγκριτική οπτικοποίηση στα δύο ηλεκτρονικά σώματα κειμένων, των λόγων του 2009 και των ετών 2012-2014, με δύο διαφορετικές τεχνικές των Voyant Tools, δίνει μια άμεση εικόνα της διαφοράς αυτής (Εικόνα 5 και Εικόνα 6).

Τα συνοδικά αποτελέσματα αυτής της πολύ απλής ανάλυσης επιβεβαίωσαν τη θέση των πολιτικών επιστημόνων σχετικά με την αλλαγή ρητορικής στους λόγους



Εικόνα 5. Συγκριτική οπτικοποίηση του όρου «λαός»: κατανομή (Trends, Voyant Tools).

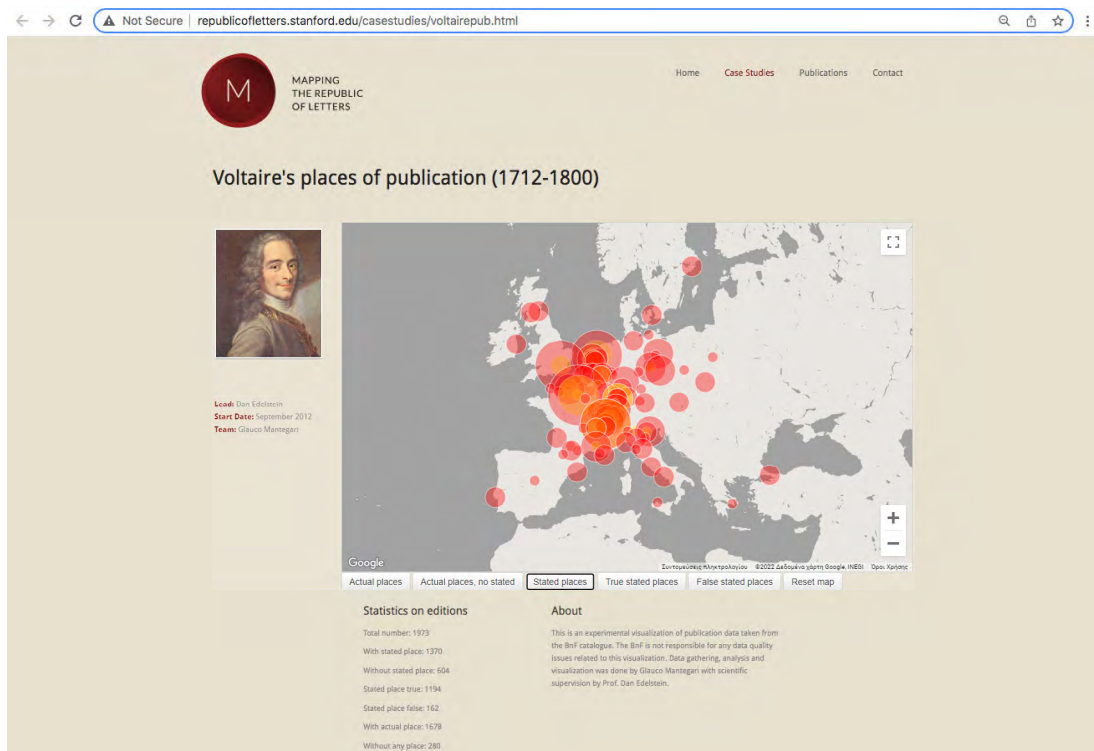


Εικόνα 6. Συγκριτική οπτικοποίηση του όρου «λαός»: γραμμές φυσαλίδων (Bubblelines, Voyant Tools).

του Αλέξη Τσίπρα από το 2009 ως το 2014, και η οπτικοποίηση ανέδειξε με άμεσο τρόπο την αλλαγή.

Η αμεσότητα σχετίζεται με τη «ρητορική ισχύ» που διαθέτουν οι γραφικές απεικονίσεις καθαυτές (Drucker 2016, 238). Τα εργαλεία οπτικοποίησης επικαθορίζουν την ερευνητική διαδικασία, όπως και οι βάσεις δεδομένων και όλα τα μοντέλα στα οποία βασίζεται η εξόρυξη δεδομένων, αλλά και η επιλογή των δεδομένων καθαυτή. Ως προς τα εργαλεία οπτικοποίησης, αυτό οφείλεται στο γεγονός ότι προέρχονται ως επί το πλείστον από τις θετικές επιστήμες, η στόχευση και η συλλογιστική των οποίων διαφέρει πολύ από εκείνες των ανθρωπιστικών επιστημών, και αναπαριστούν μεμονωμένα στοιχεία των κειμένων (Drucker 2020· Sinclair και Rockwell 2016), παρουσιάζοντας ως αντικειμενική μια ενδεχομένως μερική ερμηνευτική χειρονομία. Ως εκ τούτου, η χρήση τους στις ψηφιακές ανθρωπιστικές επιστήμες πρέπει να λαμβάνει σοβαρά υπόψη αυτές τις παραμέτρους, ενώ είναι πολύ σημαντική η επεξεργασία οπτικοποιήσεων που θα ανταποκρίνονται στην πολυπλοκότητα των αντικειμένων (Rockwell και Sinclair 2016a).

Τέλος, οι χάρτες είναι άλλο ένα εργαλείο που οι ψηφιακές ανθρωπιστικές επιστήμες το δανείστηκαν από τις θετικές επιστήμες, μαζί με τους πίνακες, τα γραφήματα, τα διαγράμματα και τις χρονογραμμές, όπως και μια από τις πιο προσφιλείς μεθόδους οπτικοποίησης, με στατικό ή δυναμικό τρόπο, γεωχωρικών δεδομένων (Geospatial Data). Στην εικόνα 7 βλέπουμε έναν χάρτη των δημοσιεύσεων του Βολταίρου, καταρτισμένο στο πλαίσιο του έργου «Mapping the Republic of Letters» του Πανεπιστημίου Στάνφορντ (τα στοιχεία των δημοσιεύσεων αντλούνται από την Εθνική Βιβλιοθήκη της Γαλλίας). Απεικονίζοντας τους διαφορετικούς τό-



Εικόνα 7. Χαρτογραφώντας τη δημοκρατία των γραμμάτων: οπτικοποίηση των δημοσιεύσεων του Βολταίρου.

πους έκδοσης του έργου του στο διάστημα 1712-1800, ο χάρτης αποτυπώνει τις μακρές περιόδους παραμονής του εκτός Γαλλίας, ηθελημένες και μη, τα προβλήματα που αντιμετώπισε με τη λογοκρισία, εξαιτίας της οποίας πολλά έργα του εκδίδονταν στην Ολλανδία ή την Ελβετία, τη διάδοση του έργου του όσο ζούσε και τα πρώτα χρόνια μετά τον θάνατό του (Εικόνα 7).²⁰

Για την αναπαράσταση και τη μελέτη των δεδομένων σε ένα ηλεκτρονικό σώμα κειμένων είναι συνήθως αναγκαία η αυτόματη, ημιαυτόματη ή η χειρωνακτική προσθήκη της πληροφορίας. Ο χαρακτηρισμός (markup) αφορά τη διάρθρωση, την κατηγοριοποίηση και τη μορφοποίηση των κειμένων, ενώ συχνά γίνεται με γλώσσα XML (Extensible Markup Language), αλλά και με το προσφιλέστερο στις ψηφιακές ανθρωπιστικές επιστήμες σχήμα TEI (Text Encoding Initiative).²¹ Μια σειρά από γλωσσικές επισημειώσεις είναι αυτοματοποιημένες, όπως η λημματοποίηση (lemmatization), η επισημείωση μερών του λόγου (Part-of-Speech ή PoS tagging), η συντακτική ανάλυση (parsing), η σημασιολογική και η κειμενική επισημείωση

20. Στην πειραματική αυτή απεικόνιση, τα δεδομένα συνέλεξε, ανέλυσε και οπτικοποίησε ο Glauco Mantegari, υπό την εποπτεία του καθηγητή Dan Edelstein.

21. Text Encoding Initiative, <http://www.tei-c.org>. Για μια εισαγωγή, βλ. Text Encoding Initiative by Example Project, <http://teibyexample.org>. Τικτοπούλου (2015) και Burnard (2014).

(semantic και discourse annotation, Γούτσος και Φραγκάκη 2015, 34-36), αλλά και η αναγνώριση ονοματικών οντοτήτων (named entities recognition) – χωρίς αυτό να σημαίνει ότι το αποτέλεσμα τους δεν χρήζει ελέγχου και επιμέλειας, με δεδομένες πάντα τις διαφορές φυσικής και τυπικής γλώσσας και του αντίκτυπού τους. Στα ελληνικά, ορισμένα από τα εργαλεία αυτά είναι διαθέσιμα στην ελληνική πλατφόρμα της υποδομής γλωσσικών πόρων και εργαλείων CLARIN (Common Language Resources and Technology Infrastructure), που υποστηρίζει την ανοιχτότητα και τη συλλογικότητα στην έρευνα των ψηφιακών ανθρωπιστικών επιστημών από το 2017, στο πλαίσιο της εθνικής υποδομής Απολλωνίς.²² Περισσότερα εργαλεία διατίθενται από το Ινστιτούτο Επεξεργασίας του Λόγου Ερευνητικό Κέντρο Αθηνά.²³

Πολλοί τύποι γλωσσικής επισημείωσης διευκολύνουν την ανάλυση κειμένου. Χαρακτηρισμένο και επισημειωμένο είναι το «Συγκριτικό αφηγηματικό ηλεκτρονικό σώμα κειμένων του γερμανικού λογοτεχνικού μοντερνισμού» (KOLIMO) της ομάδας των Berenike Herrmann και Gerhard Lauer (2017), το οποίο αποτελεί ένα εν προόδω έργο και περιλαμβάνει περισσότερα από 40.000 ποικίλου είδους πεζογραφικά, αφηγηματικά έργα. Όπως σημειώνεται στον ιστότοπο του έργου, ενώ πρόκειται για «σώμα λογοτεχνικών κειμένων», περιλαμβάνει πολλά μη λογοτεχνικά κείμενα· και ενώ αφορά τον μοντερνισμό, σε μια περίοδο που ορίζεται κατά προσέγγιση από το 1880 ως το 1930, περιλαμβάνει επίσης πολλά παλαιότερα κείμενα. Οι επιλογές αυτές σχετίζονται με το θεωρητικό και μεθοδολογικό μοντέλο της έρευνας: θεωρήθηκε απαραίτητη η μελέτη των μη λογοτεχνικών κειμένων, προκειμένου να οριστεί η λογοτεχνικότητα στη συγκεκριμένη περίοδο, όπως και των παλαιότερων έργων προκειμένου να προσδιοριστεί ο μοντερνισμός. Τέλος, η εστίαση στην αφηγηματική πεζογραφία αιτιολογείται με βάση τη στόχευση του έργου, που είναι η υφολογική ανάλυση, δεδομένων των διαφορετικών υφολογικών χαρακτηριστικών της ποίησης και του θεάτρου. Παρά ταύτα, στην παρούσα φάση του έργου, συνεχίζουν να υπάρχουν στη βάση ευάριθμα ποιητικά και δραματικά κείμενα, τα οποία στη συνέχεια αναμένεται να απαλειφθούν. Το συγκεκριμένο ηλεκτρονικό σώμα κειμένων, σε μια μορφή του, είναι ελεύθερα διαθέσιμο στους χρήστες και επιτρέπει αναζητήσεις.²⁴

22. CLARIN, <https://www.clarin.gr/el>. Το CLARIN:EL μαζί με το DARIAH-GR/ΔΥΑΣ συναποτελούν από το 2017 την Εθνική Υποδομή για τις Ψηφιακές Ανθρωπιστικές Επιστήμες και Τέχνες και για τη Γλωσσική Έρευνα και Καινοτομία «Απολλωνίς».

23. Institute for Language and Speech Processing Natural Language Processing Web Services (ILSP LP), <http://nlp.ilsp.gr/soaplab2-axis>.

24. Στο ακόλουθο βίντεο, ο Gerhard Lauer δίνει ένα καλό παράδειγμα συγκρότησης ενός εκτεταμένου, αξιόπιστου και ηλεκτρονικού σώματος κειμένων παρουσιάζοντας αναλυτικά πώς συγκροτήθηκε το ηλεκτρονικό σώμα λογοτεχνικών κειμένων στην περίπτωση του KOLIMO· βλ. <https://www.youtube.com/watch?v=Hq3eU5CSrNY>.

Το παράδειγμα του ΚΟΛΙΜΟ καθιστά απολύτως σαφές ότι ένα ψηφιακό έργο στις ανθρωπιστικές επιστήμες μπορεί να προκύψει μόνο μετά από μελετημένο σχεδιασμό και αυστηρή οργάνωση, βασισμένο σε εδραία γνώση του γνωστικού αντικειμένου, με κριτική επιλογή κειμένων και εργαλείων, και με οργάνωση των μεταδεδομένων, δηλαδή των δεδομένων που χαρακτηρίζουν τα δεδομένα επιτρέποντας τη βέλτιστη ανάκτησή τους.²⁵

Υπάρχουν όμως ορισμένοι τύποι ειδικής, ευέλικτης επισημείωσης που δεν γίνεται αυτοματοποιημένα, αλλά χειρωνακτικά ή ημιαυτόματα, όπως η επισημείωση της μεταφοράς που σχεδιάζεται στο συγκεκριμένο ηλεκτρονικό σώμα κειμένων. Τη δυνατότητα ευέλικτης επισημείωσης κειμένων δίνει το λογισμικό CATMA, που αναπτύχθηκε από το Πανεπιστήμιο του Αμβούργου, με επικεφαλής τον Jan Chris Meister (Meister κ.ά. 2019), και υποστηρίζει την ερμηνευτική, «αδογμάτιστη» πολιτισμική επισημείωση – δηλαδή την επισημείωση με βάση το εκάστοτε ερευνητικό ερώτημα. Το CATMA συνδυάζει την ανάλυση κειμένου και την επισημείωση στο πλαίσιο του ευέλικτου ερμηνευτικού κύκλου, δεν έχει κανένα περιορισμό σε επίπεδο γλώσσας και είναι ιδιαιτέρως φιλικό στον χρήστη.

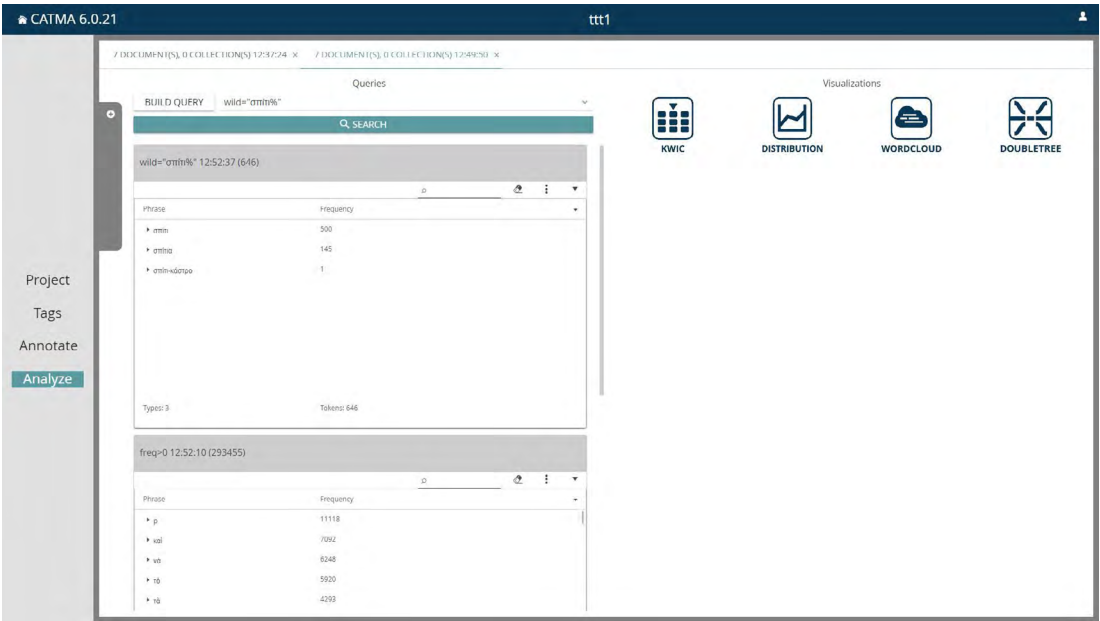
Θα δώσουμε ένα παράδειγμα ευέλικτης επισημείωσης με το CATMA, πάλι από τη λογοτεχνία, το οποίο αφορά ειδικότερα το έργο της Μέλπως Αξιώτη.²⁶ Συγκροτήσαμε ηλεκτρονικό σώμα κειμένων με τα λογοτεχνικά της κείμενα, πεζά και ποιητικά, χωρίς τα Χρονικά και τα δοκίμια που περιλαμβάνονται στο έργο *Μια καταγραφή στην περιοχή της λογοτεχνίας*, τα οποία θα ενταχθούν στο ηλεκτρονικό σώμα κειμένων σε δεύτερη φάση. Επίσης, δεν περιλαμβάνεται το γαλλόφωνο μυθιστόρημά της *République-Bastille*, το οποίο αποτελεί μείζονος σημασίας κείμενο, καθώς συνδέει την πρώτη με την τελευταία περίοδο του έργου της, αλλά θέτει το πρόβλημα ότι το πρωτότυπο είναι γαλλικό και η μετάφρασή του θα συνεξεταζόταν με πρωτότυπα κείμενα. Διατρέχοντας τον κατάλογο λέξεων, διαπιστώνει κανείς αμέσως ότι μία από τις λέξεις με ιδιαιτέρως μεγάλη συχνότητα εμφάνισης είναι η λέξη «σπίτι» (Εικόνα 8).

Οι συμφραστικοί πίνακες για τη λέξη «σπίτι» και στη συνέχεια για τη λέξη «σπίτια» (μη λημματοποιημένα δεδομένα) φαίνονται στις εικόνες 9 και 10.

Με δεδομένη τη συχνότητα εμφάνισης της λέξης, αλλά και τη γνώση μας για το έργο της Αξιώτη, διατυπώνουμε την υπόθεση ότι ο όρος «σπίτι», ο οποίος τιτλοφορεί άλλωστε το προτελευταίο έργο της, είναι έννοια σημαντική και πολύσημη, που το διατρέπει συνολικά· επομένως, παρουσιάζει ενδιαφέρον μια πιλοτική διερεύνηση της έννοιας με βάση αυτές τις εμφανίσεις, λόγου χάρη σε σχέση με

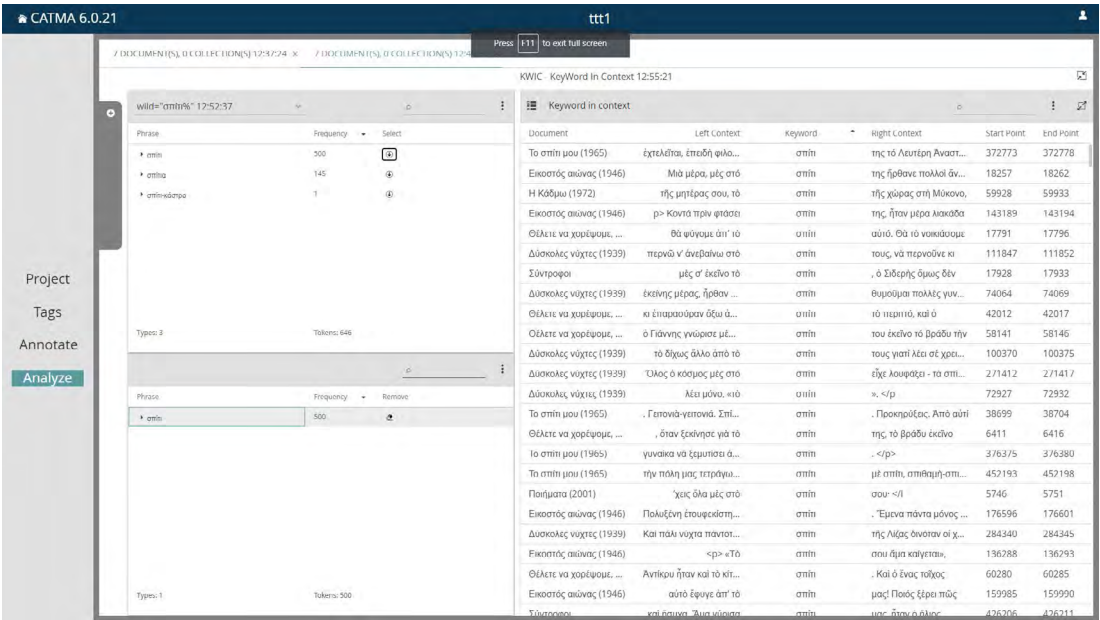
25. Για τα μεταδεδομένα και τη σημασία τους, βλ. Drucker (2021, 52-61).

26. Η ανάλυση και η επισημείωση των Απάντων της Αξιώτη είναι ένα έργο σε εξέλιξη που υλοποιούμε πιλοτικά μαζί με τον συνάδελφο Διονύση Γούτσο.

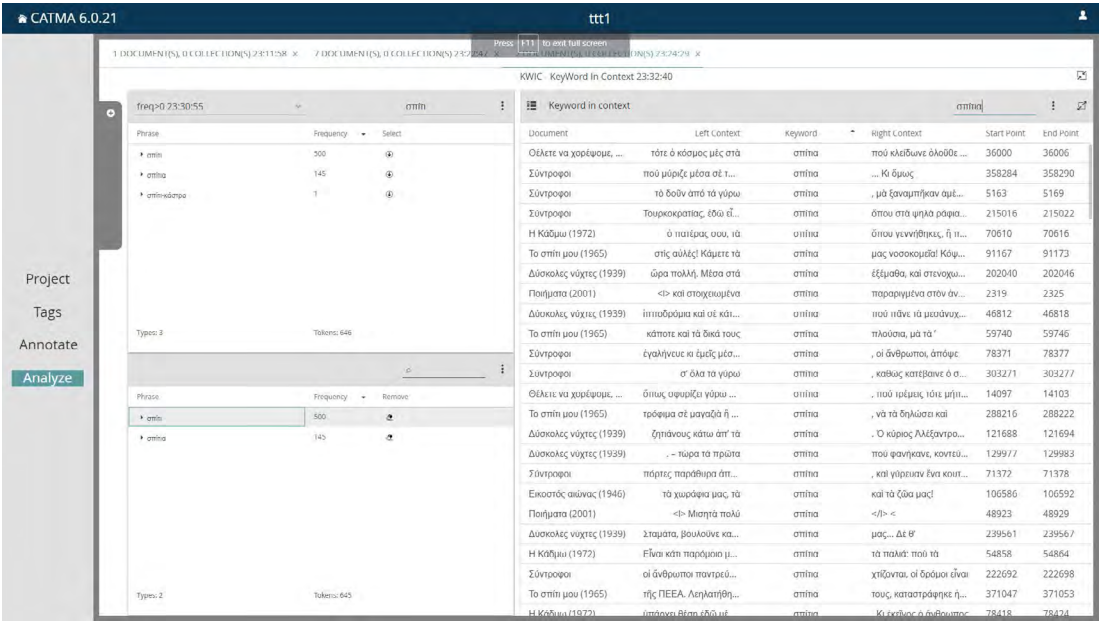


Εικόνα 8. Συχνότητα εμφάνισης της λέξης «σπίτι» στα λογοτεχνικά έργα της Αξιώτη (CATMA).

τις συνδηλωτικές της κάθε φορά σημασίες, τη μοναξιά και τον νόστο, τον φόβο και την αγωνία, την απέχθεια. Επισημειώσαμε τις εμφανίσεις με δύο δέσμες ετικετών, «θετικός» και «αρνητικός χώρος», και διαφορετικές ετικέτες για τις επιμέρους συνδηλώσεις. Η εικόνα 11 αφορά το πρώτο της έργο, τις Δύσκολες νύχτες, και προέρ-

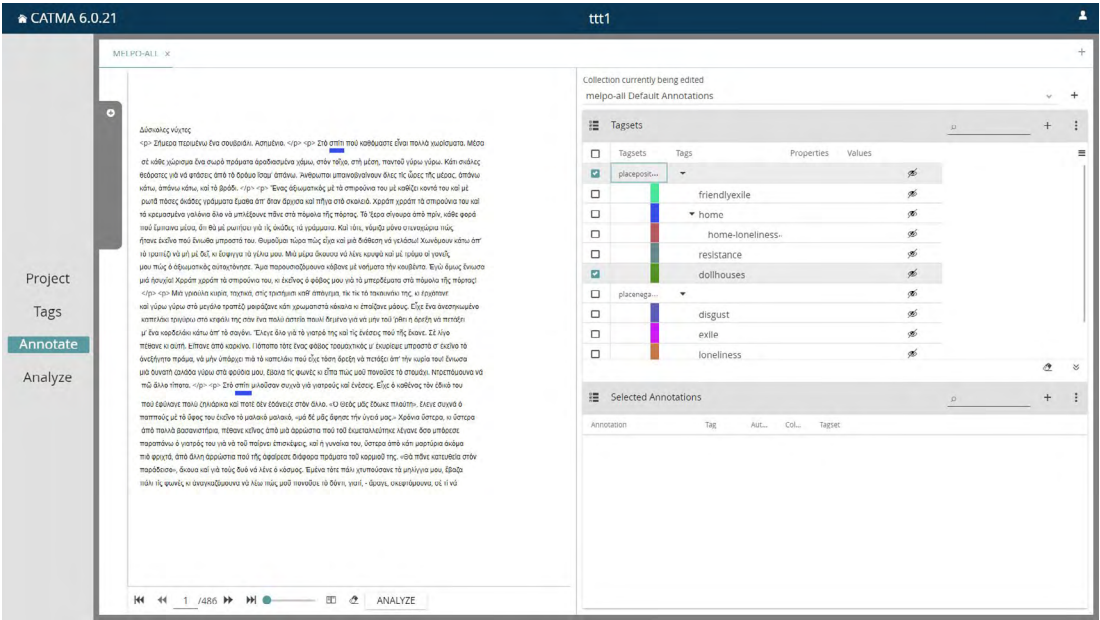


Εικόνα 9. Συμφραστικός πίνακας για τη λέξη «σπίτι» στα λογοτεχνικά έργα της Μέλπως Αξιώτη (CATMA).

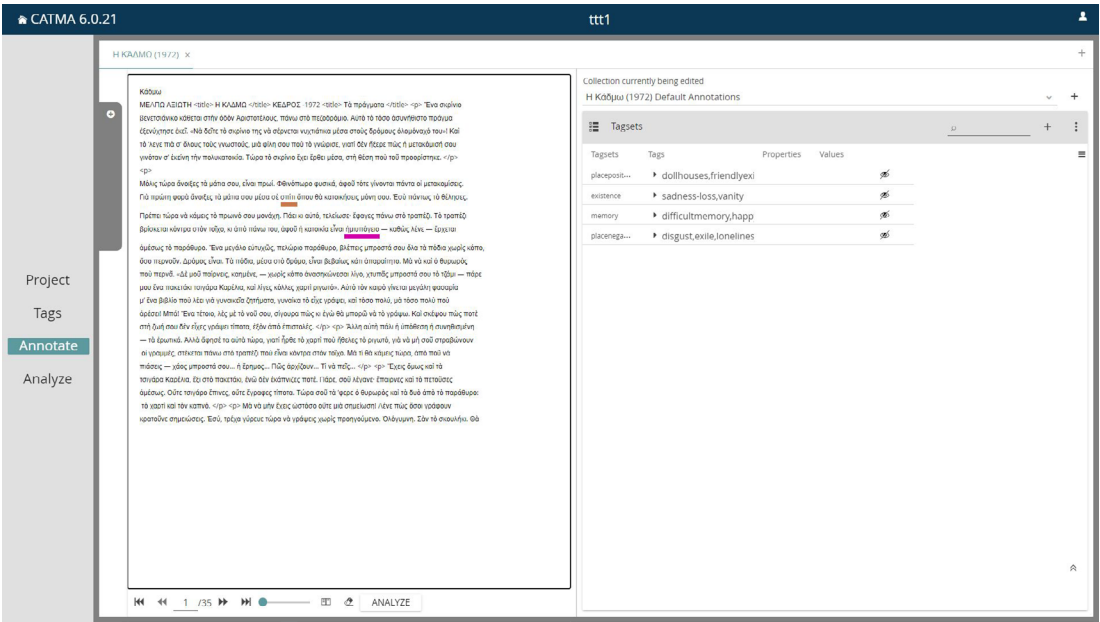


Εικόνα 10. Συμφραστικός πίνακας για τη λέξη «σπίτια» στα λογοτεχνικά έργα της Μέλπωσ Αξιώτη (CATMA).

χεται από το ηλεκτρονικό σώμα των λογοτεχνικών απάντων της σε ενιαίο κείμενο. Η επόμενη (Εικόνα 12) αφορά το τελευταίο έργο της Αξιώτη, την Κάδμω, και προέρχεται από ηλεκτρονικό σώμα με το κάθε έργο σε διαφορετικό αρχείο – οι διαφορετικοί τρόποι συγκρότησης των ηλεκτρονικών σωμάτων κειμένων αποτελούν



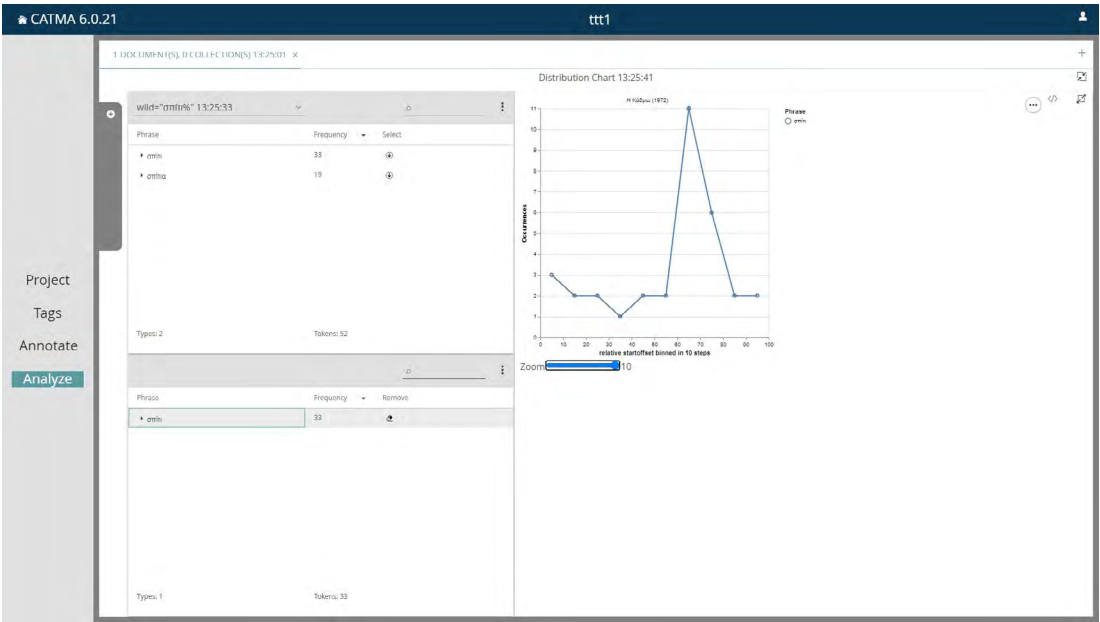
Εικόνα 11. Επισημείωση της λέξης «σπίτι» στις Δύσκολες νύχτες της Μέλπωσ Αξιώτη (CATMA).



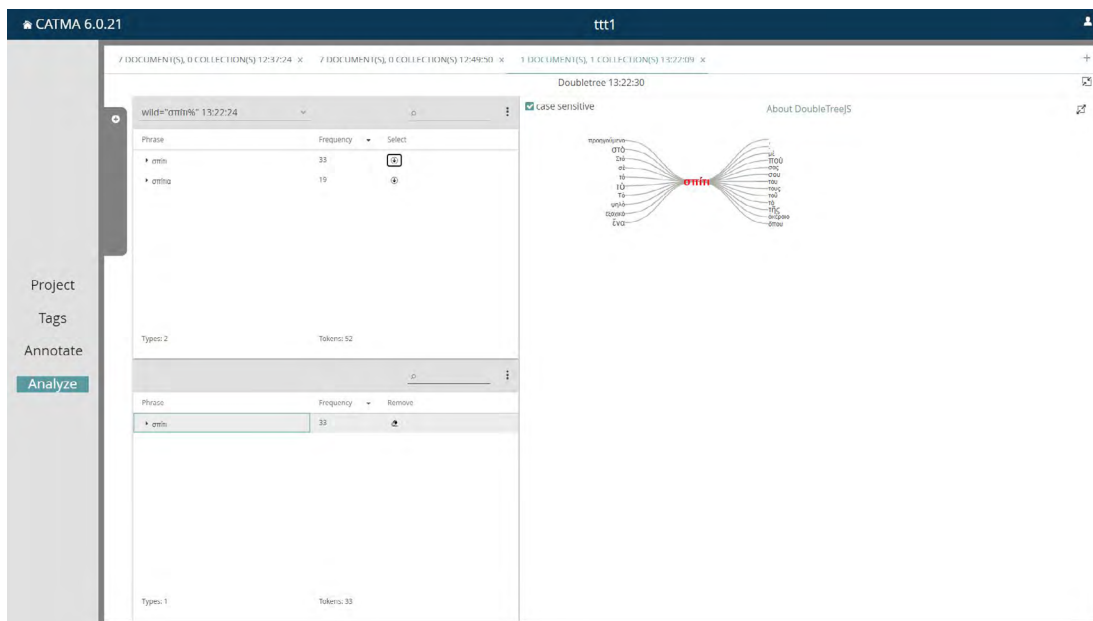
Εικόνα 12. Επισημείωση της λέξης «σπίτι» στην Κάδμω της Μέλπως Αξιώτη (CATMA).

δοκιμές σε σχέση με την αποτελεσματικότητα της επισημείωσης και της ανάλυσης, με δεδομένους περιορισμούς.

Τα επισημειωμένα δεδομένα, τα οποία μπορούν να εξαχθούν σε διάφορες μορφές για περαιτέρω χρήση, αποτελούν τη βάση για την ποιοτική έρευνα. Στις



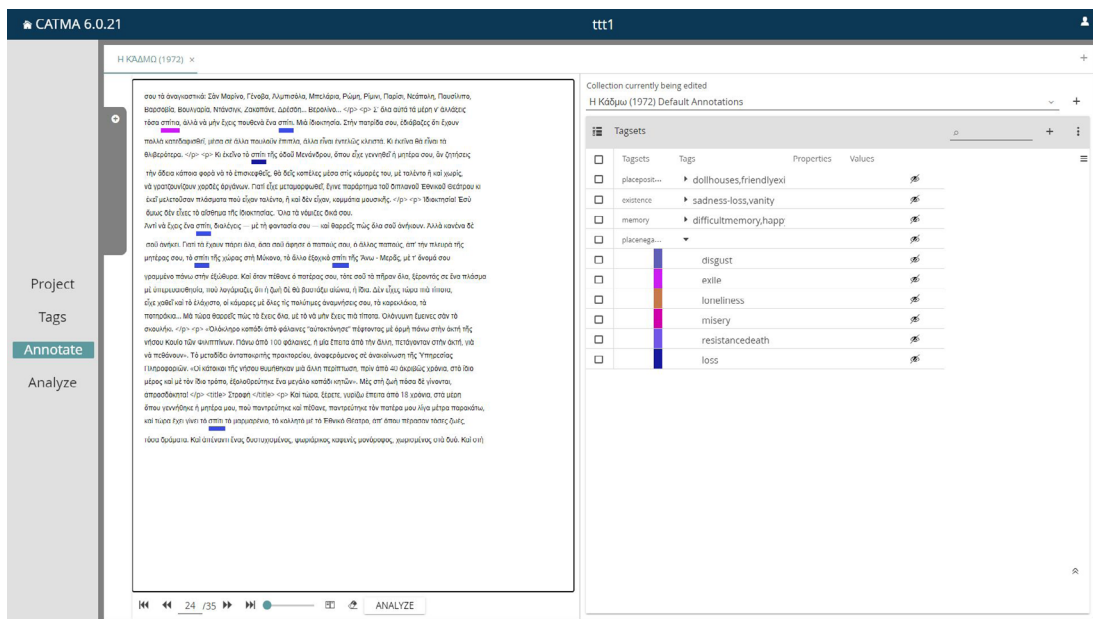
Εικόνα 13. Κατανομή της λέξης «σπίτι» στην Κάδμω της Μέλπως Αξιώτη (CATMA).



Εικόνα 14. Οπτικοποίηση άμεσου περικειμένου της λέξης «σπίτι» στην *Κάδμω* της Μέλπως Αξιώτη (CATMA).

εικόνες 13 και 14, οπτικοποιούνται οι εμφανίσεις της λέξης «σπίτι» στην *Κάδμω*, σε επίπεδο κατανομής και σε σχέση με το άμεσο κειμενικό περιβάλλον τους.

Τέλος, στην **εικόνα 15**, εμφανίζονται οι επισημειώσεις σε μια σελίδα της *Κάδ-μως*, όπου γίνεται φανερό η αντίθεση ανάμεσα στα «σπίτια» του κατατρεγμού και της εξορίας, και στο «σπίτι» ως εστία. Ωστόσο, το πιο σημαντικό συμπέρασμα που



Εικόνα 15. Επισημείωση των τύπων «σπίτι» και «σπίτια» στην Κάδμω της Μέλπως Αξιώτη (CATMA).

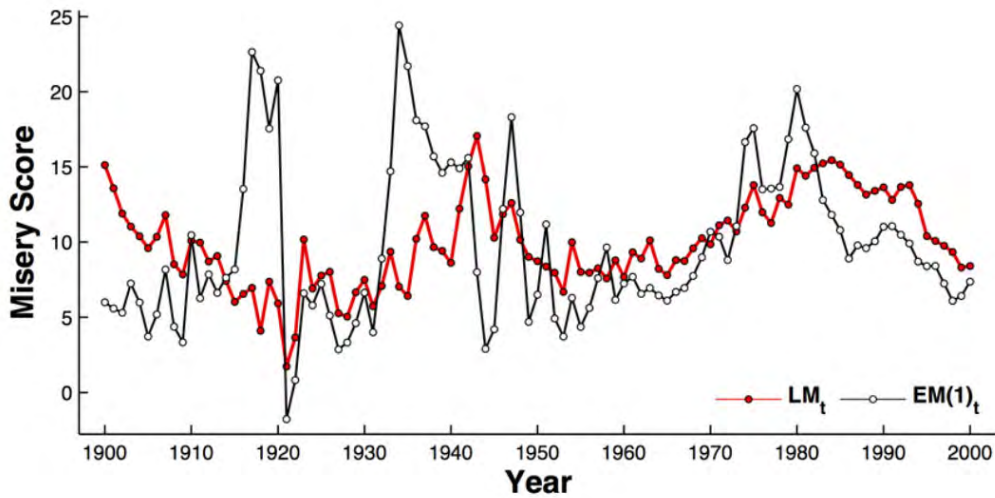
προέκυψε από την ανάλυση είναι ότι η συμβολική, αλλά και αφηγηματική, λειτουργία του σπιτιού υπερβαίνει κατά πολύ το αρχικό σχήμα μας, και χρειάζεται να επιστρέψουμε στο κείμενο και να προχωρήσουμε σε πιο πολύπλοκη επισημείωση, που θα διακρίνει τις πρωτοπρόσωπες από τις τριτοπρόσωπες αναφορές, λόγου χάρη, αλλά και θα προσθέτει νέες ετικέτες σε σχέση με την αφηγηματική λειτουργία του σπιτιού, όταν για παράδειγμα προσωποποιείται. Μείζον ζήτημα αποτελεί η αναφορά στην έννοια του σπιτιού χωρίς τη χρήση του συγκεκριμένου λεξικού τύπου, ενώ η όποια ερμηνεία ενδιαφέρει μέσα στο ευρύτερο πλαίσιο της πολιτικής στράτευσης και της εξορίας της Αξιώτη, αλλά και της θεματικής και αισθητικής εξέλιξης του έργου της.

Με δεδομένη τη χρησιμότητα και τη χρήση της στη λογοτεχνία, είναι ευνόητο ότι η κειμενομετρία έχει εφαρμογές και στη μελέτη της λογοτεχνικής μετάφρασης (Dimitroulia και Goutsos 2017), ενώ προσφάτως εφαρμόζεται επίσης στο πλαίσιο της ίδιας της μεταφραστικής πράξης και της αξιολόγησής της (Youdale 2020) – πολλά εργαλεία, λόγου χάρη, επιτρέπουν την επισημείωση και παράλληλων κειμένων.

Εξόρυξη δεδομένων από κείμενα: Μια διαφορετική προσέγγιση

Στην εξόρυξη δεδομένων από κείμενα αναλύονται μαζικά κειμενικά δεδομένα (big textual data) με εργαλεία, μεθόδους επεξεργασίας φυσικής γλώσσας και μοντέλα που κατασκευάζονται, ελέγχονται και καθοδηγούν την έρευνα. Προϋπόθεση για την εφαρμογή των αλγορίθμων και την επιτυχή ανακάλυψη νέας γνώσης είναι ο χαρακτηρισμός των ηλεκτρονικών σωμάτων κειμένων, αναλόγως με το μοντέλο ανάλυσης και το επιδιωκόμενο αποτέλεσμα, ενώ τα αποτελέσματα επίσης οπτικοποιούνται ποικιλοτρόπως. Επομένως, πολλά από όσα αναφέραμε για τον χαρακτηρισμό και την επισημείωση, αλλά και την οπτικοποίηση, αφορούν επίσης την εξόρυξη δεδομένων από κείμενα. Η κύρια διαφορά της κειμενομετρίας από την εξόρυξη κειμενικών δεδομένων αφορά την αδυναμία του ερευνητή να έχει διαρκή πρόσβαση στο σύνολο του υλικού και στα μοντέλα στα οποία βασίζεται με τους ερμηνευτικούς περιορισμούς που αυτά εισάγουν, σε μια κατεύθυνση εντελώς διαφορετική από την ευέλικτη ερμηνευτική.

Οι εφαρμογές εξόρυξης δεδομένων από κείμενα στις ψηφιακές ανθρωπιστικές επιστήμες είναι πολλές και ενδιαφέρουσες. Για παράδειγμα, οι Bentley, Acerbi, Ormerod και Lampos (2014) χρησιμοποίησαν μια σειρά από ηλεκτρονικά σώματα κειμένων, τα «Google n-grams», λεξικά και προηγμένα εργαλεία για να μελετήσουν



Εικόνα 16. Η καμπύλη του δείκτη κειμενικής δυσφορίας με κόκκινο χρώμα (LM) σε σχέση με τον δείκτη της οικονομικής κρίσης (EM). Πηγή: Bentley, Acerbi, Ormerod και Lampos 2014.

τη σχέση των οικονομικών κρίσεων στις ΗΠΑ με την έκφραση της δυσφορίας και των αρνητικών αισθημάτων (sentiment analysis) σε αγγλόφωνα κείμενα κατά τον 20ό αιώνα. Εντόπισαν την οικονομική κρίση και δυσπραγία στον 20ό αιώνα, με κύρια αναφορά στην κρίση του 1929 και στην ενεργειακή κρίση της δεκαετίας του 1970, και στο πλαίσιο του τριγωνισμού εξέτασαν επίσης γερμανόφωνα κείμενα σε σχέση με τον αντίστοιχο δείκτη της γερμανικής οικονομικής κρίσης. Κατέληξαν στο ενδιαφέρον συμπέρασμα ότι μια απόσταση περίπου δέκα ετών χωρίζει την οικονομική ύφεση και την κειμενική έκφραση της δυσθυμίας, απόσταση που εξηγείται πολυπαραγοντικά, όπως για παράδειγμα με την έκφραση παιδικών βιωμάτων, την ωρίμανση των συγγραφέων κ.λπ. (Εικόνα 16).

Ωστόσο, από το πολύ ενδιαφέρον και σύνθετο αυτό έργο, που βασίζεται στη μοντελοποίηση δεδομένων, προκύπτει η ερμηνευτική διάσταση που προϋπάρχει της ανάλυσης και την καθοδηγεί, όπως για παράδειγμα στο επίπεδο του ορισμού των συναισθημάτων, των οποίων η διαμόρφωση είναι πολύπλοκη και πολυπαραγοντική. Η πολυσημία και η συνδηλωτική πολυπλοκότητα της γλώσσας στη λογοτεχνία μπορεί πολύ δύσκολα να μοντελοποιηθεί, σε αντίθεση με τους οικονομικούς δείκτες. Το σώμα κειμένων του Google Books, πάλι, είναι αδιαμφισβήτητο για ποικίλους λόγους μη εξαντλητικό και συγκροτείται με βάση συγκεκριμένα κριτήρια, αλλά έστω κι έτσι η ανάλυση λειτουργεί ομοιογενοποιητικά στο πλαίσιο μιας εξ αποστάσεως ανάγνωσης για έργα που μπορεί να έχουν πολύ διαφορετικά χαρακτηριστικά, σημαίνοντα για τον ίδιο τον δείκτη της δυσθυμίας-δυσφορίας. Οι παρατηρήσεις αυτές δεν επιδιώκουν να ακυρώσουν τη σημασία ανάλογων εγχειρημάτων, που οδηγούν σε πολύ ενδιαφέροντα συμπεράσματα και ανοίγουν δρό-

μους στην παραδοσιακή έρευνα, αλλά να δείξουν την ιδιαιτερότητα της ανάλυσης πολιτισμικών κειμένων και τα όρια της μοντελοποίησης στο πλαίσιο της, τα οποία πρέπει να λαμβάνουμε υπόψη στις επιλογές μας (Drucker 2021, 19-23).

Ιδιαίτερο ενδιαφέρον παρουσιάζει η εξόρυξη κειμενικών δεδομένων για τον εντοπισμό κειμενικής ομοιότητας, δηλαδή διακειμενικότητας (Allen κ.ά. 2010· Forstall κ.ά. 2015· Ganascia, Glaudes και Lungo 2014), ενώ επίσης χρησιμοποιείται ευρέως στην απόδοση συγγραφικής πατρότητας και την υφομετρία (stylometry) ή υπολογιστική υφολογία (computational stylistics, Rybicki, Eder και Hoover 2016· Μικρός 2015). Οι Μαρμαρινού-Πολίτου, Μικρός και Δημητρούλια (2012) διερεύνησαν την πιθανότητα να αποδοθεί στον Αλέξανδρο Παπαδιαμάντη η πατρότητα ανώνυμων μεταφράσεων που είχαν δημοσιευτεί σε εφημερίδες. Για τις ανάγκες του έργου, συγκροτήθηκαν διαφορετικά ηλεκτρονικά σώματα κειμένων, των πρωτότυπων έργων του Παπαδιαμάντη, των βέβαιων μεταφράσεών του, των ανώνυμων μεταφράσεων και έργων των συγχρόνων του, ενώ το μοντέλο δοκιμάστηκε αρχικά σε άλλα ηλεκτρονικά σώματα κειμένων παρόμοιας λογικής. Τα αποτελέσματα της ανάλυσης συνέπεσαν σε μεγάλο βαθμό με τα ανάλογα συμπεράσματα της παραδοσιακής φιλολογικής έρευνας, κάτι που αποτέλεσε ένδειξη για την αξιοπιστία του μοντέλου.

Όλα τα παραδείγματα που προαναφέραμε είναι ενδεικτικά και έχουν στόχο να φωτίσουν τις μεθόδους και τα εργαλεία στην ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου. Οι ανάγκες των ερευνητών όμως ποικίλλουν, η επιλογή των κατάλληλων εργαλείων είναι εξίσου σημαντική με τη συγκρότηση του σώματος κειμένων, και έτσι είναι πολύ σημαντικό να μπορούν να αναζητήσουν τα εργαλεία που τους ταιριάζουν σε πύλες όπως λόγου χάρη η TAPOR, η οποία εξειδικεύεται σε εργαλεία ανάλυσης κειμένου.

Ασχέτως της επιλογής των εργαλείων, πάντως, ή και της αναδιαμόρφωσής τους ακόμη στο πλαίσιο του ερμηνευτικού κύκλου και της ευέλικτης ερμηνευτικής με βάση τα ζητούμενα της ερμηνείας, είναι δεδομένο ότι η χρήση της τεχνολογίας δεν οδηγεί από μόνη της και εξ ορισμού σε νέα ερευνητικά δεδομένα – εξού και η κριτική που ασκείται σε πολλά έργα των ψηφιακών ανθρωπιστικών επιστημών, ότι δηλαδή δεν έχουν στόχευση και αξία χρήσης. Το πιο σημαντικό είναι να προβληματιζόμαστε και να στοχαζόμαστε πάνω στους τρόπους με τους οποίους η τεχνολογία μπορεί να στηρίξει ερμηνευτικά την έρευνά μας, ειδικότερα σε σχέση με ερευνητικές υποθέσεις οι οποίες χωρίς την «επαυξητική» της ισχύ δεν θα μπορούσαν ποτέ να εξεταστούν. Άλλωστε, ως προς την ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου ισχύει ότι «όπως και σε πάρα πολλές πτυχές ψηφιακών έργων, η αξία [της] έγκειται στον διάλογο με τις παραδοσιακές μεθόδους» (Drucker 2021, 115). Ελπίζουμε, λοιπόν, ότι έγινε φανερό πως προκύπτει από τη συλλογική δουλειά και την ανοιχτότητα.

Συμπεράσματα

Η χρήση της τεχνολογίας στις ψηφιακές ανθρωπιστικές επιστήμες θέτει πολλά και σημαντικά ζητήματα, που αφορούν την επιστημολογική και την ηθική θεώρηση της τεχνολογίας και της ψηφιακότητας, των ψηφιακών έργων και των δεδομένων, τις μεθόδους και την ερμηνευτική τους κατεύθυνση (Gefen 2015· Berra 2014· Ramsay και Rockwell 2012· Drucker 2014· Kirschenbaum 2008). Δεν υπάρχει αμφιβολία ότι το εργαλείο λειτουργεί με πολλούς τρόπους διαμεσολαβητικά στην έρευνα εν γένει και αλλάζει την ερευνητική διαδικασία και την ερμηνευτική προσέγγιση (Berry 2012, 1), κάτι που ισχύει ειδικότερα και στην ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου. Είναι εξίσου βέβαιο, όμως, ότι ο ερευνητής, είτε χρησιμοποιεί απλά είτε σύνθετα εργαλεία, παραμένει στο επίκεντρο και ορίζει τις παραμέτρους του ερευνητικού του έργου.

Η χρήση εργαλείων της ηλεκτρονικά υποστηριζόμενης ανάλυσης κειμένου μπορεί να οδηγήσει την ελληνική κοινότητα των ψηφιακών ανθρωπιστικών επιστημών σε νέες ενδιαφέρουσες, ερμηνευτικές προσεγγίσεις, παρά την έλλειψη διαθέσιμων πόρων και τη σχετική έλλειψη εργαλείων. Ως εκ τούτου, προτεραιότητες της κοινότητας θα πρέπει να είναι η δημιουργία και ο διαμοιρασμός βιώσιμων πόρων και εργαλείων, όπως και η χρήση των εργαλείων αυτών με επίγνωση της ιδιαιτερότητας των ψηφιακών ανθρωπιστικών επιστημών. Γενικότερα, είναι πολύ σημαντική η διεύρυνση του διαλόγου μεταξύ διαφορετικών παραδόσεων, όπως και η ανάπτυξη της ελληνικής σχετικής βιβλιογραφίας. Τα δύο τελευταία ζητούμενα δεν αφορούν μόνον την ηλεκτρονικά υποστηριζόμενη ανάλυση κειμένου, αλλά το σύνολο των ψηφιακών ανθρωπιστικών επιστημών, και θα είναι καθοριστικά για την εξέλιξή τους.

Βιβλιογραφία

- Albaric, Michel. 2004. «Hugues de Saint-Cher et les concordances bibliques latines (xiii^e-xviii^e siècles)». Στο *Hugues de Saint-Cher* († 1263), *bibliote et théologien: Études réunies*, επιμέλεια Louis-Jacques Bataillon, Gilbert Dahan και Pierre-Marie Gy, 467-479. Τέρναουτ: Brepols.
- Allen, Timothy, Charles Cooney, Stéphane Douard, Russell Horton, Robert Morrissey, Mark Olsen, Glenn Roe, Robert Voyer. 2010. «Plundering Philosophers: Identifying Sources of the *Encyclopédie*». *Journal of the Association for History and Computing* 13 (Άνοιξη), τχ. 1. Επίσκεψη 14 Αυγούστου 2018. <http://hdl.handle.net/2027/spo.3310410.0013.107>.
- Argamon, Shlomo. 2009. «Words, Patterns and Documents: Experiments in Machine Learning and Text Analysis». *Digital Humanities Quarterly* 3, τχ. 2. Επίσκεψη 14 Ιουλίου 2021. <http://www.digitalhumanities.org/dhq/vol/3/2/000041/000041.html>.
- Barber, John. 2016. «Sound and Digital Humanities: Reflecting on a DHSI Course». *Digital Humanities Quarterly* 10, τχ. 1. Επίσκεψη 14 Ιουλίου 2021. <http://www.digitalhumanities.org/dhq/vol/10/1/000239/000239.html>.
- Bentley, Alexander E., Alberto Acerbi, Paul Ormerod και Vasileios Lamos. 2014. «Books Average Previous Decade of Economic Misery». *PLoS ONE* 9, τχ. 1. <https://doi.org/10.1371/journal.pone.0083147>.
- Berra, Aurélien. 26 Μαΐου 2014. «Connaître aujourd'hui: L'épistémologie problématique des humanités numériques». Εναρκτήρια ομιλία στο συνέδριο «DHnord2014. Humanités numériques: des outils, des méthodes, une culture». Maison européenne des sciences de l'homme et de la société, 26-28 Μαΐου 2014. Βίντεο, 77 λ. Επίσκεψη 14 Ιουλίου 2021. <http://publi.meshs.fr/ressources/connaître-aujourd'hui-lepistemologie-problematique-des-humanites-numeriques>.
- Berry, David M., επιμ. 2012. *Understanding Digital Humanities*. Λονδίνο: Palgrave-Macmillan.
- Biber, Douglas. 2010. «Corpus-Based and Corpus-driven Analyses of Language Variation and Use». *The Oxford Handbook of Linguistic Analysis*, επιμέλεια Bernd Heine και Heiko Narrog, 159-191. Οξφόρδη: Oxford University Press.
- Brier, Alan, και Bruno Hopp. 2011. «Computer Assisted Text Analysis in the Social Sciences». *Qual Quant* 45: 103-128. <https://doi.org/10.1007/s11135-010-9350-8>.
- Brunet, Étienne. 2011. *Ce qui compte: Écrits choisis*, τ. 2, *Méthodes statistiques*. Επιμέλεια Céline Poudat. Paris: Honoré Champion.
- Budzise-Weaver, Tina. 2016. «Developing a Qualitative Coding Analysis of Visual Artwork for Humanities Research». *Digital Humanities Quarterly* 10, τχ. 4. Επίσκεψη 14 Ιουλίου 2021. <http://digitalhumanities.org:8081/dhq/vol/10/4/000275/000275.html>.
- Burnard, Lou. 2014. «Introduction». Στο *What is the Text Encoding Initiative? How to Add Intelligent Markup to Digital Resources*. Μασσαλία: OpenEdition Press. <https://doi.org/10.4000/books.oep.679>.
- Burton, Dolores M. 1981α. «Automated Concordances and Word Indexes: The Early Sixties and the Early Centers». *Computers and the Humanities* 15 (Αύγουστος), τχ. 2: 83-100. <https://doi.org/10.1007/BF02404202>.
- . 1981β. «Automated Concordances and Word Indexes: The Process, the Programs, and the Products». *Computers and the Humanities* 15 (Νοέμβριος), τχ. 3: 139-154. <https://doi.org/10.1007/BF02404180>.

- Busa, Roberto. 1980. «The Annals of Humanities Computing: The Index Thomisticus». *Computers and the Humanities* 14 (Οκτώβριος), τχ. 2: 83-90. Επίσκεψη 14 Ιουλίου 2021. <https://www.alice.id.tue.nl/references/busa-1980.pdf>.
- . 1999. «Picture a Man... Busa Award Lecture, Debrecen, Hungary, 6 July 1998». *Literary and Linguistic Computing* 14 (Απρίλιος), τχ. 1: 5-9. <https://academic.oup.com/dsh/article-abstract/14/1/5/945098?redirectedFrom=fulltext>.
- Γούτσος, Διονύσης, και Γεωργία Φραγκάκη. 2014. «Πρόσφατη γλωσσική αλλαγή στα ελληνικά: Σχεδιασμός του Διαχρονικού Σώματος Ελληνικών Κειμένων του 20ού αιώνα». Στο *Selected Papers of the 11th International Conference on Greek Linguistics*, επιμέλεια G. Kotzoglou, K. Nikolou, E. Karantzola, K. Frantzi, I. Galantomos, M. Georgalidou, V. Kourti-Kazoullis, Ch. Papadopoulou, E. Vlachou, 318-329. Ρόδος: Πανεπιστήμιο Αιγαίου.
- . 2015. *Εισαγωγή στη γλωσσολογία σωμάτων κειμένων*. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. <http://hdl.handle.net/11419/1932>.
- Casanova, Pascale. 1999. *La république mondiale des lettres*. Παρίσι: Seuil. Ελλ. έκδ.: *Η παγκόσμια πολιτεία των γραμμάτων*. Μετάφραση Έφρη Γιαννοπούλου. Αθήνα: Πατάκης, 2006.
- Castelfranchi, Yuri. 2017. «Computer-aided Text Analysis: An Open-aided Laboratory for Social Sciences». *Journal of Science Communication* 16, τχ. 2: 1-11. Επίσκεψη 14 Ιουλίου 2021. https://jcom.sissa.it/sites/default/files/documents/JCOM_1602_2017_C04_en.pdf.
- Clement, Tanya. 2013. «Text Analysis, Data Mining, and Visualizations in Literary Scholarship». Στο *MLA Commons: Literary Studies in the Digital Age. An Evolving Anthology*. Επιμέλεια Kenneth M. Price και Ray Siemens. Επίσκεψη 14 Ιουλίου 2021. <https://dlsanthology.mla.hcommons.org/text-analysis-data-mining-and-visualizations-in-literary-scholarship/>.
- . 2016α. «Towards a Rationale of Audio-Text». *Digital Humanities Quarterly* 10, τχ. 2. Επίσκεψη 14 Ιουλίου 2021. <http://www.digitalhumanities.org/dhq/vol/10/3/000254/000254.html>.
- . 2016β. «Where is Methodology in Digital Humanities?». Στο *Debates in Digital Humanities* 2016. Επιμέλεια Matthew K. Gold και Lauren F. Klein. Μινεσότα: University of Minnesota Press. Επίσκεψη 14 Ιουλίου 2021. <http://dhdebates.gc.cuny.edu/debates/text/65>.
- Condillac, Etienne Bonnot de. 1798. *Essai sur l'origine des connaissances humaines*. Παρίσι: Ch. Houel, imprimeur. Αγγλ. έκδ.: *Essay on the Origin of Human Knowledge*. Μετάφραση Hans Aarsleff. Κέμπριτζ, Ηνωμένο Βασίλειο: Cambridge University Press, 2001.
- Culpeper, Jonathan. 2002. «Computers, Language and Characterisation: An Analysis of Six Characters in *Romeo and Juliet*». Στο *Conversation in Life and in Literature: Papers from the ASLA Symposium, Association Suedoise de Linguistique Appliquee (ASLA)*, τ. 15, 11-30. Επιμέλεια Ulla Melander-Marttala, Carin Ostman και Merja Kytö. Ουψάλα: Universitetsstryckeriet. Επίσκεψη 14 Ιουλίου 2021. https://lexically.net/wordsmith/corpus_linguistics_links/Keywords-Culpeper.pdf.
- . 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 (Μάρτιος), τχ. 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.
- de Swaan, Abram. 2001. *Words of the World: The Global Language System*. Κέμπριτζ: Polity Press.
- Δημητρούλια, Τιτίκα. 2015. «Παραγωγή κειμένου». Στο Τιτίκα Δημητρούλια και Κατερίνα Τικτοπούλου, *Ψηφιακές λογοτεχνικές σπουδές*, 40-94. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. <http://hdl.handle.net/11419/5827>.

- Dimitroulia, Titika, και Dionysis Goutsos. 2017. «Introduction: Special Issue on Corpora and Literary Translation». *InTRAlinea*. Επίσκεψη 14 Ιουλίου 2021. <http://www.intraline.org/specials/article/2260>.
- Drucker, Johanna. 2011. «Humanities Approaches to Graphical Display». *Digital Humanities Quarterly* 5, τχ. 1. Επίσκεψη 14 Ιουλίου 2021. <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.
- . 2014. *Graphesis*. Κέμπριτζ, Μασαχουσέτη: Harvard University Press.
- . 2016. «Graphical Approaches to the Digital Humanities». Στο *A New Companion to Digital Humanities*, επιμέλεια Susan Schreibman, Ray Siemens και John Unsworth, 238-250. Οξφόρδη: Blackwell.
- . 2017. «Information, Visualization and/as Enunciation». *Journal of Documentation* 73, τχ. 5: 903-916. <https://doi.org/10.1108/JD-01-2017-0004>.
- . 2020. *Visualization and Interpretation Humanistic: Approaches to Display*. Κέμπριτζ, Μασαχουσέτη: MIT Press.
- . 2021. *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*. Λονδίνο και Νέα Υόρκη: Routledge.
- Drucker, Johanna, David Kim, Iman Salehian και Anthony Bushong. 2014. «Data Mining and Text Analysis». Στο *Introduction to Digital Humanities Course Book: Concepts, Methods, and Tutorials for Students and Instructors*, 43-45. Επίσκεψη 14 Ιουλίου 2021. [https://ia801202.us.archive.org/4/items/IntroductionToDigitalHumanities/Introduction to Digital Humanities. Concepts, Methods, and Tutorial for Students and Instructors.pdf](https://ia801202.us.archive.org/4/items/IntroductionToDigitalHumanities/Introduction%20to%20Digital%20Humanities.%20Concepts,%20Methods,%20and%20Tutorial%20for%20Students%20and%20Instructors.pdf).
- Forstall, Christopher, Neil Coffee, Thomas Buck, Katherine Roache και Sarah Jacobson. 2015. «Modeling the Scholars: Detecting Intertextuality through Enhanced Word-Level N-Gram Matching». *Literary and Linguistic Computing* 30 (Δεκέμβριος), τχ. 4: 503-515. <https://doi.org/10.1093/lc/fqu014>.
- Fraser, Michael. 1996. «A Hypertextual History of Humanities Computing». Επίσκεψη 14 Ιουλίου 2021. <http://users.ox.ac.uk/~ctitext2/history>.
- Ganascia, Jean-Gabriel, Pierre Glaudes και Andrea Del Lungo. 2014. «Automatic Detection of Reuses and Citations in Literary Texts». *Literary and Linguistic Computing* 29 (Σεπτέμβριος), τχ. 3: 412-421. Επίσκεψη 14 Ιουλίου 2021. <https://hal.archives-ouvertes.fr/hal-00977310/document>.
- Gavriliidou Maria, Maria Koutsombogera, Anastasios Patrikakos και Stelios Piperidis. 2012. *The Greek Language in the Digital Age / Η ελληνική γλώσσα στην ψηφιακή εποχή*. META-NET White Paper Series «Europe's Languages in the Digital Age». Χαϊδελβέργη: Springer. <http://www.meta-net.eu/whitepapers/volumes/e-book/greek.pdf>.
- Gefen, Alexandre. 2015. «Les enjeux épistémologiques des humanités numériques». *Socio* 4: 61-74. <https://doi.org/10.4000/socio.1296>.
- Gius, Evelyn, Jan Christoph Meister, Malte Meister, Marco Petris, Christian Bruck, Janina Jacke, Mareike Schumacher, Dominik Gerstorfer, Marie Flüh και Jan Horstmann. 2021. CATMA 6 (Version 6.1). Zenodo. <https://doi.org/10.5281/zenodo.1470118>.
- Graham, Shawn, Scott Weingart και Ian Milligan. 2012. «Getting Started with Topic Modeling and MALLET». *The Programming Historian*. Επίσκεψη 14 Ιουλίου 2021. <https://programminghistorian.org/en/lessons/topic-modeling-and-mallet>. <https://doi.org/10.46430/phen0017>.

- Herrmann, Berenike J., και Gerhard Lauer. 2017. «KOLIMO: A Corpus of Literary Modernism for Comparative Analysis». Επίσκεψη 14 Ιουλίου 2021. <https://kolimo.uni-goettingen.de/about>.
- Hockey, Susan M. 2004. «The History of Humanities Computing». Στο *A Companion to Digital Humanities*, επιμέλεια Susan Schreibman, Ray Siemens και John Unsworth, 3-19. Οξφόρδη: Blackwell.
- Hoover, David L. 2013. «Textual Analysis». Στο *MLA Commons: Literary Studies in the Digital Age. An Evolving Anthology*, επιμέλεια Kenneth M. Price και Ray Siemens. Επίσκεψη 14 Ιουλίου 2021. <https://dlsanthology.mla.hcommons.org/textual-analysis/>.
- Jockers, Matthew L., και Ted Underwood. 2015. «Text-Mining the Humanities». Στο *A New Companion to Digital Humanities*, επιμέλεια Susan Schreibman, Ray Siemens και John Unsworth, 291-306. Οξφόρδη: Blackwell.
- Kirschenbaum, Matthew. 2008. *Mechanisms: New Media and the Forensic Imagination*. Κέμπριτζ, Μασαχουσέτη: MIT Press.
- . 2010. «What is Digital Humanities and What's It Doing in English Departments». *ADE Bulletin* 150: 55-61. Επίσκεψη 14 Ιουλίου 2021. <https://www.uvic.ca/humanities/english/assets/docs/kirschenbaum.pdf>.
- Kucher, Kostiantyn, και Andreas Kerren. 2015. «Text Visualization Techniques: Taxonomy, Visual Survey, and Community Insights». Στο *Proceedings of the 8th IEEE Pacific Visualization Symposium (PacificVis)*, επιμέλεια Shixia Liu, Gerik Scheuermann και Shigeo Takahashi, 117-121. Νιου Τζέρσεϊ: Institute of Electrical and Electronics Engineers (IEEE). Επίσκεψη 14 Ιουλίου 2021. <https://ieeexplore.ieee.org/document/7156366>.
- Lauer, Gerhard. 2017. «Doing Literary Studies with Corpora: The Case of KOLIMO». *Digital Humanities Lunch* (31 Μαρτίου). Επίσκεψη 14 Ιουλίου 2021. <http://scriptores.pl/dhlunch/en/31-01-2017>.
- Lebart Ludovic, Bénédicte Pincemin και Céline Poudat. 2019. *Analyse des données textuelles*. Κεμπέκ: Presses de l'Université du Québec.
- Lejeune, Christophe. 2010. «Montrer, calculer, explorer, analyser: Ce que l'informatique fait (faire) à l'analyse qualitative», *Recherches Qualitatives* 9: 15-32.
- Lejeune, Christophe, και Aurélien Béné. 2012. «Lexicométrie pour l'analyse qualitative: Pourquoi et comment résoudre le paradoxe?». Στο *11e Journées internationales d'analyse statistique des données textuelles (JADT)*, *Lexicométrica*, επιμέλεια Anne Dister, Dominique Longrée, Gérald Purnelle, 591-602. Επίσκεψη 14 Ιουλίου 2021. <https://hal-utt.archives-ouvertes.fr/hal-02923821/document>.
- Liu, Alan. 25 Απριλίου 2015. *Map of Digital Humanities*. Επίσκεψη 14 Ιουλίου 2021. <https://prezi.com/hjkj8ztj-clv/map-of-digital-humanities>.
- . 18 Οκτωβρίου 2020. «Digital Humanities and New Media Studies in the UCSB English Department: Map of Digital Humanities and New Media Studies in the English Department, 1994 to present.» Επίσκεψη 30 Ιανουαρίου 2023. <https://alanyliu.org/digital-humanities-new-media-studies-in-the-ucsb-english-department/>.
- Manovich, Lev. 2015. «Data Science and Digital Art History». *International Journal for Digital Art History* 1 (Ιούνιος): 12-35. <https://doi.org/10.11588/dah.2015.1.21631>.
- Μαρμαρινού-Πολίτου, Ελένη, Γεώργιος Μικρός και Τιτίκα Δημητρούλια. 2012. «Εφαρμογή υφομετρικών τεχνικών στην αναγνώριση πατρότητας κειμένου: Πρωτότυπα έργα και μεταφράσεις του Παπαδιαμάντη». Στο *Πρακτικά του Γ' Διεθνούς Συνεδρίου για τον Αλ. Παπαδιαμάντη*.

- ντη, τ. 2, 363-394. Αθήνα: Δόμος. Επίσκεψη 14 Ιουλίου 2021. <https://www.papadiamantis.net/aleksandros-papadiamantis/meletes/praktika-synedrion>.
- McCallum, Andrew K. 2002. «MALLET: A Machine Learning for Language Toolkit». Επίσκεψη 14 Ιουλίου 2021. <http://mallet.cs.umass.edu>.
- McEnery, Tony, και Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Κέμπριτζ, Ηνωμένο Βασίλειο: Cambridge University Press.
- McGann, Jerome. 2004. «Marking Texts of Many Dimensions». Στο *A Companion to Digital Humanities*, επιμέλεια Susan Schreibman, Ray Siemens και John Unsworth, 198-217. Οξφόρδη: Blackwell. Επίσκεψη 14 Ιουλίου 2021. <http://digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-3-4>.
- Meister, Jan Chris. 2023. «From TACT to CATMA». Στο *On Making in the Digital Humanities: Essays on the Scholarship of Digital Humanities Development in Honour of John Bradley*, επιμέλεια Julianne Nyhan και Geoffrey Rockwell. Υπό έκδοση. Προδημοσίευση προσβάσιμη (Νοέμβριος 2022) στο http://jcmeister.de/downloads/texts/Meister_2020-TACT-to-CATMA.pdf.
- Meister, Jan Chris, Jan Horstmann, Marco Petris, Janina Jacke, Christian Bruck, Mareike Schumacher και Marie Flüh. 2019. CATMA: Computer Assisted Text Markup and Analysis. <https://catma.de>.
- Μικρός, Γεώργιος. 2015. *Υπολογιστική υφολογία*. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. <http://hdl.handle.net/11419/4860>.
- Νικήσιανης, Νίκος, Θωμάς Σιώμος, Γιάννης Σταυρακάκης και Τιτίκα Δημητρούλια. 2016. «Λαϊκισμός εναντίον αντιλαϊκισμού στον ελληνικό τύπο, 2014-2015». *Σύγχρονα θέματα* 132-133: 52-70.
- Nikisianis, Nikos, Thomas Siomos, Yannis Stavrakakis, Grigoris Markou, Titika Dimitroulia. 2018. «Populism vs. Anti-populism in the Greek Press, 2014-2015: The Essex School Meets Corpus Linguistics». Στο *Discourse, Culture and Organization: Inquiries into Relational Structures of Power*, επιμέλεια Tomas Marttila, 267-295. Μπέζινγκστοκ, Χαμόαιρ: Palgrave Macmillan.
- O'Halloran, Kay, L. 2015. «Multimodal Digital Humanities». Στο *International Handbook of Semiotics*, επιμέλεια Peter P. Trifonas, 383-409. Ντόρτρεχτ: Springer.
- Pincemin, Bénédicte. 2020. «La textométrie en question». Ειδικό αφιέρωμα, *Le Français Moderne / Revue de Linguistique Française* 88, τχ. 1: 1-14. PDF. Επίσκεψη 14 Ιουλίου 2021. <https://shs.hal.science/halshs-02902088/document>. Έντυπη έκδοση στο Véronique Magri, επιμ. 2020. «Linguistique et traitements quantitatifs». Ειδικό αφιέρωμα, *Le Français Moderne: Revue de Linguistique Française* 88, τχ. 1: 26-43.
- Pollach, Irene. 2012. «Taming Textual Data: The Contribution of Corpus Linguistics to Computer-Aided Text Analysis». *Organizational Research Methods* 15, τχ. 2 (Απρίλιος): 263-287. <https://doi.org/10.1177/1094428111417451>.
- Ramsay, Stephen, και Geoffrey Rockwell. 2012. «Developing Things: Notes toward an Epistemology of Building in the Digital Humanities». Στο *Debates in Digital Humanities* 2012, επιμέλεια Matthew K. Gold. Μινεσότα: University of Minnesota Press. Επίσκεψη 14 Ιουλίου 2021. <https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb-36bfd1e/section/c733786e-5787-454e-8f12-e1b7a85cac72>.
- Rastier, François. 2011. *La mesure et le grain: Sémantique de corpus*. Παρίσι: Honoré Champion.
- Rockwell, Geoffrey. 2003. «What is Text Analysis, Really?». *Literary and Linguistic Computing* 18, τχ. 2: 209-219. <https://doi.org/10.1093/lc/18.2.209>.

- Rockwell, Geoffrey, και Stéfan Sinclair. 2014. «Towards an Archaeology of Text Analysis Tools». Εισήγηση στο Digital Humanities Conference 2014, Λωζάνη. Επίσκεψη 14 Ιουλίου 2021. https://www.researchgate.net/publication/273449857_Towards_an_Archaeology_of_Text_Analysis_Tools.
- . 2016α. «DH in Practice – Visualising Text». Επίσκεψη 14 Ιουλίου 2021. Βίντεο, 5 λ. 48 δ. <https://www.youtube.com/watch?v=uamyLcWtECg>.
- . 2016β. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Κέμπριτζ, Μασαχουσέτη, και Λονδίνο: MIT Press.
- . 2018. «Too Much Information and the KWIC». *Fudan Journal of the Humanities and Social Sciences* 11 (Δεκέμβριος), τχ. 4: 443-452. <https://doi.org/10.1007/s40647-018-0230-2>.
- Rybicki, Jan, Maciej Eder και David L. Hoover. 2016. «Computational Stylistics and Text Analysis». Στο *Doing Digital Humanities: Practice, Training, Research*, επιμέλεια Constance Crompton, Richard Lane και Ray Siemens, 123-144. Λονδίνο και Νέα Υόρκη: Routledge.
- Schreibman Susan, Ray Siemens και John Unsworth. 2016. *A New Companion to Digital Humanities*. Οξφόρδη: Blackwell.
- Siemens, Ray, Claire Warwick, Richard Cunningham, Teresa Dobson, Alan Galey, Stan Ruecker, Susan Schreibman, INKE Team. 2009. «Codex Ultor: Toward a Conceptual and Theoretical Foundation for New Research on Books and Knowledge Environments». *Digital Studies: Le Champ Numerique* 1, τχ. 2. <https://www.digitalstudies.org/article/id/7194/>.
- Sinclair, Stéfan, και Rockwell Geoffrey. 2016. «Text Analysis and Visualization: Making Meaning Count». Στο *A New Companion to Digital Humanities*, επιμέλεια Susan Schreibman, Ray Siemens και John Unsworth, 274-290. Οξφόρδη: Blackwell.
- Sinclair, Stéfan, Stan Ruecker και Milena Radzikowska. 2013. «Information Visualization for Humanities Scholars». Στο *MLA Commons: Literary Studies in the Digital Age. An Evolving Anthology*, επιμέλεια Kenneth M. Price και Ray Siemens. Επίσκεψη 14 Ιουλίου 2021. <https://dlsanthology.mla.hcommons.org/information-visualization-for-humanities-scholars>.
- Tasman, Paul. 1957. «Literary Data Processing». *IBM Journal of Research and Development* 1 (Ιούλιος), τχ. 3: 249-256. <https://doi.org/10.1147/rd.13.0249>.
- Τικτοπούλου, Αικατερίνη. 2015. «Κωδικοποιήσεις κειμένου». Στο Ξανθίππη Δημητρούλια και Αικατερίνη Τικτοπούλου, *Ψηφιακές λογοτεχνικές σπουδές*, 150-188. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. <http://hdl.handle.net/11419/5827>.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Άμστερνταμ και Φιλαδέλφεια: John Benjamins.
- Φραντζή, Κατερίνα. 2012. *Εισαγωγή στην επεξεργασία σωμάτων κειμένων*. Αθήνα: Ίων.
- Underwood, Ted. 2015. «Seven Ways Humanists are Using the Computer to Understand Text». Επίσκεψη 14 Ιουλίου 2021. <https://tedunderwood.com/2015/06/04/seven-ways-humanists-are-using-computers-to-understand-text/>.
- Unsworth, John. 2001. «Knowledge Representation in Humanities Computing». Εναρκτήρια διάλεξη στο National Endowment for the Humanities (3 Απριλίου). Επίσκεψη 14 Ιουλίου 2021. <http://johnunsworth.name/KR/>.
- Wiedemann, Gregor. 2013. «Opening up to Big Data: Computer-Assisted Analysis of Textual Data in Social Sciences». *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 14 (Μάιος), τχ. 2. <https://doi.org/10.17169/fqs-14.2.1949>.
- . 2015. «Text Mining #1: Extending the Method Toolbox: Text Mining for Social Science

and Humanities Research». *Europeana Research*. Επίσκεψη 14 Ιουλίου 2021. <http://research.europeana.eu/blogpost/extending-the-method-toolbox-text-mining-for-social-science-and-humanities-research>.

———. *Text Mining for Qualitative Data Analysis in the Social Sciences: A Study on Democratic Discourse in Germany*. Βισμπάντεν: Springer VS.

Winter, Thomas Nelson. 1999. «Roberto Busa, S. J., and the Invention of the Machine-Generated Concordance». *Faculty Publications, Classics and Religious Studies Department*, τ. 70. Επίσκεψη 14 Ιουλίου 2021. <http://digitalcommons.unl.edu/classicsfacpub/70>.

Youdale, Roy. 2020. *Using Computers in the Translation of Literary Style: Challenges and Opportunities*. Λονδίνο και Νέα Υόρκη: Routledge.

Δικτυογραφία

Η τελευταία επίσκεψη σε ηλεκτρονικούς πόρους πραγματοποιήθηκε στις 31 Ιανουαρίου 2023.

Anthony, L. «AntConc». www.laurenceanthony.net/software/antconc.
CLARIN-ERIC και DARIAH-EU. *Digital Humanities Course Registry*. <https://dhcr.clarin-dariah.eu/>.
Corpus Thomisticum. <http://www.corpusthomisticum.org/it/index.age>.
Digital Humanities Quarterly. <http://www.digitalhumanities.org/dhq>.
Digital Scholarship in the Humanities. <https://academic.oup.com/dsh>.
Digital Studies / Le champ numérique. <https://www.digitalstudies.org>.
Εθνική Υποδομή Γλωσσικών Πόρων και Τεχνολογιών CLARIN:EL. <https://www.clarin.gr/el/node/2>.
Εθνική Υποδομή για τις Ψηφιακές Ανθρωπιστικές Επιστήμες και Τέχνες και για τη Γλωσσική Έρευνα και Καινοτομία «Απολλωνίς». <https://apollonis-infrastructure.gr/the-infrastructure/>.
European Association for Digital Humanities (EADH). <https://eadh.org>.
Github. <https://github.com>.
Google. «Books Ngram Viewer». <https://books.google.com/ngrams>.
Hyperbase. <http://hyperbase.unice.fr/hyperbase>.
Institute for Language and Speech Processing, Natural Language Processing. Εργαλεία Ινστιτούτου Επεξεργασίας του Λόγου. <http://nlp.ilsp.gr/soaplab2-axis>.
IraMuTeQ. <http://www.iramuteq.org>.
Jstor Analyzer. <https://www.jstor.org/analyze>.
Le Trameur. <http://www.tal.univ-paris3.fr/trameur>.
Lexico 5. <http://lexi-co.com>.
Mallet. <http://mallet.cs.umass.edu>.
Populismus (Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης). <https://goo.gl/WdyrGd>.
Sketch Engine. <https://www.sketchengine.eu>.
Tapor. <http://tapor.ca/home>.
Tesseract. <http://tesseract.caset.buffalo.edu>.
Text Encoding Initiative. <http://www.tei-c.org> και <http://teibyexample.org>.
Text Visualization Browser. <https://textvis.lnu.se>.
Voyant Tools. <https://voyant-tools.org>.
Wordsmith Tools. <https://lexically.net/wordsmith>.
Zotero. «Doing Digital Humanities: A DARIAH Bibliography». Βιβλιογραφία έργων ψηφιακών ανθρωπιστικών επιστημών DARIAH-EU. https://www.zotero.org/groups/113737/doing_digital_humanities_-_a_dariah_bibliography.