

*Μοντελοποίηση και αναπαράσταση  
πολιτισμικής πληροφορίας  
Έμφαση στην αναπαράσταση εθνογραφικών δεδομένων*



*Διδακτορική Διατριβή*

*Μανόλης Πεπονάκης*



# Μοντελοποίηση και αναπαράσταση πολιτισμικής πληροφορίας

Έμφαση στην αναπαράσταση εθνογραφικών δεδομένων

Διδακτορική Διατριβή

Μανόλης (Εμμανουήλ) Πεπονάκης 

Ιόνιο Πανεπιστήμιο  
Σχολή Επιστήμης της Πληροφορίας & Πληροφορικής  
Τμήμα Αρχειονομίας, Βιβλιοθηκονομίας και Μουσειολογίας

Κέρκυρα, 2025

#### Τριμελής συμβουλευτική επιτροπή

- Σαράντος Καπιδάκης (Καθηγητής, *Πανεπιστήμιο Δυτικής Αττικής*)
- Ειρήνη Τουνταςάκη (Καθηγήτρια, *Πάντειο Πανεπιστήμιο*)
- Martin Doerr (Διευθυντής Ερευνών, *Ινστιτούτο Έρευνας και Τεχνολογίας*)

#### Επταμελής εξεταστική επιτροπή

- Σαράντος Καπιδάκης (Καθηγητής, *Πανεπιστήμιο Δυτικής Αττικής*)
- Ειρήνη Τουνταςάκη (Καθηγήτρια, *Πάντειο Πανεπιστήμιο*)
- Martin Doerr (Διευθυντής Ερευνών, *Ινστιτούτο Έρευνας και Τεχνολογίας*)
- Γιώργος Παπαϊωάννου (Καθηγητής, *Ιόνιο Πανεπιστήμιο*)
- Μάριος Πούλος (Καθηγητής, *Ιόνιο Πανεπιστήμιο*)
- Μιχάλης Σφακάκης (Καθηγητής, *Ιόνιο Πανεπιστήμιο*)
- Σοφία Στάμου (Επίκουρη Καθηγήτρια, *Ιόνιο Πανεπιστήμιο*)

## Πίνακας περιεχομένων

Πίνακας περιεχομένων .....	5
Ευρετήρια: σχημάτων, πινάκων, γραφημάτων και κώδικα .....	7
Πρόλογος .....	9
Ευχαριστίες .....	12
Περίληψη .....	13
Summary .....	16
1 Εισαγωγή: η ανθρωπολογία στην ψηφιακή εποχή .....	19
2 Αντικείμενο, στόχοι, δομή και συμβολή της διατριβής .....	25
2.1 Ερευνητικό ερώτημα, αντικείμενο και στόχοι της διατριβής .....	25
2.2 Δομή της διατριβής .....	26
2.3 Συμβολή της διατριβής .....	27
3 Μεθοδολογία .....	29
3.1 Εναλλαγή μεθόδων .....	31
3.2 Στρατηγική αναζήτησης και τεχνικές αναζήτησης πηγών .....	31
3.3 Έλεγχοι εγκυρότητας και επαληθευσσιμότητα .....	33
3.4 Περί ορολογίας .....	33
4 Υπολογιστική επεξεργασία εθνογραφικών δεδομένων: ιστορική αναδρομή, παρόν και προοπτικές .....	35
4.1 Υπολογιστική ανθρωπολογία και εθνογραφία: ορισμοί .....	36
4.2 Πρωτογενή δεδομένα και επεξεργασμένα δεδομένα .....	38
4.3 Επεξεργασία δεδομένων: στατιστική ανάλυση, μοτίβα (patterns) και επισημειώσεις (tagging) .....	39
4.3.1 Big data .....	44
4.4 Απευθείας συνεπαγωγή από τα δεδομένα: το παράδειγμα της Prolog .....	48
4.4.1 Συγγένεια και σχετικό λογισμικό .....	50
4.5 Αναζήτηση πληροφορίας .....	52
4.5.1 Βάσεις δεδομένων για Εθνογραφία και Ανθρωπολογία .....	53
4.6 Η σημασία της μοντελοποίησης των δεδομένων .....	56
4.6.1 Διαλειτουργικότητα .....	58
4.6.2 Επανάχρηση και διαμοιρασμός των δεδομένων .....	59
4.6.3 Προσπάθειες αναπαράστασης ανθρωπολογικής γνώσης .....	61
4.7 Συμπέρασμα κεφαλαίου .....	62
5 Αναπαράσταση και οργάνωση της γνώσης: φυσικές και τεχνητές γλώσσες .....	65
5.1 Η φυσική γλώσσα ως φορέας νοήματος .....	66

5.2	Φυσική γλώσσα, γραφή και ανθρωπολογική γνώση .....	69
5.2.1	Εκφραστικότητα: η γραφή ως αναπαράσταση της φυσικής γλώσσας και η πληροφορική ως γλώσσα αναπαράστασης πληροφορίας .....	70
5.3	Σημασιολογικός ιστός και διασυνδεδεμένα δεδομένα: ως φορείς νοήματος	72
5.3.1	Η RDF (και η RDFs) ως γλώσσες αναπαράστασης .....	76
5.3.2	Οντολογίες και OWL .....	80
5.4	Τα Συστήματα Οργάνωσης της Γνώσης (ΣΟΓ) ως τεχνητές ‘γλώσσες’ και ως φορείς νοήματος.....	84
5.4.1	Το SKOS ως μέσο αναπαράστασης των Συστημάτων Οργάνωσης της Γνώσης (ΣΟΓ).....	87
5.5	Ανάλυση αντιπροσωπευτικών συστημάτων οργάνωσης και αναπαράστασης γνώσης .....	90
5.5.1	Σύνολα δεδομένων και μεθοδολογία ανάλυσής τους .....	91
5.5.2	Συγκεντρωτικά αποτελέσματα .....	94
5.5.3	Συγκριτική ανάλυση επιλεγμένων ΣΟΓ και αντιπαράβολή με οντολογική προσέγγιση .....	100
5.5.4	Κριτική ανάλυση αποτελεσμάτων .....	108
5.6	Συμπέρασμα κεφαλαίου.....	110
6	Οντολογική αναπαράσταση Εθνογραφίας και Κοινωνικής Ανθρωπολογίας .....	115
6.1	Η οντολογία ως πολιτισμική οπτική .....	116
6.2	Μοντέλο αναπαράστασης με χρήση επιπέδων ολοκλήρωσης (layers of integration) .....	117
6.2.1	Διαφορετικές οντολογικές κατηγορίες ανά πολιτισμό .....	117
6.2.2	Πρώτο επίπεδο: αναπαριστώντας έναν πολιτισμό (εθνογραφία).....	122
6.2.3	Δεύτερο επίπεδο: γενικεύοντας σε ένα αφαιρετικό επίπεδο (κοινωνική ανθρωπολογία).....	124
6.3	Συμπέρασμα κεφαλαίου.....	128
7	Συζήτηση .....	133
7.1	Προαπαιτούμενα / προϋποθέσεις για την αναπαράσταση της ανθρωπολογικής γνώσης στο σημασιολογικό ιστό .....	135
7.2	Οφέλη για την ανθρωπολογία.....	136
7.3	Generative AI, Large Language Models, Deep Learning και μορφές Τεχνητής Νοημοσύνης.....	139
7.4	Μελλοντική έρευνα .....	146
8	Συμπεράσματα.....	147
	Βιβλιογραφικές αναφορές.....	155
	Γλωσσάρι.....	175

## Ευρετήρια: σχημάτων, πινάκων, γραφημάτων και κώδικα

### Ευρετήριο σχημάτων

Σχήμα 1: λέξεις και έννοιες βάσει της λογικής (της δεύτερης φάσης) του Wittgenstein	67
Σχήμα 2: <i>semantic web stack</i> .....	73
Σχήμα 3: παράδειγμα δόμησης δεδομένων σε RDF (γραφική απεικόνιση). ....	77
Σχήμα 4: πραγματικός κόσμος και εννοιολόγηση κλάσεων.....	81
Σχήμα 5: οι κατηγορίες των ΣΟΓ ως γλώσσες ευρετηρίασης σύμφωνα με τον Hjørland	85
Σχήμα 6: γραφική αναπαράσταση σε SKOS.....	89
Σχήμα 7: σχηματική απεικόνιση των βασικών κλάσεων του CIDOC-CRM και των μεταξύ τους σχέσεων .....	107
Σχήμα 8: παραδείγματα ανάγκης περιορισμού πληθικότητας στις σχέσεις γάμου .....	118
Σχήμα 9: γραφική απεικόνιση (μέσω Protégé) των σχέσεων νυμφεύεται/παντρεύεται και έχει σύζυγο.....	121
Σχήμα 10: φορμαλιστικός ορισμός εγκληματικής πράξης και αυτόματης ένταξης σε αυτή την κλάση μιας πράξης η οποία αποτελεί έγκλημα μόνο σε κάποιους πολιτισμούς.....	126

### Ευρετήριο γραφημάτων, πινάκων και εικόνων

Γράφημα 1: η αύξηση των επιστημονικών εργασιών για <i>big data</i> (πηγή Scopus) .....	45
Γράφημα 2: κατανομή, ανά επιστημονικό τομέα, του όρου <i>big data</i> (πηγή Scopus).....	46
Γράφημα 3: επιστημονικές δημοσιεύσεις -ανά έτος- για <i>linked data</i> , <i>semantic web</i> , <i>ontologies</i> και <i>Knowledge representation</i> (πηγή Scopus) .....	75
Γράφημα 4: συν-εμφανίσεις όρων στην επιστημονική βιβλιογραφία.....	76
Γράφημα 5: σύνθετες έννοιες ανά ΣΟΓ .....	99
Γράφημα 6: επιστημονικές δημοσιεύσεις για νευρωνικά δίκτυα, βαθιά μάθηση και μεγάλα γλωσσικά μοντέλα (πηγή δεδομένων Scopus) .....	139

Πίνακας 1: κατανομή των μερών του λόγου (ΜτΛ) για κάθε ΣΟΓ, βάσει της αυτόματης κατάταξης του NLP tagger του IEL. ....	97
--	----

Πίνακας 2: τα είκοσι πιο κοινά συντακτικά μοτίβα (διατεταγμένα κατά συχνότητα) .....	98
--	----

Εικόνα 1: το ChatGPT για τη σελίδα του Αριστοτέλη στο Google Scholar .....	142
Εικόνα 2: το ChatGPT για χώρες που ξεκινούν από ελληνικό Β.....	143
Εικόνα 3: το ChatGPT προτείνει τη χρήση ανύπαρκτου στοιχείου του χώρου ονομάτων ( <i>namespace</i> ) της RDF .....	144

## Ευρετήριο αποσπασμάτων κώδικα

Απόσπασμα κώδικα 1: παράδειγμα συνεπαγωγής σε <i>Prolog</i> .....	50
Απόσπασμα κώδικα 2: το Σχήμα 3 σε <i>RDF/XML</i> κώδικα .....	79
Απόσπασμα κώδικα 3: μια <i>OWL</i> ιδιότητα (σε <i>RDF/XML</i> ) με τα χαρακτηριστικά της .....	82
Απόσπασμα κώδικα 4: σύνδεση στιγμιότυπου με κλάση αντί με στιγμιότυπο .....	83
Απόσπασμα κώδικα 5: <i>RDF/XML</i> αναπαράσταση δύο εγγραφών σε <i>SKOS</i> .....	90
Απόσπασμα κώδικα 6: παράδειγμα επισημειώσεων <i>ΜτΛ</i> ( <i>POS tagging</i> ) από <i>ΙΕΛ</i> .....	93
Απόσπασμα κώδικα 7: η σχέση « <i>νυμφεύεται</i> » -ως <i>OWL</i> ιδιότητα- σε περιβάλλον μόνιμης μονογαμίας (χωρίς γάμους ομόφυλων ζευγαριών) .....	120
Απόσπασμα κώδικα 8: η σχέση « <i>νυμφεύεται</i> » -ως <i>OWL</i> ιδιότητα- σε περιβάλλον πολυανδρίας .....	121
Απόσπασμα κώδικα 9: ορισμός του εγκλήματος - σε <i>OWL</i> - ως Συμπεριφορά που επισύρει Ποινή .....	126

## Πρόλογος

Κατά τη διάρκεια της νεωτερικότητας, σε αρκετές περιπτώσεις, η επιστήμη πήρε το ρόλο της θρησκείας ως το μέσο που θα οδηγήσει την ανθρωπότητα στην απόλυτη αλήθεια. Η “πίστη” αυτή θεωρείτο, από τους “πιστούς” της, ως το ακριβώς αντίθετο της θρησκευτικής πίστης αφού το εφελτήριο ήταν οι επιστημονικές μέθοδοι, μέσω των οποίων επιτυγχάνεται η επαληθευσιμότητα των συμπερασμάτων. Η βασική αρχή αυτής της προσέγγισης θα μπορούσε να συνοψιστεί στη φράση: δεν ισχυρίζομαι ότι ισχύει κάτι επειδή έτσι πιστεύω, αλλά επειδή μπορώ να αποδείξω ότι ισχύει και αυτό οδηγεί σε μια αντικειμενική αλήθεια.

Όστόσο ο μεταμοντερνισμός έδειξε ότι δεν υπάρχει αντικειμενικότητα, παρά μόνο συλλογικές υποκειμενικότητες. Η επιστήμη είναι και αυτή μια οπτική και, όπως σημειώνει ο έγκριτος επιστημολόγος Thomas Kuhh, η έρευνα δεν είναι παρά μια επίπονη προσπάθεια να περιοριστεί η φύση στο εννοιολογικό πλαίσιο που παρέχει η σύγχρονη επαγγελματική εκπαίδευση (T. S. Kuhh, 1970, σελ. 5). Άρα η διένεξη, για το αν η επιστήμη θα προσφέρει το μέσο για την αντικειμενική αλήθεια, ανήκει στον προηγούμενο αιώνα. Και η φιλοσοφική θέση του «ignoramus et ignorabimus» απλά τίθεται σε νέα βάση, όπου η δυσκολία -ή και αδυναμία- ορισμού μιας έξω-ανθρώπινης (ή έστω πανανθρώπινης) αντικειμενικής πραγματικότητας δεν οδηγεί νομοτελειακά στην παραδοχή πως δεν υπάρχει η δυνατότητα να εκφραστούν “επιστημονικές αλήθειες” με καθολική ισχύ.

Η όραση είναι ένα ενδεικτικό παράδειγμα για τη σχέση μεταξύ “αντικειμενικής πραγματικότητας” και ανθρώπινης αντίληψης. Για τον άνθρωπο διαχρονικά η όραση αποτέλεσε μία από τις πιο σημαντικές αισθήσεις για την αλληλεπίδρασή του με το περιβάλλον. Στην εποχή μας αυτό έχει γιγαντωθεί και όλα φαντάζουν πιο εύληπτα με εικόνες. Πράγματα έξω από τις ανθρώπινες αισθήσεις, όπως οι ακτίνες X (οι γνωστές ακτινογραφίες) ή οι υπέρηχοι, μετατρέπονται σε εικόνες. Αλλά και πράγματα εντός του φάσματος των ανθρώπινων αισθήσεων, όπως η θερμοκρασία, μετατρέπονται επίσης σε εικόνες (θερμικές κάμερες). Αυτές, λοιπόν, οι αναπαραστάσεις αποτελούν μια αντικειμενική πραγματικότητα ή μια εξανθρωπισμένη πραγματικότητα; Αν οι άνθρωποι διέθεταν το επίπεδο νοημοσύνης που ήδη διαθέτουν, αλλά τις αισθήσεις ενός τυφλοπόντικα, τι νόημα θα είχαν τα προηγούμενα μοντέλα και πόσο θα τους επέτρεπαν να κατανοήσουν το φυσικό κόσμο και να αλληλεπιδράσουν με αυτόν;

Στην εξαιρετική TED παρουσίαση με τίτλο «Your brain hallucinates your conscious reality» ο διαπρεπής νευροεπιστήμονας Anil Seth περιγράφει τον τρόπο με τον οποίο ο ανθρώπινος εγκέφαλος “δημιουργεί” την πραγματικότητα για τον φυσικό κόσμο που τον περιβάλλει και σημειώνει: «Instead of perception depending largely on signals coming into the brain from the outside world, it depends as, if not more, on perceptual prediction flowing in the opposite direction. We don't just passively perceive the world; we actively generate it. The world we experience comes as much, if not more, from the inside out as from the outside in» (Seth, 2017). Συνεπώς ο ανθρώπινος νους δεν αντιλαμβάνεται

απευθείας την πραγματικότητα, αλλά την αντιλαμβάνεται (ή και την δημιουργεί σύμφωνα με τον Seth) μέσω των μοντέλων που έχει για αυτήν. Τα μοντέλα αυτά -παρότι καταφανώς πατούν πάνω σε μια βιολογική διάσταση- είναι, σε μεγάλο βαθμό, πολιτισμικές κατασκευές.

Προφανώς τα μοντέλα είναι απλουστευμένες αναπαραστάσεις της πραγματικότητας, αλλά οι άνθρωποι δεν μπορούν να συλλάβουν την πραγματικότητα ως έχει παρά μόνο μέσω των μοντέλων που έχουν για αυτήν. Η ανθρώπινη πραγματικότητα δεν είναι αυτό που αντικειμενικά υπάρχει αλλά αυτό που οι άνθρωποι μπορούν να κατανοήσουν. Και τα μοντέλα τους, όσο γίνονται πιο σύνθετα, τόσο περισσότερο τους επιτρέπουν να κατανοήσουν μια πιο σύνθετη πραγματικότητα. Άρα, τα μοντέλα, παρότι απλοποιούν την πραγματικότητα, δίνουν ταυτόχρονα έναν τρόπο να την αντιληφθούμε, αφού τα μοντέλα προσομοιώνουν την πραγματικότητα και μέσω αυτών ο άνθρωπος μπορεί να αντιληφθεί διαφορετικές “πραγματικότητες” οι οποίες προκύπτουν από το εκάστοτε μοντέλο μέσω του οποίου ο παρατηρητής προσλαμβάνει τον κόσμο. Ενδεικτικό παράδειγμα είναι η γραφή. Η γραφή δεν είναι εγγενές χαρακτηριστικό της φυσικής γλώσσας αλλά ένα μοντέλο -ή πολλά διαφορετικά μοντέλα αφού τα ιδεογράμματα πολύ μικρή σχέση έχουν με τα αλφάβητα- το οποίο επιτρέπει μια συγκεκριμένη αναπαράσταση της φυσικής γλώσσας. Όμως ο σύγχρονος εγγράμματος άνθρωπος δυσκολεύεται πάρα πολύ να αντιληφθεί τη γλώσσα χωρίς να τη συνδέσει με τη γραφή της και τους φορμαλισμούς που κομίζει το μοντέλο που έχουν αναπτύξει οι μελετητές καθεμίας γλώσσας<sup>1</sup>.

Και επειδή κάποια/ος θα μπορούσε να υποστηρίξει ότι το παράδειγμα με τη φυσική γλώσσα και τη γραφή της αναφέρεται, ούτως ή άλλως, σε κάτι εγγενώς ανθρώπινο, ας δούμε ένα παράδειγμα από τον αμιγώς φυσικό κόσμο και τα μοντέλα που έχουμε για αυτόν. Ας αναρωτηθούμε, λοιπόν, τι ώρα είναι αυτή τη στιγμή στο βόρειο ή στο νότιο πόλο αντίστοιχα; Είναι αυτονόητο ότι δεν είναι δυνατόν να υπολογιστεί κάτι τέτοιο, αφού το μοντέλο του διαμοιρασμού της γήινης σφαίρας σε 24 μεσημβρινούς επιτρέπει στο άτομο που βρίσκεται ακριβώς πάνω στο βόρειο (ή στο νότιο) πόλο να βρίσκεται ταυτόχρονα και στις 24 ζώνες ώρας της γης! Άρα, παρότι το μοντέλο του εικοσιτετράωρου περιγράφει ένα αμιγώς φυσικό φαινόμενο, πολύ μικρή πρακτική αξία θα είχε αν οι άνθρωποι κατοικούσαν κυρίως γύρω από τους πόλους. Μάλιστα η ιδέα της μέτρησης του χρόνου, στο πλαίσιο της κλασικής εικοσιτετράωρης βάσης, αποτέλεσε και διαφημιστικό τέχνασμα για περιοχές της βόρειας Νορβηγίας (Henley, 2019) που βρίσκονται εντός του αρκτικού κύκλου, άρα πολύ κοντά στο βόρειο πόλο.

Και αν συνδυάσουμε τη γραφή με το χρόνο, ως παράδειγμα μοντέλων πολιτισμικής κατασκευής, έχουμε ενδιαφέροντα ευρήματα για τον τρόπο με τον οποίο οι άνθρωποι εκφράζουν μια χρονική ακολουθία στο χώρο. Για τους πολιτισμούς που γράφουν από αριστερά προς τα δεξιά (π.χ. Ευρωπαίοι) η έννοια της χρονικής ακολουθίας πηγαιίνει από τα αριστερά προς τα δεξιά. Ενώ για εκείνους που γράφουν από τα δεξιά προς τα

---

<sup>1</sup> Περισσότερα για τη γραφή, ως μοντέλο αναπαράστασης της φυσικής γλώσσας, αναλύονται στο κεφάλαιο «5 Αναπαράσταση και οργάνωση της γνώσης: φυσικές και τεχνητές γλώσσες».

αριστερά (π.χ. Άραβες ή Εβραίους) η χρονική ακολουθία πηγαίνει από τα δεξιά προς τα αριστερά (Burns κ.ά., 2019). Ενώ οι Αβορίγινες, που δεν έχουν γραφή αλλά κινούνται σε ένα σύστημα συντεταγμένων που βασίζεται στα τέσσερα σημεία του ορίζοντα, όταν τους ζητήθηκε να βάλουν στη σειρά φωτογραφίες βάσει της ηλικίας του εικονιζόμενου προσώπου, τις τοποθετούσαν με φορά από την ανατολή προς τη δύση. Άρα, σε αντίθεση με τους έχοντες γραφή, το μοντέλο τους για την χωρική αναπαράσταση του χρόνου είναι ανεξάρτητο από την κατεύθυνση του δικού τους σώματος και βασίζεται σε έναν απόλυτο γεωχωρικό προσανατολισμό (Boroditsky & Gaby, 2010).

Προφανώς και τρέφει αυταπάτες εκείνη/ος που θεωρεί ότι θα δημιουργήσουμε ένα και μόνο μοντέλο το οποίο θα καλύπτει όλο το εύρος της πραγματικότητας που μπορεί να αντιληφθεί ο ανθρώπινος νους. Η μοντελοποίηση είναι μια αέναη διαλεκτική σχέση ανάμεσα στην έρευνα και την αποτύπωση της. Αυτή η αποτύπωση μπορεί να φαντάζει ως παγίωση της γνώσης, αλλά, επί της ουσίας, συνεχίζει να αποτελεί μια πρόσκαιρη οπτική της πραγματικότητας.

Ένα επιπλέον πρόβλημα της μοντελοποίησης είναι ότι τα μοντέλα που αναπτύσσονται από τις θετικές επιστήμες και εκείνα των κοινωνικών και ανθρωπιστικών επιστημών, πολλές φορές, δεν εδράζονται στις ίδιες αρχές. Είναι πάνω από μισός αιώνας από τότε που τονίστηκε το παραπάνω πρόβλημα στο έργο «The two cultures and the scientific revolution» (Snow, 1961) και η απουσία σύγκλισης διατηρείται σε υψηλό βαθμό. Αυτή η απουσία σύγκλισης των διαφορετικών οπτικών αποτέλεσε και ένα σημαντικό κίνητρό μου για την παρούσα διατριβή. Διότι ξεκίνησα ευελπιστώντας να συστηματοποιήσω τις δυνατότητες σύνδεσης που έβλεπα από τις φαινομενικά ετερόκλητες σπουδές μου (MSc στην επιστήμη της πληροφορίας, πτυχίο κοινωνικής ανθρωπολογίας και πτυχίο βιβλιοθηκονομίας και συστημάτων πληροφόρησης) και την επαγγελματική και ερευνητική μου εμπειρία στο Εθνικό Ίδρυμα Ερευνών και το Εθνικό Κέντρο Τεκμηρίωσης, επιχειρώντας την εναρμόνισή τους σε ένα λειτουργικό σύνολο.

## Ευχαριστίες

Στην εκπόνηση μιας διδακτορικής διατριβής, ειδικά αν αυτή εκπονείται από έναν μεσήλικα με χρόνια προϋπηρεσίας σε ερευνητικούς φορείς, έχει ήδη συμβάλει πλήθος ατόμων εντός και εκτός του τυπικού εκπαιδευτικού συστήματος. Από τις/ους “δασκάλους/ους” των πανεπιστημιακών εδράνων μέχρι τις/τους συναδέλφους στους φορείς της έρευνας. Οι απόψεις όλων αυτών των ανθρώπων έχουν επιδράσει, στον έναν ή τον άλλο βαθμό και με τον ένα ή τον άλλο τρόπο, τόσο στις γνώσεις του όσο και την οπτική του για την έρευνα. Προφανώς η απλή απαρίθμηση αυτών των ονομάτων μάλλον συντείνει στη συσκότιση της συμβολής τους παρά στην ανάδειξη αυτής. Έτσι θα περιοριστώ να ευχαριστήσω ονομαστικά μόνο στα άτομα που είχαν άμεση συμβολή στην εξέλιξη αυτής της διατριβής.

Τον καθηγητή του ΠΑΔΑ (καθηγητής στο Ιόνιο όταν ξεκινήσαμε αυτή την προσπάθεια) Σαράντο Καπιδάκη για τον κεντρικό ρόλο στο συντονισμό αυτής της έρευνας, την καθοδήγησή του και γενικά την κατανόησή του για όλων των ειδών τα ζητήματα που έπρεπε να αντιμετωπιστούν. Την καθηγήτρια του Παντείου Ειρήνη Τουντασάκη η οποία αποδέχτηκε την πρόκληση να συμμετάσχει σε ένα εγχείρημα που είχε πολύ περισσότερο πληροφορική παρά ανθρωπολογία. Η συμβολή της υπήρξε ιδιαίτερα κρίσιμη και η βοήθειά της πολυεπίπεδη. Τον Διευθυντή Ερευνών του ITE Martin Doerr -πλέον επίτιμο επικεφαλής του Κέντρου Πολιτισμικής Πληροφορικής- η βαθιά γνώση του οποίου πάνω σε θέματα αναπαράστασης γνώσης είχε καθοριστική σημασία τόσο για την εξέλιξη της παρούσας διατριβής όσο και για τη φιλοσοφική θεώρηση με την οποία πλέον αντιμετωπίζω την έρευνα για την αναπαράσταση της έρευνας.

Την Άννα Μάστορα η οποία, πέρα από την άμεση συνεργασία σε σημεία κοινού ενδιαφέροντος που αφορούσαν στην ανάλυση των συστημάτων οργάνωσης της γνώσης και τη χρήση τεχνολογιών επεξεργασίας φυσικής γλώσσας (NLP), είχε ενεργή συμμετοχή, σε όλα τα στάδια τόσο της έρευνας όσο και της συγγραφής του κειμένου της διατριβής, με πραγματικά καθοριστική συμβολή. Το Μιχάλη Σφακάκη για το διττό του ρόλο. Πρώτον ως τον άνθρωπο που στα πρώτα επαγγελματικά μου βήματα στο Εθνικό Ίδρυμα Ερευνών, όταν δεν γνώριζα ούτε τι είναι Πληροφορική αλλά ούτε και μοντελοποίηση(!), με εισήγαγε, ως προϊστάμενος μου, στο χώρο της ανάπτυξης λογισμικού και με δίδαξε πολλά. Και δεύτερον, χρόνια μετά, για τη βοήθειά του -ως πρόεδρος πλέον του τμήματος Αρχαιονομίας, Βιβλιοθηκονομίας και Μουσειολογίας- στο κλείσιμο του τυπικού κύκλου αυτής της διατριβής.

Τέλος, θα ήθελα να ευχαριστήσω το προσωπικό της βιβλιοθήκης του Παντείου Πανεπιστημίου -και ιδιαίτερα την προϊσταμένη Κωνσταντία Κακάλη- για τη βοήθεια στην απομακρυσμένη πρόσβαση σε πηγές ειδικά κατά την περίοδο της καραντίνας λόγω κορονοϊού.

## Περίληψη

Η σχέση της Κοινωνικής Ανθρωπολογίας με την Πληροφορική εκτείνεται σε δύο διαστάσεις. Από τη μία, αναφορικά με το πώς η Ανθρωπολογία μελετά την τεχνολογία ως πολιτισμική διαδικασία, και από την άλλη, σχετικά με το πώς η Ανθρωπολογία χρησιμοποιεί την τεχνολογία για να επιτελέσει τον σκοπό της. Για την πρώτη περίπτωση, δηλαδή για το πώς η Ανθρωπολογία μελετά την τεχνολογία ως πολιτισμική διαδικασία, εισάγονται έννοιες όπως Ψηφιακή Εθνογραφία (Digital Ethnography), Netnography και Εικονική Εθνογραφία (Virtual Ethnography). Αναφορικά με τη δεύτερη διάσταση της επίδρασης, δηλαδή εκείνη της χρήσης της Πληροφορικής ως εργαλειώδες μέσο ή ως μεθοδολογικό εργαλείο για την ανθρωπολογία, εισάγεται ο όρος Υπολογιστική Ανθρωπολογία (Computational Anthropology). Σημειώνεται ότι και οι δύο διαστάσεις αποτελούν υποκλάδους της Κοινωνικής Ανθρωπολογίας. Για την πρώτη προφανώς είναι αυτονόητο, όμως και η δεύτερη αποτελεί τομέα της Κοινωνικής Ανθρωπολογίας διότι κάθε Χ υπολογιστική επιστήμη είναι, εξ ορισμού, τμήμα αυτής της Χ-επιστήμης, παρότι για την ανάπτυξή της είναι απαραίτητη προϋπόθεση η συνέργεια των ατόμων που την υπηρετούν με την Πληροφορική. Η παρούσα διατριβή ασχολείται με την δεύτερη αυτή διάσταση.

Το βασικό ερευνητικό ερώτημα θα μπορούσε να συμπυκνωθεί ως εξής: πώς η ανθρωπολογία αξιοποιεί την πληροφορική και την επιστήμη της πληροφορίας και, ειδικότερα, πώς οι τεχνολογίες του σημασιολογικού ιστού θα μπορούσαν να συμβάλλουν είτε στη βελτιστοποίηση αυτής της χρήσης είτε στην δημιουργία νέων χρήσεων. Μέσα σε αυτό το πλαίσιο τίθενται τρεις κύριοι στόχοι για την διατριβή. Πρώτος στόχος: να παρέχει μια αναλυτική επισκόπηση της σχέσης της πληροφορικής με την ανθρωπολογία, εστιάζοντας στο πώς η πληροφορική χρησιμοποιείται από την ανθρωπολογία ώστε η δεύτερη να επιτύχει τους σκοπούς της. Δεύτερος στόχος: αφού καταδείξει το θεμελιώδη ρόλο της φυσικής γλώσσας ως μέσο αναπαράστασης της ανθρωπολογικής έρευνας, να παρουσιάσει τεχνητές γλώσσες -όχι κειμενοκεντρικές αλλά εννοιολογικές- κατάλληλες για την αναπαράσταση των εθνογραφικών δεδομένων και της ανθρωπολογικής γνώσης. Τρίτος στόχος: να παρουσιάσει τις βασικές αρχές για την αναπαράσταση της γνώσης στο σημασιολογικό ιστό (γενικώς) και να εστιάσει στην αποτύπωση της ανθρωπολογικής γνώσης (ειδικώς) σκιαγραφώντας τις βασικές προϋποθέσεις. Επιχειρώντας να υλοποιήσει τους παραπάνω στόχους εφαρμόζει μια διεπιστημονική προσέγγιση, η οποία επιθυμεί να συγκεράσει τρεις διαφορετικές επιστημονικές περιοχές: την κοινωνική ανθρωπολογία, την πληροφορική και την επιστήμη της πληροφορίας. Την τελευταία κυρίως ως προς την συμβολή της στην οργάνωση και την αναπαράσταση της γνώσης. Όλα τα παραπάνω νοηματοδοτούνται μέσα στα πλαίσια της ανοικτής επιστήμης όπου τόσα τα δεδομένα που υπόκεινται σε επεξεργασία όσο και τα συμπεράσματα που προκύπτουν από αυτά μπορούν να κωδικοποιηθούν με τη χρήση ανοικτών προτύπων και να διαλειτουργήσουν με άλλα δεδομένα.

Με δεδομένο ότι η έρευνα που προηγήθηκε της συγγραφής του κειμένου της διατριβής ήταν διεπιστημονική, η πολυσυλλεκτικότητα των επιστημονικών τομέων συντέινει στην αύξηση των μεθόδων που χρησιμοποιήθηκαν στις διαφορετικές φάσεις της έρευνας, αλλά και στους δυνητικούς συνδυασμούς μεθόδων. Η διατριβή έχει ως σημείο εκκίνησης τη θεμελιωμένη θεωρία (Grounded Theory). Η θεμελιωμένη θεωρία είναι μια γενική μεθοδολογία η οποία επιτρέπει την ανάπτυξη θεωριών οι οποίες θεμελιώνονται σε δεδομένα τα οποία έχουν συλλεχθεί και αναλυθεί με συστηματικό τρόπο. Στην περίπτωση όμως της συγκεκριμένης διατριβής τα δεδομένα είναι μελέτες, κείμενα, λογισμικά, πληροφοριακά συστήματα, οντολογίες, συστήματα οργάνωσης της γνώσης και δομές δεδομένων. Για να ενταχθούν αυτά σε ένα σύνολο ακολουθήθηκε η μέθοδος της μετα-σύνθεσης (meta-synthesis). η παρούσα μελέτη επιχειρεί να εντάξει σε ένα λειτουργικό σύνολο τα αποτελέσματα από μια σειρά διαφορετικών -αλλά αλληλένδετων- μελετών, λογισμικών, δομών δεδομένων κτλ. Όμως υποστηρίζει μάλλον μια ερμηνευτική παρά μια συγκεντρωτική αντίληψη. Έτσι καταλήγει σε αυτό που ονομάζεται κριτική ερμηνευτική σύνθεση (critical interpretive synthesis).

Μέσα αυτό το μεθοδολογικό και ερμηνευτικό πλαίσιο, και για την αποτελεσματική κάλυψη του στόχου της, η μελέτη κινήθηκε σε δύο βασικές συνιστώσες: από τη μία εξέτασε ζητήματα που αφορούν στη χρήση της πληροφορικής ώστε να βοηθηθούν οι ανθρωπολόγοι στην επεξεργασία των δεδομένων τους και την εξαγωγή συμπερασμάτων από αυτά, ενώ από την άλλη εξέτασε τρόπους για την αποτύπωση αυτών των συμπερασμάτων με μηχαναγνώσιμο τρόπο.

Για τη μελέτη της πρώτης συνιστώσας (δηλαδή το πώς μπορούμε να καταλήξουμε σε συμπεράσματα) εξέτασε μεθόδους και εργαλεία που μπορούν να λειτουργήσουν επικουρικά για την εξαγωγή συμπερασμάτων: από βάσεις δεδομένων έως λογισμικά ποιοτικής ανάλυσης δεδομένων (CAQDAS), αλλά και συστήματα αυτόματης συνεπαγωγής, εστιάζοντας στην Prolog ως μέσο μελέτης της συγγένειας. Η ανάλυση έδειξε ότι υπάρχει μια συνεχιζόμενη μετάβαση από τη χρήση των υπολογιστών ως απλά εργαλεία μετρήσεων στη χρήση της πληροφορικής ως εργαλείο ανάπτυξης νέων μεθόδων για την ανάλυση πολλών διαφορετικών τύπων δεδομένων -και όχι κατ' ανάγκη αριθμητικών- ώστε να συνδράμει στην εξαγωγή τεκμηριωμένων συμπερασμάτων. Επίσης έγινε εμφανές ότι, παρά τις διαφωνίες μέρους της κοινότητας των ανθρωπολόγων, υπάρχει μια σαφής στροφή προς την αξιοποίηση των ψηφιακών δεδομένων για την διεξαγωγή της ανθρωπολογικής έρευνας.

Για τη μελέτη της δεύτερης συνιστώσας (δηλαδή για το πώς μπορούν να αποτυπωθούν τα συμπεράσματα της ανάλυσης, ήτοι η αποκτηθείσα γνώση ανεξάρτητα από τον τρόπο που αυτή έχει αποκτηθεί) εξέτασε -εφαρμόζοντας, μεταξύ άλλων, τεχνικές επεξεργασίας φυσικής γλώσσας- συστήματα οργάνωσης και αναπαράστασης γνώσης με γνώμονα τις τεχνολογίες του σημασιολογικού ιστού και, αμφισβήτησε τη φυσική γλώσσα ως αποκλειστικό μέσο για την αποτύπωση των αποτελεσμάτων της ανθρωπολογικής έρευνας, διερευνώντας υπολογιστικές κωδικοποιήσεις και αναπαραστάσεις που είναι κατάλληλες για τα εθνογραφικά δεδομένα και την ανθρωπολογική γνώση.

Κατέδειξε ότι η φυσική γλώσσα δεν είναι η μόνη εναλλακτική για την αναπαράσταση των εθνογραφικών δεδομένων και της ανθρωπολογικής γνώσης. Το περιβάλλον των γράφων γνώσης και του σημασιολογικού ιστού αποτελεί, επίσης, πρόσφορο έδαφος για την αποτύπωσή τους. Η διατριβή κατέληξε προτείνοντας ένα μοντέλο που επιτρέπει την δημιουργία δύο επιπέδων ολοκλήρωσης (integration). Το πρώτο επίπεδο αφορά στην περιγραφή ενός πολιτισμού (με κίνδυνο υπεραπλούστευσης, ας το ονομάσουμε Εθνογραφία), ενώ το δεύτερο επίπεδο (πάλι με κίνδυνο υπεραπλούστευσης, ας το ονομάσουμε Κοινωνική Ανθρωπολογία) επιτρέπει τη γενίκευση και συμβάλει αποτελεσματικά στη συγκριτική ανθρωπολογία.

Η οντολογική αναπαράσταση που προτείνεται κάνει αυτό που (θα έπρεπε να) κάνει και η Ανθρωπολογία: διασαφηνίζει ρητά αυτό που υποβόσκει ως γνώση στα άτομα. Με άλλα λόγια η αναπαράσταση της γνώσης, μέσω οντολογιών, προσιδιάζει με την κοινωνική ανθρωπολογία στο σημείο του ότι προσπαθεί να καταγράψει, με σαφή και επιστημονικό τρόπο, τα όσα ήδη είναι γνωστά σε μια ομάδα ατόμων και να χρησιμοποιήσει αυτή την καταγραφή για να βγάλει συμπεράσματα ερμηνεύοντας τη γνώση που κατέγραψε.

## Summary

### **Title of the PhD Dissertation:**

#### ***Modeling and Representing Cultural Information: Emphasis on Ethnographic Data***

The relationship between social anthropology and computer science extends in two dimensions. On the one hand, it concerns how anthropology studies technology as a cultural process; on the other hand, it concerns how anthropology uses technology to fulfil its purpose. In the former case, i.e. how anthropology studies technology as a cultural process, concepts such as digital ethnography, netnography, and virtual ethnography are introduced. Regarding the second dimension, namely the use of computer science as a tool or methodological instrument for anthropology, the term 'computational anthropology' is introduced. It should be noted that both dimensions are subfields of social anthropology. The first is self-evident, but the second is also part of social anthropology because any computational science is, by definition, part of that science, even though progress in this area requires collaboration with computer scientists. The dissertation deals with the second dimension.

The main research question can be summarized as follows: how does anthropology leverage computer science and information science, and specifically, how could semantic web technologies contribute to optimizing this usage or creating new uses? Within this framework, three main goals have been set for the dissertation. The first is to provide an extended overview of the relationship between computer science and anthropology, focusing on how computer science is used by anthropology to achieve its objectives. The second goal, after demonstrating the fundamental role of natural language as a means of representing anthropological research, is to present artificial languages that are concept-centric rather than text-centric and suitable for representing ethnographic data and anthropological knowledge. The analysis includes the evaluation of specific knowledge organization systems and knowledge representation techniques in terms of their suitability for use in anthropology. The third goal is to present the basic principles for representing knowledge in the semantic web in general and focus on capturing anthropological knowledge in particular, outlining the basic prerequisites. In attempting to achieve these goals, the study takes an interdisciplinary approach, aiming to reconcile three different scientific areas: social anthropology, computer science, and information science. The latter primarily concerns its contribution to the organization and representation of knowledge. All of the above is contextualized within the framework of open science, where both the data being processed and the conclusions derived from them can be encoded using open standards and made interoperable with other data

As the research preceding the writing of this dissertation was interdisciplinary, the diversity of scientific fields contributed to an increase in the methods used in different phases of the research, as well as potential combinations of methods. The dissertation begins with Grounded Theory a general methodology that allows theories to be

developed based on systematically collected and analyzed data. In this dissertation, the data comprise studies, texts, software, information systems, ontologies, knowledge organization systems, and data structures. To integrate these into a whole, the meta-synthesis method was deployed. This study seeks to integrate the results from a series of different—yet interconnected—studies, software, data structures, etc., into a functional whole. However, it takes more of an interpretative rather than a comprehensive view. Thus, it concludes with what is called critical interpretive synthesis.

Within this methodological and interpretive framework, the study proceeded in two main components to effectively meet its goal. On the one hand, it examined issues related to using computer science to assist anthropologists in processing their data and drawing conclusions from it. On the other hand, it examined ways to capture these conclusions in a machine-readable manner.

For the study of the first component (i.e., how we can draw conclusions), methods and tools that can assist in this process were examined, ranging from databases to qualitative data analysis software (CAQDAS), as well as automatic inference systems, focusing on Prolog as a tool for studying kinship. The analysis showed that there is a continuous transition from using computers as simple tools for metrics to using computer science as a tool for developing new methods for analyzing many different types of data—not necessarily quantitative—to assist in drawing evidence-based conclusions. Despite disagreements within the anthropological community, it also became apparent that there is a clear shift towards utilizing digital data for conducting anthropological research.

To study the second component, which concerns how analytical conclusions and acquired knowledge can be captured regardless of their origin, knowledge organization systems and knowledge representation techniques were examined, with a focus on semantic web technologies. This examination challenges the exclusive use of natural language in recording anthropological research results by investigating computational representations suitable for ethnographic data and anthropological knowledge.

It was demonstrated that natural language is not the only alternative for representing ethnographic data and anthropological knowledge. The environment of knowledge graphs and the semantic web also provides fertile ground for their representation. The dissertation concluded by proposing a model that allows for the creation of two levels of integration. The first level concerns the description of a culture (with the risk of oversimplification, let's call this Ethnography), while the second level (again with the risk of oversimplification, let's call this Social Anthropology) allows for generalization and effectively contributes to comparative anthropology.

The proposed ontological representation does what anthropology (should) do: it explicitly clarifies what underlies the knowledge within individuals. In other words, representing knowledge through ontologies aligns with social anthropology in that it attempts to document what is already known within a group of individuals in a clear and scientific manner and use this documentation to draw conclusions by interpreting the knowledge it has captured.



# 1 Εισαγωγή: η ανθρωπολογία στην ψηφιακή εποχή

Στις αρχές του 21<sup>ου</sup> αιώνα οι Τεχνολογίες Πληροφοριών και Επικοινωνίας-ΤΠΕ (Information and Communication Technologies-ICT) έχουν διεισδύσει στην καθημερινότητα των πολιτών, ιδιαίτερα στον Δυτικό κόσμο, και έχουν καταφέρει να επηρεάσουν σε σημαντικό βαθμό την ανθρώπινη δραστηριότητα. Η πληθώρα των διαθέσιμων συσκευών (devices), σε συνδυασμό με την ευρυζωνική τεχνολογία, το υπολογιστικό νέφος (cloud computing) και την αναδυόμενη δημιουργική τεχνητή νοημοσύνη (Generative AI) έχουν αλλάξει, αλλά και εξακολουθούν να μεταβάλλουν, τόσο τον τρόπο με τον οποίο η πραγματικότητα γίνεται αντιληπτή όσο και τον τρόπο με τον οποίο οι άνθρωποι διαντιδρούν με αυτήν.

Ειδικότερα στον χώρο της Κοινωνικής Ανθρωπολογίας, η επανάσταση της πληροφορικής έχει ασκήσει σημαντικές επιρροές ως προς δύο βασικές διαστάσεις. Από τη μία, αναφορικά με το πώς η Ανθρωπολογία μελετά την τεχνολογία ως πολιτισμική διαδικασία, και από την άλλη, σχετικά με το πώς η Ανθρωπολογία χρησιμοποιεί την τεχνολογία για να επιτελέσει τον σκοπό της. Δανειζόμενοι τα λόγια του Hakken «θα πρέπει κανείς να διακρίνει τους ανθρωπολόγους ανάμεσα σε εκείνους που μελετούν την πληροφορική τεχνολογία ως πολιτισμική διαδικασία και εκείνους που ενδιαφέρονται για τους υπολογιστές επειδή τους αντιμετωπίζουν κυρίως ως εργαλείο μέσο ή για την πληροφορική επειδή αποτελεί μεθοδολογικό εργαλείο για την ανθρωπολογία» (Hakken, 1993, σελ. 112).

Για την πρώτη περίπτωση, δηλαδή το πώς η Ανθρωπολογία μελετά την τεχνολογία ως πολιτισμική διαδικασία, εισάγονται στο πεδίο έννοιες όπως Ψηφιακή Εθνογραφία (Digital Ethnography)<sup>2</sup> και Εικονική Εθνογραφία (Virtual Ethnography). Σημαντική συμβολή στην καθιέρωση του πεδίου της Ψηφιακής Εθνογραφίας έχει ο Michael Wesch, ο οποίος εστιάζει στη δύναμη των νέων μέσων. Μία από τις πιο επιδραστικές εργασίες του είναι μια παρουσίαση που έκανε στο Κογκρέσο των Ηνωμένων Πολιτειών για το

---

<sup>2</sup> Κάποιοι συγγραφείς χρησιμοποιούν τον όρο Netnography ως υποκατηγορία του Digital Ethnography. Η Netnography θεωρεί τους χρήστες του Διαδικτύου ως online κοινότητα ενώ η ψηφιακή εθνογραφία μπορεί απλά να χρησιμοποιεί το ψηφιακό περιβάλλον ως συμπλήρωμα μιας εκτός διαδικτύου έρευνας (Delli Paoli & D’Auria, 2025).

YouTube (Wesch, 2008). Αυτή η παρουσίαση ηθελημένα δεν ακολούθησε την τυπική δομή ενός ακαδημαϊκού κειμένου ακριβώς επειδή επιθυμούσε να δείξει τη δύναμη του μέσου (εν προκειμένω του βίντεο) στην επικοινωνήση της γνώσης. Μια πιο σύντομη εκδοχή αυτής της παρουσίασης, με τίτλο “*Web 2.0 ... The Machine is Us/ing Us*”, η οποία έγινε viral με περισσότερο από δώδεκα εκατομμύρια προβολές, αποτελεί μια ενδιαφέρουσα εισαγωγή στο πεδίο (Wesch, 2007). Από την άλλη το πεδίο της Εικονικής Εθνογραφίας ορίστηκε από την Hine (Hine, 2000), η οποία ασχολήθηκε με το κατά πόσο όσα συμβαίνουν στις διάφορες πλατφόρμες του εικονικού κόσμου γίνονται αντιληπτά από τα υποκείμενα με τον ίδιο τρόπο όπως αυτά που συμβαίνουν στον πραγματικό κόσμο. Σημαντική είναι η συμβολή σε αυτό το πεδίο και του Boellstorff, ο οποίος διεξήγαγε επιτόπια έρευνα στο διαδικτυακό παιχνίδι Second Life. Το avatar (δηλαδή ο ψηφιακός χαρακτήρας του χρήστη, το alter ego του στο παιχνίδι) “έζησε” ανάμεσα στους “κατοίκους” του παιχνιδιού και διεξήγαγε την παραδοσιακή για την ανθρωπολογική έρευνα μεθοδολογία της συμμετοχικής παρατήρησης. Μέσω αυτής κατέληξε στο συμπέρασμα ότι οι εικονικοί κόσμοι (virtual worlds) μπορούν να αλλάξουν τις ιδέες για την ταυτότητα και την κοινωνία στον πραγματικό κόσμο (Boellstorff, 2015). Μια σύνοψη για την πληροφορική ως πολιτισμική διαδικασία στη σύγχρονη της εκδοχή, καταγράφεται στο «*Digital Methods for Ethnography: Analytical Concepts for Ethnographers Exploring Social Media Environments*» όπου ο συγγραφέας σχολιάζει ότι οι ψηφιακές μέθοδοι προσκαλούν τους ερευνητές να ακολουθήσουν το μέσο, με τρόπο ώστε να αναλογιστούν το διαδίκτυο όχι τόσο σαν αντικείμενο μελέτης αλλά περισσότερο ως πηγή νέων μεθόδων και γλωσσών για την κατανόηση της σύγχρονης κοινωνίας (Caliandro, 2017, σελ. 3).

Τα παραπάνω καταδεικνύουν ότι η Τεχνολογία Πληροφοριών (Information Technology) μπορεί να μετασηματίσει το αντικείμενο της ανθρωπολογικής ανάλυσης, που είναι οι ανθρώπινες κοινωνίες. Ακόμη και οι μη εξειδικευμένοι σε θέματα πληροφορικής τείνουν να σκέφτονται με όρους πληροφορικής κάτω από συγκεκριμένες συνθήκες. Και δεν αναφερόμαστε στην ποσοτικοποίηση και την έμφαση σε προβολές και likes στα κοινωνικά δίκτυα, αλλά σε περιπτώσεις όπου προσπαθούν να χειραγωγήσουν τους αλγόριθμους που χρησιμοποιούν τα λογισμικά. Ένα παράδειγμα αυτής της συμπεριφοράς δίνεται από τους Devendorf και Goodman, όπως αναφέρεται από τον Seaver (Seaver, 2017, σελ. 4), όπου σχολιάζεται ότι οι χρήστες μιας online υπηρεσίας γνωριμιών προσπαθούν να “ξεγελάσουν” τον αλγόριθμο του συστήματος (στο βαθμό που τον κατανοούν), ώστε να επιτύχουν σε μεγαλύτερο βαθμό το επιθυμητό αποτέλεσμα, δηλαδή να γίνουν αποδέκτες της καταλληλότερης πρότασης για επιλογή συντρόφου. Από τα παραπάνω γίνεται σαφές πως έννοιες όπως οι αλγόριθμοι, δηλαδή ακολουθίες εντολών με αρχή και τέλος, που έχουν σκοπό την εκτέλεση συγκεκριμένων εργασιών ή την επίλυση κάποιου προβλήματος, είναι τόσο ενσωματωμένοι στην καθημερινότητα, ώστε δικαιολογημένα κάποιοι ανθρωπολόγοι ισχυρίζονται ότι «άνθρωποι και αλγόριθμοι προσαρμόζονται οι μεν στους δε» και ότι «υπάρχει μια συνεχώς αυξανόμενη αλληλεξάρτηση μεταξύ ανθρώπων και αλγορίθμων» (Arplin & Fischer, 2015, σελ. 5). Για αυτόν το λόγο, άλλωστε, η διαφάνεια της λειτουργίας των αλγορίθμων (algorithmic transparency) έχει καταστεί πολύ κρίσιμη: πρέπει να είναι γνωστό και κατανοητό το πώς λειτουργεί ο αλγόριθμος τόσο για να αποφευχθούν

μεροληπτικά αποτελέσματα (prevent bias) όσο και για να είναι σαφής η διαδικασία μέσω της οποίας ο αλγόριθμος κατέληξε στο X αποτέλεσμα.

Αναφορικά με τη δεύτερη διάσταση της επίδρασης, δηλαδή εκείνη της χρήσης της Πληροφορικής ως εργαλειακό μέσο ή ως μεθοδολογικό εργαλείο για την ανθρωπολογία, παρατηρούνται δύο βασικές μεταβολές. Αφενός, ο όγκος των δεδομένων που μπορούν να έχουν στη διάθεσή τους οι ανθρωπολόγοι εκτινάσσεται, όπως ιστορικά καταγραφής (logs) από servers και smart phones, αλλά και πληροφορίες που δημοσιοποιούν οι ίδιοι οι χρήστες (π.χ. βίντεο και φωτογραφίες). Αφετέρου, η διαχείριση των δεδομένων (είτε είναι γεννημένα ψηφιακά είτε όχι) μεταβάλλεται και εμπλουτίζεται εξαιτίας των νέων υπολογιστικών δυνατοτήτων. Και παρότι αρκετοί ανθρωπολόγοι, επικαλούμενοι θεωρητικές και μεθοδολογικές δικαιολογίες, αρνούνται την εξέταση αυτού του υλικού (Παπαηλία & Πετρίδης, 2015), η ανθρωπολογία -αν θέλει να καταστήσει τον σύγχρονο άνθρωπο αντικείμενο της έρευνάς της- τότε δεν μπορεί να αποφύγει την εμπλοκή της στη διαχείριση του ψηφιακού αποτυπώματος που αφήνει η δράση των ανθρώπων.

Ο τρόπος προσέγγισης αυτού του τεράστιου ψηφιακού αποτυπώματος, το οποίο ηθελήμενα ή αθέλητα αφήνουν οι χρήστες, μπορεί να εξετασθεί με τη χρήση δύο διαφορετικών (όχι κατ' ανάγκη αλληλοαποκλειόμενων) μεθοδολογικών προσεγγίσεων. Η πρώτη έχει να κάνει με την οπτική του ίδιου του χρήστη. Σε αυτή την περίπτωση δεν απαιτείται ιδιαίτερη εξειδίκευση στην πληροφορική από την πλευρά αυτού που κάνει την έρευνα. Έτσι ο ερευνητής μπορεί να παρακολουθεί και να καταγράφει με δικές του μεθόδους τα τεκταινόμενα στον ψηφιακό κόσμο και να προσπαθεί να τα ερμηνεύσει. Για παράδειγμα, αν κάποιος θέλει να αναλύσει μια κοινότητα στο Facebook το μόνο που χρειάζεται από ψηφιακές δεξιότητες είναι να γνωρίζει τη χρήση της εν λόγω πλατφόρμας. Σε αυτό το πλαίσιο εντάσσεται το έργο του Boellstorff που αναφέρθηκε προηγουμένως. Η δεύτερη προσέγγιση έχει να κάνει με τη μαζική διαχείριση υποσυνόλων αυτού του ψηφιακού αποτυπώματος χρησιμοποιώντας υπολογιστικές μεθόδους. Σε αυτή την περίπτωση η ανάλυση γίνεται απευθείας στο υλικό (logs κτλ.) που είναι αποθηκευμένο στους υπολογιστές. Αν όμως κάποιος θέλει να αναλύσει μέσω της δεύτερης μεθόδου την ίδια ομάδα του Facebook που αναφέρθηκε στην προηγούμενη περίπτωση, τότε οι γνώσεις του σε θέματα πληροφορικής θα πρέπει να είναι σημαντικά πιο εξειδικευμένες αφού η υπολογιστική επεξεργασία δεν θα γίνει μέσω απλής χρήσης του Facebook αλλά με λογισμικά τα οποία, κατά περίπτωση, υπακούουν σε πιο απαιτητικές μεθοδολογίες. Οι Geiger και Ribes ονομάζουν την ανάλυση του ψηφιακού αποτυπώματος για τους σκοπούς της εθνογραφίας "trace ethnography" (Geiger & Ribes, 2011). Τη μέθοδο αυτή ακολουθεί ο Turner σε μια πρόσφατη εργασία του όπου αναλύει δεδομένα από το GitHub (Turner, 2019). Ενώ η Barkhatona προτείνει μια μεθοδολογική προσέγγιση η οποία ενσωματώνει την ποσοτική ανάλυση σε εθνογραφικές μελέτες διαδικτυακών κοινοτήτων. Εξετάζοντας τα ίχνη των αλληλεπιδράσεων σε ψηφιακά περιβάλλοντα, καταδεικνύει τα δυνητικά οφέλη της υπολογιστικής ανάλυσης στην ενίσχυση της εθνογραφίας στο διαδίκτυο, εστιάζοντας στο πώς ποσοτικές μέθοδοι -οι οποίες συχνά θεωρούνται αντισυμβατικές στον τομέα της εθνογραφίας- συμπληρώνουν αποτελεσματικά την ποιοτική κατανόηση (Barkhatona, 2023).

Αυτός ο τεράστιος όγκος πληροφοριών συνέβαλε ώστε η σύγχρονη δυτική κοινωνία να χαρακτηρίζεται, πολύ συχνά, ως «κοινωνία της πληροφορίας». Η πληροφορία θεωρείται βασικός πυλώνας και διακινείται σε μεγάλο βαθμό μέσω υπολογιστών. Όμως ένας μεγάλος αριθμός ερευνητών (κάποιοι από αυτούς αναφέρονται και στο κείμενο της παρούσας διατριβής) αντιμετωπίζουν τους υπολογιστές, λίγο ως πολύ, ως αυτόνομες οντότητες οι οποίες λειτουργούν με έναν τρόπο διαφορετικό ή και ανταγωνιστικό ως προς τον ανθρώπινο. Αντίθετα, για το συγγραφέα της ανά χειράς διατριβής, οι υπολογιστές νοούνται ως ανθρώπινα κατασκευάσματα και ο τρόπος λειτουργίας τους προβάλλει τον τρόπο σκέψης των δημιουργών τους. Η ενσωμάτωση της χρήσης τους σε τόσους τομείς της σύγχρονης ζωής έχει άμεση συσχέτιση με τον τρόπο σκέψης που αναπτύχθηκε στη δύση τους τελευταίους αιώνες. Οι Fauconnier and Turner ξεκινούν το ιδιαίτερα επιδραστικό, από άποψη ετεροαναφορών, έργο τους “The Way We Think” με τη φράση «ζούμε στην εποχή του θριάμβου της φόρμας» (Fauconnier & Turner, 2003, σελ. 3). Και σύμφωνα με τους Totaro & Ninno, αυτή ακριβώς η ηγεμονία του φορμαλισμού οδηγεί στην τεχνολογία, για αυτό και υποστηρίζουν πως «ο πολιτισμός της μηχανοποίησης δεν προέρχεται σαν το ‘υπεροικοδόμημα’ που δημιουργήθηκε από την ύπαρξη των μηχανών. Μάλλον το αντίθετο είναι αληθές, με την έννοια ότι η ανάπτυξη μιας νοοτροπίας που εστιάζει σε διαδικασίες φορμαλισμού διευκόλυνε στο σχεδιασμό και τη διασπορά της μηχανοποίησης» (Totaro & Ninno, 2014, σελ. 33).

Συγκεκριμένα για την ανθρωπολογία, οι Fortun et al ισχυρίζονται ότι η διαμόρφωση της κοινωνιοτεχνικής υποδομής (sociotechnical infrastructure) του 21<sup>ου</sup> αιώνα ανανεώνει τρία βασικά ζητήματα: α) το τι είναι ανθρωπολογική θεωρία και πώς αυτή παράγεται, β) ποιες είναι οι μορφές που σμιλεύουν και μεταφέρουν αυτήν τη θεωρία και γ) την πιθανότητα, και την ανάγκη, για νέες μορφές συνεργασίας (Fortun κ.ά., 2017, σελ. 13). Όμως μέχρι σήμερα, ο μόνος τρόπος για να επικοινωνηθεί η γνώση στο πεδίο της κοινωνικής ανθρωπολογίας ήταν η φυσική γλώσσα. Τώρα, δίνεται η δυνατότητα να γίνει και μέσω άλλου τύπου δομών. Ο Lyon σχολιάζει για αυτές τις δομές πως «φορμαλιστικές αναπαραστάσεις των εθνογραφικών δεδομένων είναι ζωτικής σημασίας για τη βιωσιμότητα και την υγεία της ανθρωπολογίας ως κοινωνικής επιστήμης» καθώς ενισχύουν τη διαφάνεια (transparency) όσων υποστηρίζονται από τους μελετητές και επιτρέπουν την επαληθευσσιμότητά (verifiability) τους αυξάνοντας, με τον τρόπο αυτόν, και την αξιοπιστία (reliability) των ευρημάτων (Lyon, 2013, σελ. 46). Στην παρούσα έρευνα αναγνωρίζονται οι παραπάνω διαστάσεις του φορμαλισμού, αλλά λαμβάνεται σοβαρά υπόψη η άποψη του Fischer πως οι μέθοδοι αυτές σε καμία περίπτωση δεν είναι κατάλληλες για όλων των ειδών τα προβλήματα. Όμως, έχουν εφαρμογή σε ένα μεγάλο αριθμό προβλημάτων, σε όλο το φάσμα της ανθρωπολογίας καταλαμβάνοντας το ευρύ πεδίο που ορίζεται μεταξύ της οπτικής που προέρχεται από τις Ανθρωπιστικές επιστήμες έως και της οπτικής των Θετικών επιστημών (M. D. Fischer, 2006, σελ. 5). Ενώ, παράλληλα με την υιοθέτηση τεχνικών που προέρχονται από τον χώρο της πληροφορικής, είναι πιθανότερο να αποφευχθεί η κοινοτοπία (triviality) η οποία συχνά συνδέεται με τις φορμαλιστικές περιγραφές του κοινωνικού γίνεσθαι από τη μία, και τις άγονες γενικεύσεις από την άλλη (M. D. Fischer & Finkelstein, 1991, σελ. 119).

Παρότι η τεχνολογία της πληροφορίας δεν έχει ιδιαίτερη μεγάλη διείσδυση στις κοινωνικές επιστήμες, τα τελευταία χρόνια υπάρχει μια αυξητική τάση. Σύμφωνα με τη βάση επιστημονικών δημοσιεύσεων Web of Science το κλασικό επιστημονικό περιοδικό του χώρου "Social Science Computer Review" αύξησε τον βαθμό επίδρασής του (impact factor) από το 2007 έως και το 2024, από 0,405 σε 3 (3,9 ο πενταετής). Ενώ ο εκδοτικός οίκος Springer ξεκίνησε το 2018 την έκδοση του περιοδικού με τίτλο «Journal of Computational Social Science» και το 2020 η Institute of Electrical and Electronics Engineers (IEEE) το περιοδικό «Journal of Social Computing». Όμως και παραδοσιακά περιοδικά του χώρου της ανθρωπολογίας ασχολούνται με το θέμα, αν και με κάπως ευρύτερη προσέγγιση. Για παράδειγμα, στο καλοκαιρινό τεύχος του 2018 του Cultural Anthropology υπάρχει η συλλογή άρθρων με τίτλο "Algorithms and Automation" η οποία φιλοξενεί τέσσερα άρθρα για το θέμα (Ekbia & Nardi, 2018; Lowrie, 2018; Nafus, 2018; Seaver, 2018). Ενώ και περιοδικά με πιο γενική ύλη αφιερώνουν τεύχη σε συναφή θέματα, όπως το "special theme on algorithms in Culture" στο Big Data & Society πάλι το 2018.

Βέβαια η ιδέα της χρήσης υπολογιστών στην ανθρωπολογία είναι πολύ παλιά (Hymes, 1965), (Burton, 1970), (Burton, 1973), (Wolfe, 1978), αλλά οι όποιες προσπάθειες ήταν μεμονωμένες και αποσπασματικές, εκτός ελαχίστων εξαιρέσεων. Ήδη από το 1973 ο Burton γράφει πως όλο και περισσότεροι ανθρωπολόγοι αναπτύσσουν τα δικά τους προγράμματα για τα εξειδικευμένα προβλήματα της ανθρωπολογίας (Burton, 1973, σελ. 337). Στα μέσα της δεκαετίας του 1980, η πλειοψηφία των ακαδημαϊκών τμημάτων ανθρωπολογίας στις ΗΠΑ διαθέτουν υπολογιστές ως ειδικό εξοπλισμό για τη διεξαγωγή των ακαδημαϊκών τους δραστηριοτήτων. Το ποσοστό των τμημάτων που δίνουν στοιχεία για χρήση υπολογιστών σε επίπεδο διδακτορικών ξεπερνάει το 70% (O'Neil, 1987, σελ. 453). Όσον αφορά στην εποχή μας, η Antonijević δείχνει ότι οι υπολογιστές χρησιμοποιούνται ευρέως για να αυτοματοποιήσουν μια σειρά από διαδικασίες που κάνουν οι κοινωνικοί επιστήμονες, όπως το να διαβάσουν μια έρευνα από αρχεία μορφής pdf, να χρησιμοποιούν υπηρεσίες νέφους (cloud) για να αποθηκεύσουν αρχεία, ή να χρησιμοποιούν το Zotero ως εργαλείο διαχείρισης βιβλιογραφίας (reference management tool) (Antonijević, 2020). Παρόλα αυτά, εξακολουθεί να ισχύει αυτό που σημείωνε ο Wagner τριάντα χρόνια πριν: «υπάρχει ουσιαστική χρήση των υπολογιστών στην ανθρωπολογία, αλλά σχετικά περιορισμένη υπολογιστική ανθρωπολογία» (Wagner, 1989, σελ. 419). Για το λόγο αυτό οι Behrens & Read καλούν, ήδη από το 1993, τους ανθρωπολόγους όχι να χρησιμοποιήσουν τους υπολογιστές για να αυτοματοποιήσουν όσα έκαναν πριν, αλλά για να τους βοηθήσουν να σκεφτούν με διαφορετικό τρόπο (Behrens & Read, 1993, σελ. 429). Όμως, με εξαίρεση τον τομέα της συγγένειας, όπου έχουν γίνει σημαντικά βήματα, η συμμετοχή της ανθρωπολογικής κοινότητας παραμένει σχετικά μικρή. Οι πρωτοπόροι του πεδίου δεν χάνουν ευκαιρία να αναφέρονται στις δυνατότητες που υπάρχουν οι οποίες είναι συμβατές τόσο με 'επιστημονικές' όσο και με μεταμοντέρνες προσεγγίσεις (M. Fischer, 2004, σελ. 154) ενώ αντιμετωπίζουν τις υπολογιστικές μεθόδους ως εργαλεία για την παραγωγή θεωρίας και όχι μόνο ως εργαλεία που έχουν προκύψει από ή για την υποστήριξη θεωριών (M. D. Fischer κ.ά., 2013, σελ. 12).

Η εισαγωγή των προηγμένων τεχνολογιών στα πολιτιστικά συμφραζόμενα συνεπάγεται επίσης την εισαγωγή ενός διαφορετικού συστήματος γνώσης, ενός νέου τρόπου σκέψης, μιας νέας γλώσσας, ενός νέου τρόπου για την ερμηνεία και την καθιέρωση της τάξης στον κόσμο (Combi, 1992, σελ. 43). Άρα είναι σημαντική η δημιουργία, μέσα από διεπιστημονικές προσεγγίσεις, ενός κοινού σημείου θέασης το οποίο θα επιτρέψει την ανάδυση του πεδίου της υπολογιστικής κοινωνικής επιστήμης και την αναγκαία ανάπτυξη ενός νέου παραδείγματος (paradigm) για την εκπαίδευση νέων ερευνητών (Lazer κ.ά., 2009, σελ. 722). Οι νέοι ερευνητές θα πρέπει να αναπτύξουν υπολογιστική σκέψη (computational thinking) με τη λογική της Wing, δηλαδή μια βασική δεξιότητα που δεν αφορά μόνο τους επιστήμονες της πληροφορικής, και η οποία εστιάζει στην εννοιολόγηση, και όχι στον προγραμματισμό. Πρόκειται δηλαδή για έναν τρόπο που οι άνθρωποι χρησιμοποιούν τις ιδέες τους για να προσεγγίσουν και να λύσουν προβλήματα (Wing, 2006). Στη συνέχεια της διατριβής καταγράφεται η ιστορία χρήσης των υπολογιστών στο πεδίο της ανθρωπολογίας και, κυρίως, αναδεικνύονται οι δυνατότητες της υπολογιστικής επεξεργασίας για λογαριασμό της ανθρωπολογίας ως αποτέλεσμα αυτής της «υπολογιστικής σκέψης». Αυτή η οπτική κάνει τον Seaver, ανθρωπολόγο που ισχυρίζεται ότι οι αλγόριθμοι είναι εγγενώς πολιτισμικοί, να σημειώσει ότι δεν είμαστε απομακρυσμένοι παρατηρητές αλλά ενεργοί συμμετέχοντες οι οποίοι παράγουν αλγορίθμους μέσα από την έρευνά τους (Seaver, 2017, σελ. 5).

## 2 Αντικείμενο, στόχοι, δομή και συμβολή της διατριβής

Στο παρόν κεφάλαιο σκιαγραφείται το βασικό αντικείμενο της διατριβής, οι στόχοι που έχει θέσει καθώς και η διάρθρωσή της.

### 2.1 Ερευνητικό ερώτημα, αντικείμενο και στόχοι της διατριβής

Το βασικό ερευνητικό ερώτημα που τίθεται θα μπορούσε να συνοψιστεί ως εξής: πώς η ανθρωπολογία αξιοποιεί την πληροφορική και την επιστήμη της πληροφορίας και, ειδικότερα, πώς οι τεχνολογίες του σημασιολογικού ιστού θα μπορούσαν να συμβάλλουν είτε στη βελτιστοποίηση αυτής της χρήσης είτε στην δημιουργία νέων χρήσεων. Μέσα σε αυτό το πλαίσιο τίθενται τρεις κύριοι στόχοι για την διατριβή.

- Πρώτος στόχος: αναλυτική επισκόπηση της σχέσης της πληροφορικής με την ανθρωπολογία, εστιάζοντας στο πώς η πληροφορική χρησιμοποιείται ήδη από τη ανθρωπολογία ώστε η δεύτερη να επιτύχει τους σκοπούς της.
- Δεύτερος στόχος: κατάδειξη του θεμελιώδους ρόλου της φυσικής γλώσσας, ως μέσο αναπαράστασης της ανθρωπολογικής έρευνας και παρουσίαση τεχνητών γλωσσών οι οποίες είναι κατάλληλες για την αναπαράσταση των εθνογραφικών δεδομένων και της ανθρωπολογικής γνώσης. Οι εν λόγω γλώσσες δεν είναι κειμενοκεντρικές αλλά εννοιοκεντρικές.
- Τρίτος στόχος: παρουσίαση βασικών αρχών για την αναπαράσταση της γνώσης στο σημασιολογικό ιστό (γενικώς), με ειδικό ενδιαφέρον στην αποτύπωση της ανθρωπολογικής γνώσης σκιαγραφώντας τις βασικές προϋποθέσεις για να αναπαρασταθεί η κοινωνική ανθρωπολογία με όρους σημασιολογικού ιστού.

Προς επίτευξη των παραπάνω στόχων εφαρμόζεται διεπιστημονική προσέγγιση, η οποία επιθυμεί να συγκεράσει τρεις διαφορετικές επιστημονικές περιοχές: την κοινωνική ανθρωπολογία, την πληροφορική και την επιστήμη της πληροφορίας` την τελευταία κυρίως ως προς την συμβολή της στην οργάνωση και την αναπαράσταση της γνώσης.

Απώτερος στόχος της διατριβής είναι η διερεύνηση μεθόδων οι οποίες επιτρέπουν τη μείωση της ασάφειας, χωρίς όμως αυτό να οδηγεί σε μείωση της εκφραστικότητας. Η μείωση της ασάφειας κρίνεται καθοριστική για τη μείωση της αμφισημίας και κατ'

επέκταση της αβεβαιότητας. Διότι αυτή η μείωση, οδηγεί με τη σειρά της σε μειωμένα περιθώρια 'παρερμηνείας' από τον αποδέκτη του μηνύματος<sup>3</sup>. Για να επιτευχθεί αυτό απαιτείται η αύξηση του φορμαλισμού του μηνύματος με προκαθορισμένη σαφή μέθοδο. Διότι όσο τα πράγματα είναι "χειροποίητα" υποβόσκει, σε μεγάλο βαθμό, άρρητη γνώση. Όπως το θέτουν οι Carlson & Anderson οι "craft-like methods" συνδυάζονται με πολύ "tacit knowledge" και αυτό οδηγεί σε μειωμένη επανάχρηση (data re-use) (Carlson & Anderson, 2007), με αποτέλεσμα να μειώνεται ο αθροιστικός χαρακτήρας της συσσώρευσης γνώσης που εγγενώς διαθέτει η Επιστήμη.

## 2.2 Δομή της διατριβής

Οι τρεις στόχοι που τέθηκαν στην προηγούμενη υποενότητα δημιουργούν τρεις κύριους άξονες γύρω από τους οποίους περιστρέφεται τόσο το ερευνητικό εγχείρημα όσο και το κείμενο της διατριβής.

- Ο πρώτος άξονας αφορά σε μια αναλυτική επισκόπηση του τρόπου με τον οποίο χρησιμοποίησαν την πληροφορική οι ανθρωπολόγοι, εστιάζοντας στην τελευταία 35ετία. Τα αποτελέσματα αυτής της ανάλυσης παρουσιάζονται στο κεφάλαιο «4 Υπολογιστική επεξεργασία εθνογραφικών δεδομένων: ιστορική αναδρομή, παρόν και προοπτικές».
- Ο δεύτερος άξονας σχετίζεται με την αναπαράσταση και οργάνωση της γνώσης συνολικά, εξειδικεύοντας στη συνέχεια, στην ανθρωπολογική γνώση. Συγκρίνεται η φυσική γλώσσα, ως μέσο αναπαράστασης γνώσης, με τεχνητές γλώσσες όπως ο σημασιολογικός ιστός και τα συστήματα οργάνωσης της γνώσης. Τα αποτελέσματα αυτής της προσέγγισης παρουσιάζονται στο κεφάλαιο «5 Αναπαράσταση και οργάνωση της γνώσης: φυσικές και τεχνητές γλώσσες».
- Ο τρίτος άξονας περιστρέφεται γύρω από τις δυνατότητες που παρέχει ο σημασιολογικός ιστός για την αναπαράσταση της ανθρωπολογικής γνώσης και καταλήγει στην πρόταση ενός μοντέλου αναπαράστασης το οποίο βασίζεται στην ολοκλήρωση της πληροφορίας (information integration) μέσω διαφορετικών επιπέδων αφαίρεσης. Η εν λόγω προσέγγιση παρουσιάζεται στο κεφάλαιο «6 Οντολογική αναπαράσταση Εθνογραφίας και Κοινωνικής Ανθρωπολογίας ».

Κάθε ένας από τους παραπάνω άξονες παρατίθεται σε ένα αυτόνομο κεφάλαιο το οποίο διαθέτει τα δικά του -αυτοτελή- συμπεράσματα. Αφού αναλυθούν οι τρεις παραπάνω συνιστώσες, ακολουθεί στο κεφάλαιο «7 Συζήτηση» μια παράθεση ζητημάτων που προκύπτουν από την συγκριτική θεώρηση των επιμέρους προσεγγίσεων. Για να επέλθει η ολοκλήρωση με το κεφάλαιο «8 Συμπεράσματα », όπου παρουσιάζονται τα συμπεράσματα που διατρέχουν το σύνολο των τριών αξόνων.

---

<sup>3</sup> Προφανώς όλη αυτή η προσέγγιση εδράζεται στην «Θεωρία της πληροφορίας» (Information Theory) του Claude Shannon όπου το κεντρικό ζητούμενο είναι το πόση/ποια πληροφορία είναι απαραίτητη για την αποτελεσματική μεταφορά ενός μηνύματος ('Θεωρία πληροφορίας', 2023).

## 2.3 Συμβολή της διατριβής

Η διατριβή εντάσσεται στο ευρύτερο πεδίο της αναδυόμενης επιστημονικής περιοχής των «υπολογιστικών κοινωνικών επιστημών» (computational social sciences) και διερευνά τη θέση της κοινωνικής ανθρωπολογίας εντός αυτής.

Η μελέτη όμως των επιμέρους αντικειμένων που σχετίζονται με τον τελικό στόχο οδήγησε σε μια σειρά ευρημάτων. Έτσι η διατριβή:

- Ορίζει τι σημαίνει ο επιθετικός προσδιορισμός «υπολογιστική» για την ανθρωπολογία και την εθνογραφία, αφού αποθησαύρισε τους ετερόκλητους ορισμούς.
- Παρέχει εκτενή ιστορική επισκόπηση του πεδίου της χρήσης των υπολογιστών στην ανθρωπολογία από την δεκαετία του 1970 και μετά, δίνοντας έμφαση στα τελευταία τριάντα πέντε χρόνια.
- Αναδεικνύει την αξία της μελέτης των ψηφιακών ιχνών (digital traces) με ειδική αναφορά στα big data.
- Εξετάζει συγκριτικά τους ενδεικτικούς τύπους συστημάτων οργάνωσης της γνώσης και τα αντιπαραβάλλει με τις οντολογίες. Στο στάδιο αυτό κάνει εκτεταμένη χρήση τεχνολογιών επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP).
- Επισημαίνει τη δυναμική και τις δυνατότητες αναπαράστασης της κοινωνικής ανθρωπολογίας με γνώμονα τις λογικές του σημασιολογικού ιστού και προτείνει ένα συγκεκριμένο μοντέλο για την αναπαράσταση εθνογραφικών δεδομένων και ανθρωπολογικής γνώσης.
- Καταγράφει τα οφέλη που θα έχει για την κοινωνική ανθρωπολογία η αναπαράσταση της ανθρωπολογικής γνώσης με όρους σημασιολογικού ιστού, αλλά και τις προκλήσεις που έχουν να αντιμετωπίσουν οι ανθρωπολόγοι.
- Καταδεικνύει ότι η φυσική γλώσσα δεν είναι η μόνη εναλλακτική για την αναπαράσταση της ανθρωπολογικής γνώσης.
- Προτείνει ένα μοντέλο για την αποτύπωση της ανθρωπολογικής γνώσης το οποίο επιτρέπει την ολοκλήρωση (integration) της πληροφορίας και την συγκριτική υπολογιστική ανθρωπολογία.

Πέραν του κειμένου της παρούσας διατριβής, τα αποτελέσματα της έρευνας έχουν αποτυπωθεί -έως τώρα- στις εξής δημοσιεύσεις σε έγκριτα διεθνή επιστημονικά περιοδικά:

1. Peponakis, M., Kapidakis, S., Doerr, M., & Tountasaki, E. (2024). From calculations to reasoning: History, trends and the potential of Computational Ethnography and Computational Social Anthropology. *Social Science Computer Review*, 42(1), 84–102. [DOI:10.1177/08944393231167692](https://doi.org/10.1177/08944393231167692)
2. Peponakis, M., Mastora, A., Kapidakis, S., & Doerr, M. (2019). Expressiveness and machine processability of Knowledge Organization Systems (KOS): An analysis of concepts and relations. *International Journal on Digital Libraries*, 20(4), 433–452. [DOI:10.1007/s00799-019-00269-0](https://doi.org/10.1007/s00799-019-00269-0)

3. Mastora, A., Peponakis, M., & Kapidakis, S. (2017). SKOS Concepts and Natural Language Concepts: An Analysis of Latent Relationships in KOSs. *Journal of Information Science*, 43(4), 492–508. [DOI:10.1177/200165551516648108](https://doi.org/10.1177/200165551516648108)

### 3 Μεθοδολογία

Η επιστήμη, στο σύνολό της, βασίζεται σε εμπειρικά δεδομένα. Ο Poincaré σημειώνει πως, όπως τα σπίτια φτιάχνονται από τούβλα, η επιστήμη φτιάχνεται από γεγονότα. Όμως, όπως μια στοίβα τούβλα δεν αποτελούν ένα σπίτι έτσι και η συσσώρευση γεγονότων και δεδομένων δεν αποτελεί επιστήμη (Poincaré, 1952, σελ. 141). Για πολλούς ερευνητές, η μεθοδολογία που αφορά τόσο στη συγκέντρωση όσο και στην ανάλυση των δεδομένων είναι ο κρίσιμος παράγοντας και αποτελεί το θεμέλιο λίθο της σύγχρονης επιστήμης.

Στη μεταμοντέρνα εποχή ο ιστορικός της επιστήμης Thomas Kuhn, στο κλασικό έργο του «Η δομή των επιστημονικών επαναστάσεων» (T. S. Kuhn, 1970), τεκμηριώνει το πώς μια δεδομένη επιστημολογική προσέγγιση δημιουργεί ένα 'παράδειγμα' (paradigm). Στον ιστορικό κύκλο που διαρκεί το 'παράδειγμα' δημιουργείται μια σειρά από συγκεκριμένα επιστημονικά ζητήματα τα οποία συναρτώνται με συγκεκριμένες μεθόδους και εργαλεία για την επίλυσή τους. Συνεπώς το 'παράδειγμα' είναι το γενικό πλαίσιο μέσα στο οποίο αναπτύσσονται συγκεκριμένες θεωρίες και μέθοδοι. Όμως για κάποιους ερευνητές της επιστημολογίας η προσήλωση σε συγκεκριμένες μεθόδους καταλήγει να γίνει τροχοπέδη που εμποδίζει την πρόοδο. Για να αποφευχθεί αυτό, ο Feyerabend, στο έτερο κλασικό της επιστημολογίας με τίτλο «Against method», καλεί σε επιστημολογική αναρχία. Όχι με την έννοια ότι δεν υπάρχει μέθοδος που σε οδηγεί με ασφάλεια στην απόδειξη επιστημονικών ερωτημάτων αλλά αφήνοντας πάντα ανοικτή την πιθανότητα να υπάρχει και άλλη αποδεκτή ερμηνεία μέσω της εφαρμογής άλλων μεθόδων. Για το λόγο αυτό κρίνει ως πολύ σημαντικό παράγοντα της επιστήμης την αμφισβήτηση των καθιερωμένων προτύπων (Feyerabend, 2009).

Η παρούσα διατριβή προσπάθησε να επιτύχει μια ισορροπία ανάμεσα στον εκάστοτε μεθοδολογικό φορμαλισμό και τον δημιουργικό μετασχηματισμό του μέσω επιρροών από συγγενείς -ή και όχι- μεθοδολογικές προσεγγίσεις. Με δεδομένο ότι η έρευνα που προηγήθηκε της συγγραφής του κειμένου της διατριβής ήταν διεπιστημονική, η πολυσυλλεκτικότητα των επιστημονικών τομέων συνέτεινε στην

αύξηση των μεθόδων που χρησιμοποιήθηκαν στις διαφορετικές φάσεις της έρευνας<sup>4</sup>, αλλά και στους δυνητικούς συνδυασμούς μεθόδων.

Αρχικά σημειώνεται ότι η διατριβή έχει ως σημείο εκκίνησης τη θεμελιωμένη θεωρία (Grounded Theory). Η θεμελιωμένη θεωρία είναι μια γενική μεθοδολογία η οποία επιτρέπει την ανάπτυξη θεωριών οι οποίες θεμελιώνονται σε δεδομένα τα οποία έχουν συλλεχθεί και αναλυθεί με συστηματικό τρόπο (Strauss & Corbin, 1994, σελ. 273). Ένα σημαντικό στοιχείο της θεμελιωμένης θεωρίας είναι ότι δεν είναι απαραίτητο να τεθεί εκ των προτέρων ένα ερευνητικό ερώτημα ως αφετηρία της έρευνας, αφού αυτό μπορεί να προκύψει στην πορεία από την ανάλυση των δεδομένων. Με τον τρόπο αυτό μειώνεται ο βαθμός της μεροληψίας (bias) διότι τα δεδομένα δεν συλλέγονται για να αποδείξουν κάτι αλλά οδηγούν τα ίδια στον τρόπο της ανάλυσής τους καθώς και στα ερωτήματα που είναι δυνατόν να απαντηθούν από αυτά.

Στην περίπτωση της συγκεκριμένης διατριβής τα δεδομένα είναι μελέτες, κείμενα, λογισμικά, πληροφοριακά συστήματα, οντολογίες, συστήματα οργάνωσης της γνώσης και δομές δεδομένων. Για να ενταχθούν αυτά σε ένα σύνολο ακολουθήθηκε η μέθοδος της μετα-σύνθεσης (meta-synthesis) (Walsh & Downe, 2005). Συνήθως η μετα-σύνθεση πραγματοποιείται σε έξι στάδια: καθορισμός του γενικού ερωτήματος και των κριτηρίων συμπερίληψης, επιλογή των μελετών, αξιολόγηση της ποιότητάς τους, εξαγωγή και παρουσίαση των επίσημων δεδομένων, ανάλυση των δεδομένων και έκφραση της σύνθεσης (Lachal κ.ά., 2017). Σε αυτό το πλαίσιο, η παρούσα μελέτη επιχειρεί να εντάξει σε ένα λειτουργικό σύνολο τα αποτελέσματα από μια σειρά διαφορετικών -αλλά αλληλένδετων- μελετών, λογισμικών, δομών δεδομένων κτλ. Ωστόσο διακατέχεται μάλλον από μια ερμηνευτική παρά από μια συγκεντρωτική αντίληψη. Έτσι καταλήγει σε αυτό που οι Edwards & Kaimal ονομάζουν κριτική ερμηνευτική σύνθεση (critical interpretive synthesis) (Edwards & Kaimal, 2016).

Επιχειρώντας μια σύνοψη των παραπάνω, θα μπορούσε να ειπωθεί ότι το γενικότερο μεθοδολογικό πλαίσιο της διατριβής εδράζεται στη θεωρητική προσέγγιση της «θεμελιωμένης θεωρίας» (grounded theory) και από εκεί οδηγείται στη μεθοδολογία της μετα-σύνθεσης (meta-synthesis). Όμως, δεν αρκείται στον αθροιστικό χαρακτήρα της μετα-σύνθεσης, αλλά προβαίνει τελικά σε μία κριτική ερμηνευτική σύνθεση.

Ο απώτερος στόχος είναι η ολιστική ανάλυση όπως την ορίζουν οι Gorichanaz & Latham: όλα τα μέρη που αναλύονται πρέπει να εξετάζονται όχι μόνο από μόνα τους, αλλά και μεταξύ τους ως μέρη ενός αλληλένδετου συνόλου. Επιπλέον, οι οντότητες που μπορεί να φαίνονται "ολόκληρες" από μόνες τους μπορούν, με τη σειρά τους, να θεωρηθούν προοδευτικά ως μέρη πιο πολύπλοκων συνόλων (Gorichanaz & Latham, 2016, σελ. 1116).

---

<sup>4</sup> Η εναλλαγή των μεθόδων ανά στάδιο της έρευνας θα παρουσιαστεί στην επόμενη υποενότητα.

### 3.1 Εναλλαγή μεθόδων

Το σύνολο της έρευνας μπορεί να χωριστεί σε τρία διακριτά στάδια τα οποία βρίσκονται σε απόλυτη σύμπτωση με τους τρεις άξονες που τέθηκαν στο προηγούμενο κεφάλαιο, στην ενότητα «2.1 Ερευνητικό ερώτημα, αντικείμενο και στόχοι της διατριβής». Με εξαίρεση τη βιβλιογραφική επισκόπηση -και την πολυδιάστατη βιβλιομετρική ανάλυση που την συνοδεύει-, καθένα από αυτά τα στάδια απαιτεί διαφορετική μεθοδολογική προσέγγιση για την αποτελεσματική ανάλυση των δεδομένων και των παραμέτρων που λαμβάνονται υπόψη.

Έτσι, κατά το πρώτο στάδιο -όπου εξετάζεται η χρήση της πληροφορικής από τους ανθρωπολόγους τόσο ως εργαλείο όσο και ως μέθοδος για την υποβοήθηση της ανθρωπολογικής έρευνας- η βασική μεθοδολογία που χρησιμοποιείται είναι η βιβλιογραφική επισκόπηση και οι πειραματισμοί με το σχετικό λογισμικό. Αυτό είναι σαφές ότι διεξάγεται στο πλαίσιο της θεμελιωμένης θεωρίας και της μετα-σύνθεσης που αναλύθηκε παραπάνω.

Κατά το δεύτερο στάδιο εξετάζονται συστήματα οργάνωσης και αναπαράστασης γνώσης και συγκρίνονται ως προς την εκφραστικότητα και την αποτελεσματικότητα της υπολογιστικής επεξεργασίας (expressiveness and machine processability). Στη φάση αυτή εγκαταλείπεται η θεμελιωμένη θεωρία διότι τίθεται εξ αρχής ένα σαφές ερευνητικό ερώτημα: τι μπορούν να εκφράσουν οι διάφοροι τύποι συστημάτων οργάνωσης της γνώσης και κατά πόσο αυτό που εκφράζουν μπορεί να είναι μηχαναγνώσιμο στο περιβάλλον του σημασιολογικού ιστού. Πέραν από τη βιβλιογραφική επισκόπηση του πεδίου και την ανάλυση των τύπων εννοιών και σχέσεων που περιέχουν τα ΣΟΓ, η βασική μεθοδολογία που χρησιμοποιείται είναι η εκτεταμένη χρήση τεχνολογιών επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP) για την ανάλυση των συστημάτων οργάνωσης της γνώσης.

Κατά το τρίτο στάδιο της έρευνας εξετάζονται οι δυνατότητες κωδικοποίησης της ανθρωπολογικής πληροφορίας για το σημασιολογικό ιστό με έμφαση στις οντολογικές αναπαραστάσεις. Μεθοδολογία: βιβλιογραφική επισκόπηση (με έμφαση στα πρότυπα του W3C για τον σημασιολογικό ιστό), μέθοδοι ανάπτυξης οντολογιών και χρήση του σχετικού λογισμικού. Εδώ σημαντικό ρόλο έπαιξαν οι επεξεργαστές οντολογιών και ιδιαίτερα το Protégé το οποίο χρησιμοποιήθηκε τόσο για την ανάπτυξη XML/RDF κώδικα όσο και για τον έλεγχο της εγκυρότητας και επαληθευσιμότητας των συμπερασμάτων. Βλέπε και παρακάτω υποενότητα «3.3 Έλεγχοι εγκυρότητας και επαληθευσιμότητα».

Προφανώς το τελικό στάδιο της συγγραφής του κειμένου ενσωματώνει τα παραπάνω και αποτελεί τη μετασύνθεση με την οπτική μιας κριτικής ερμηνευτικής σύνθεσης ολιστικής προσέγγισης.

### 3.2 Στρατηγική αναζήτησης και τεχνικές αναζήτησης πηγών

Ο χρόνος περάτωσης μιας διδακτορικής διατριβής είναι αρκετά μακρύς, ειδικά αν ο συγγραφέας της έχει 'σχέση μερικής απασχόλησης' με την εκπόνησή της. Ως εκ τούτου, η διαδικασία αναζήτησης των σχετικών πηγών δεν πραγματοποιείται άπαξ. Αντιθέτως πρέπει να υπάρχει ένα σχέδιο για τη συνεχή παρακολούθηση της σχετικής

βιβλιογραφίας. Επειδή η μεθοδολογία εκπόνησης της παρούσας διατριβής εφαρμόζει εκτεταμένη επισκόπηση της βιβλιογραφίας -καθώς και του σχετικού λογισμικού και υπηρεσιών ιστού (web services)- η ανάγκη αυτή γίνεται εντονότερη.

Η αναζήτηση στις βιβλιογραφικές βάσεις δεδομένων χαρακτηρίζεται από δύο σημαντικές παραμέτρους. Η πρώτη παράμετρος είναι η ανάκληση (recall) και σχετίζεται με το ποσοστό των σχετικών τεκμηρίων της βάσης τα οποία θα ανακτηθούν. Η δεύτερη παράμετρος είναι η ακρίβεια (precision) και σχετίζεται με το ποσοστό των σχετικών τεκμηρίων εντός του συνόλου των αποτελεσμάτων που ανακτήθηκαν. Ο αρμονικός μέσος αυτών των δύο, δηλαδή η ισορροπία ανάκλησης-ακρίβειας, ορίζεται από τη μετρική F-measure<sup>5</sup>. Η βασική στρατηγική αναζήτησης ήταν να κρατηθεί όσο γίνεται πιο ψηλά το ποσοστό της ανάκλησης, διότι ο στόχος ήταν να ανακτηθούν, ει δυνατόν, όλα τα σχετικά τεκμήρια. Άρα οι αναζητήσεις δεν κρίνονταν ως επιτυχημένες βάσει των κριτηρίων της αναζήτησης πληροφοριών (information retrieval), δηλαδή με υψηλό F-measure, αλλά η υψηλή ανάκληση ήταν θεμιτή ακόμη και αν αυτό είχε αρνητική επίπτωση στην ακρίβεια και, ως εκ τούτου, στο F-measure.

Όταν το ζητούμενο είναι η υψηλή ανάκληση, είναι προφανές ότι αυξάνεται ο όγκος του υλικού που πρέπει να ελεγχθεί ως προς τη σχετικότητά του. Έτσι, στο ερώτημα [*(anthropolog\* OR ethnograph\*) AND ("knowledge representation" OR owl OR rdf OR "semantic web")*] είναι πιθανό να επιστραφούν στα αποτελέσματα δημοσιεύσεις που σχετίζονται με τις κουκουβάγιες και την ανθρωπολογία εξαιτίας της διττής σημασίας της λέξης OWL: που είναι συντομογραφία του Web Ontology Language, και του αγγλικής λέξης που αντιστοιχεί στο πτηνό *κουκουβάγια*.

Τα ερωτήματα είναι προσαρμοσμένα στις προδιαγραφές και τις δυνατότητες των αντίστοιχων βάσεων. Έτσι, στο Google scholar, όπου δεν υποστηρίζεται αποκοπή (truncation), επαναλαμβάνονται οι διάφορες παραλλαγές των όρων, ενώ στη Scopus χρησιμοποιούνται οι λεγόμενοι χαρακτήρες 'μπαλαντέρ' (wildcard characters). Για παράδειγμα, η ερώτηση για άρθρα που σχετίζονται με ανθρωπολογία και πληροφορική στη Scopus μπορεί να έχει τη μορφή [*anthropolog\* AND comput\**].

Για τον εντοπισμό της βιβλιογραφίας που καλύπτει τα θέματα ενδιαφέροντος της παρούσας μελέτης, στην αρχική φάση, έγιναν πολλαπλές αναζητήσεις θέτοντας κατάλληλα κριτήρια σε βάσεις δεδομένων, όπως η Web of Science (Clarivate analytics), η Scopus, και η Google scholar. Το λογισμικό και οι υπηρεσίες ιστού εντοπίστηκαν τόσο μέσω της βιβλιογραφίας που συγκεντρώθηκε όσο και μέσω αναζήτησης σε εξειδικευμένα αποθετήρια λογισμικού όπως το Git-Hub. Όμως επειδή κεντρικό ζητούμενο ήταν η θεωρητική θεμελίωση των διαδικασιών που το λογισμικό επιχειρεί να αυτοματοποιήσει, έμφαση δίνεται σε λογισμικά που συνοδεύονται από επιστημονικές δημοσιεύσεις οι οποίες τεκμηριώνουν επί των θεωρητικών αρχών τους. Έκτοτε παρακολουθούνται συστηματικά δύο κύριες βάσεις δεδομένων: η Scopus και η Google scholar. Και στις δύο έχει ενεργοποιηθεί υπηρεσία ειδοποιήσεων (alerts) για την

---

<sup>5</sup> Το F-measure είναι ένας τυπικός δείκτης που μετράει την αποτελεσματικότητα μια επερώτησης (query) στο πλαίσιο της ανάκτησης πληροφοριών (information retrieval).

παρακολούθηση ό,τι σχετικού εκδίδεται και ανήκει στα ερευνητικά ενδιαφέροντα του συγγραφέα. Έτσι εξασφαλίζεται η αδιάλειπτη και συστηματική παρακολούθηση της σχετικής βιβλιογραφίας κατά τη διάρκεια της έρευνας.

### 3.3 Έλεγχοι εγκυρότητας και επαληθευσιμότητα

Ένα σημαντικό πλεονέκτημα των υπολογιστικών μεθόδων είναι η ύπαρξη τρόπων ελέγχου εγκυρότητας των δηλώσεων και επαληθευσιμότητας των συμπερασμάτων. Οι εν λόγω διαδικασίες δεν είναι πάντα εφικτές, ούτε γίνονται με τον ίδιο τρόπο για όλα τα συστήματα της πληροφορικής. Στην παρούσα διατριβή αναπτύχθηκε κώδικας που αφορά μοντελοποίηση δεδομένων με στόχο τον αυτόματο συμπερασμό (reasoning) ή την διευκόλυνση της επεξεργασίας των δεδομένων.

Ένας μεγάλο μέρος της έρευνας περιστρέφεται γύρω από την αναπαράσταση γνώσης με τεχνολογίες σημασιολογικού ιστού, με έμφαση στο RDF και την OWL<sup>6</sup>. Για τον έλεγχο της συντακτικής εγκυρότητας του RDF κώδικα<sup>7</sup> που υπάρχει στην παρούσα διατριβή χρησιμοποιήθηκε ο RDF validator του W3C, ο οποίος είναι διαθέσιμος στο <https://www.w3.org/RDF/Validator/>. Ενώ οι όποιες οντολογικές αναπαραστάσεις σε OWL έχουν ελεγχθεί μέσω του επεξεργαστή οντολογιών Protégé. Για τις οντολογικές αναπαραστάσεις έγιναν πειραματισμοί με διάφορα λογισμικά (όπως το Ontorion Fluent editor κ.α.). Όμως το λογισμικό που χρησιμοποιήθηκε για τις δοκιμές ήταν το Protégé. Για την επαλήθευση των όποιων συμπερασμάτων χρησιμοποιήθηκε ο HermiT OWL Reasoner, ο οποίος ενσωματώνεται στο προηγούμενο λογισμικό. Σημειώνεται ότι η επαλήθευση αφορά κυρίως στα παραδείγματα που παρουσιάζονται σε φυσική γλώσσα στο κεφάλαιο 6.

Τέλος, στη διατριβή υπάρχει αναφορά στη γλώσσα Prolog, χρήση της οποίας γίνεται ήδη για τη διεξαγωγή ανθρωπολογικής έρευνας. Στην παρούσα διατριβή χρησιμοποιήθηκε η SWI-Prolog που είναι διαθέσιμη στο <https://www.swi-prolog.org/>. Το ενδεικτικό απόσπασμα κώδικα σε Prolog που παρατίθεται, έχει ελεγχθεί ως προς την συντακτική εγκυρότητά του, αλλά και ως προς το συμπερασμό του, μέσω του on-line περιβάλλοντος ανάπτυξης κώδικα που παρέχει η πιο πάνω πλατφόρμα στο <https://swish.swi-prolog.org/>.

### 3.4 Περί ορολογίας

Η βιβλιογραφία που ελήφθη υπόψη αποτελούνταν, στη συντριπτική της πλειοψηφία, από τεκμήρια στην αγγλική γλώσσα. Δεδομένου ότι, για πλήθος όρων, δεν υπάρχει καθιερωμένη ή κοινώς αποδεκτή ελληνική ορολογία, κρίθηκε σκόπιμη η αντιστοιχία των συγκεκριμένων όρων στον αντίστοιχο αγγλικό όρο για την καλύτερη αποσαφήνιση των

<sup>6</sup> Τόσο το RDF όσο και η OWL αναλύονται διεξοδικά στο κεφάλαιο 5 όπου εξετάζεται ο σημασιολογικός ιστός.

<sup>7</sup> Σημειώνεται ότι στα παραδείγματα χρησιμοποιείται, για την παραγωγή των URIs, το domain "MyPhDExamples" το οποίο προφανώς είναι πλασματικό και δεν αντιστοιχεί σε πραγματικό domain name.

νοημάτων. Η διάκριση ανάμεσα στο «καθιερωμένη» και το «κοινώς αποδεκτή» δηλώνει ότι πολλές φορές υπάρχει 'θεσμικά' καθιερωμένη ορολογία (π.χ. μέσω του Ελληνικού Οργανισμού Τυποποίησης (ΕΛΟΤ)) όμως αυτή δεν ακολουθείται από τους συγγραφείς του εκάστοτε επιστημονικού πεδίου.

Ένα επιπλέον πρόβλημα είναι ότι πολλές φορές οι λέξεις αποκτούν διαφορετικό νόημα ανάλογα με τα συμφραζόμενα, όπως άλλωστε δείχνει και η παρούσα διατριβή στην ενότητα «5.1 Η φυσική γλώσσα ως φορέας νοήματος». Εξαιτίας του γεγονότος ότι η διατριβή συνδυάζει οπτικές από διαφορετικούς τομείς (δηλαδή διαφορετικά συμφραζόμενα) το πρόβλημα μεγεθύνεται. Ενδεικτικό παράδειγμα είναι ο αγγλικός όρος «reflexivity». Στα συμφραζόμενα της ανθρωπολογίας ο όρος αποδίδεται στα ελληνικά ως «αναστοχασμός» ή «αναστοχαστικότητα». Ο ίδιος αγγλικός όρος για το πεδίο της επιστήμης της πληροφορίας έχει τελείως διαφορετική σημασία και αποδίδεται ως «ανακλαστικότητα».

Για τους παραπάνω λόγους κρίθηκε επιβεβλημένη η προσθήκη ενός γλωσσαρίου στο τέλος του παρόντος κειμένου. Στο εν λόγω Γλωσσάρι παρουσιάζονται οι αντιστοιχίες των όρων ανάμεσα στην αγγλική και την ελληνική γλώσσα. Για μεγαλύτερη διευκόλυνση του αναγνώστη ο αγγλικός όρος πολλές φορές έπεται -εντός παρένθεσης- του ελληνικού όρου και μέσα στο κείμενο.

## 4 Υπολογιστική επεξεργασία εθνογραφικών δεδομένων: ιστορική αναδρομή, παρόν και προοπτικές

Ο σκοπός του κεφαλαίου είναι να αποτυπώσει την ιστορική εξέλιξη της χρήσης των υπολογιστών από τους ανθρωπολόγους ως εργαλεία που επικουρούν την έρευνά τους και να δείξει την τρέχουσα αλλά και τη δυνητική τους χρήση. Η πρώτη υποενότητα του κεφαλαίου (υποενότητα 4.1) επιχειρεί να ορίσει τι είναι υπολογιστική ανθρωπολογία και εθνογραφία, ενώ η δεύτερη (υποενότητα 4.2) αναζητά ορισμούς για το ποια είναι τα προς ανάλυση δεδομένα. Στη συνέχεια, τα ζητήματα επεξεργασίας χωρίζονται σε τέσσερις βασικές κατηγορίες.

- Η πρώτη κατηγορία (υποενότητα 4.3) περιλαμβάνει ζητήματα ανάλυσης δεδομένων με ημιαυτόματο τρόπο.
- Η δεύτερη κατηγορία (υποενότητα 4.4) ασχολείται με την απευθείας εξαγωγή συμπερασμάτων.
- Η τρίτη κατηγορία (υποενότητα 4.5) παρουσιάζει ζητήματα που σχετίζονται με την αναζήτηση και την ανάκτηση ανθρωπολογικής πληροφορίας.
- Και η τέταρτη κατηγορία (υποενότητα 4.6) πραγματεύεται ζητήματα μοντελοποίησης των δεδομένων.

Η αποτελεσματική επεξεργασία των δεδομένων μπορεί να επιτευχθεί με δύο προσεγγίσεις, όχι κατ' ανάγκη αντιθετικές μεταξύ τους. Η μία προσέγγιση στοχεύει στην παραγωγή αλγορίθμων οι οποίοι επιχειρούν να καταλήξουν σε συμπεράσματα επεξεργαζόμενοι τα δεδομένα στη μορφή που αυτά υπάρχουν. Που σημαίνει ότι οι αλγόριθμοι προσαρμόζονται στα υπάρχοντα δεδομένα ανεξάρτητα από την κωδικοποίηση των δεδομένων. Ένα παράδειγμα αυτής της προσέγγισης είναι οι αλγόριθμοι επεξεργασίας φυσικής γλώσσας και τα μοντέλα μηχανικής μάθησης. Μια δεύτερη προσέγγιση είναι η μοντελοποίηση των ίδιων των δεδομένων με τρόπο που θα καταστήσει αποτελεσματικότερη και πιο αξιόπιστη την επεξεργασία τους. Ένα παράδειγμα αυτής της προσέγγισης είναι η κωδικοποίηση ενός κειμένου μέσω του προτύπου Text Encoding Initiative (TEI).

Κάτω από αυτό το πρίσμα, η υποενότητα «4.3 Επεξεργασία δεδομένων: στατιστική ανάλυση, μοτίβα (patterns) και επισημειώσεις (tagging)» αφορά στην πρώτη προσέγγιση, αυτή της παραγωγής αλγορίθμων και λογισμικών, παρά στη μοντελοποίηση των δεδομένων –η μοντελοποίηση θα μας απασχολήσει στην ενότητα 4.6–, παρουσιάζοντας τις διαδικασίες που αυτοματοποιούνται με στόχο να βοηθήσουν στην επεξεργασία δεδομένων και αναζήτηση πληροφοριών για τους σκοπούς της ανθρωπολογίας. Η επεξεργασία των δεδομένων αφορά σε τρόπους και εργαλεία που χρησιμοποιούνται για την ανάλυση των δεδομένων ώστε να έχουν οι ανθρωπολόγοι τη δυνατότητα να εξαγάγουν συμπεράσματα από τα δεδομένα τους και να τεκμηριώνουν επί των συμπερασμάτων. Προφανώς εστιάζει σε διαδικασίες που αφορούν την ανθρωπολογική μελέτη αμιγώς, και όχι σε διαδικασίες οι οποίες είναι, μεν, προαπαιτούμενες, αλλά δεν αφορούν αμιγώς την ανθρωπολογία, όπως η επεξεργασία κειμένων ή η αναζήτηση σε βιβλιογραφικές βάσεις δεδομένων για τον εντοπισμό σχετικών τεκμηρίων. Ωστόσο, η διαδικασία της αναζήτησης αφορά την παρούσα ενότητα όταν πρόκειται για αναζήτηση εθνογραφικών ή ανθρωπολογικών πληροφοριών, όπως στην περίπτωση της βάσης δεδομένων Human Relations Area Files (HRAF).

Στο σημείο αυτό είναι σκόπιμο να διευκρινιστούν τα κριτήρια που ακολουθήθηκαν για τη διάκριση ανάμεσα στις υποενότητες «4.3 Επεξεργασία δεδομένων: στατιστική ανάλυση, μοτίβα (patterns) και επισημειώσεις (tagging)» και «4.5 Αναζήτηση πληροφορίας». Στην πρώτη περίπτωση το πεδίο εφαρμογής προσδιορίζεται στη χρήση που κάνει ο εθνογράφος σε ένα εργαλείο το οποίο ο ίδιος εμπλουτίζει με δεδομένα, ενώ στη δεύτερη περίπτωση ο εθνογράφος χρησιμοποιεί δεδομένα άλλων, τα οποία είναι ήδη καταχωρημένα σε ένα σύστημα, με σκοπό να εκτελέσει μια άλλη εργασία, η οποία συνήθως έχει να κάνει με μετα-ανάλυση.

## 4.1 Υπολογιστική ανθρωπολογία και εθνογραφία: ορισμοί

Η επισκόπηση της βιβλιογραφίας καταδεικνύει σημαντική ασυμβατότητα σχετικά με τον ορισμό της φύσης της υπολογιστικής εθνογραφίας ή ανθρωπολογίας. Η ασυμβατότητα αυτή προκύπτει, κυρίως, από τις διαφορετικές οπτικές μελέτης του συγκεκριμένου πεδίου. Όπως σημειώνει ο Bhargwani, γενικά οι ανθρωπολόγοι είναι διστακτικοί στο να εφαρμόσουν πολύ-επιστημονικές (multidisciplinary) μεθόδους και ακόμη περισσότερο να αναμείξουν σε αυτές ποσοτικές και φορμαλιστικές προσεγγίσεις (Bhargwani, 2006, σελ. 79). Ωστόσο, εντοπίζονται περιπτώσεις ανθρωπολόγων που ασχολούνται τόσο με τους φορμαλισμούς όσο και με την επεξεργασία τους μέσω υπολογιστών. Παράλληλα, εντοπίζεται πλήθος ερευνητών οι οποίοι δεν προέρχονται από το πεδίο της ανθρωπολογίας αλλά από άλλους επιστημονικούς τομείς, όπως η πληροφορική, οι οποίοι ασχολούνται με ανθρωπολογικά δεδομένα και τους υπολογιστικούς φορμαλισμούς αυτών. Αυτό επιτείνει την έλλειψη συμφωνίας, κυρίως γύρω από το τι σημαίνει ο επιθετικός προσδιορισμός «υπολογιστική» όταν συνοδεύει την εθνογραφία ή την ανθρωπολογία.

Ο Artmann, στο συλλογικό έργο “21st Century Anthropology: A Reference Handbook” (Birx, 2010), συνδέει την τεχνητή νοημοσύνη με την υπολογιστική ανθρωπολογία με την έννοια ότι όταν η πρώτη θα είναι σε θέση να περιγράψει το

γνωσιακό υπόβαθρο του ανθρώπινου νου με τη μορφή της προγραμματιστικά επεξεργάσιμης ψηφιακής πληροφορίας, τότε θα έχουμε υπολογιστική ανθρωπολογία (Artmann, 2010). Ενώ στο, επίσης συλλογικό έργο, “The Routledge companion to digital ethnography” η Beaulieu αντιμετωπίζει την υπολογιστική εθνογραφία περισσότερο όπως αντιμετωπίζει η Hine την εικονική εθνογραφία, δηλαδή ως εθνογραφία στον εικονικό ψηφιακό κόσμο. Και θεωρεί ότι γίνεται προβληματική η θέση του εθνογράφου ως εργαλείου της εθνογραφικής έρευνας, διότι εκλείπουν δύο σημαντικές παράμετροι της παραδοσιακής εθνογραφίας: η φυσική παρουσία του εθνογράφου και η προσωπική σχέση του με τους ανθρώπους (presence and engagement) (Beaulieu, 2017). Οι Moore κ.ά. θεωρούν ότι η υπολογιστική εθνογραφία επεκτείνει την παραδοσιακή ποιοτική μεθοδολογική εργαλειοθήκη της εθνογραφίας προσθέτοντας υπολογιστικές μεθόδους. Συνεχίζει να αξιοποιεί τη δύναμη της εθνογραφίας όσον αφορά την κατανόηση των βιωμένων εμπειριών των ατόμων/ομάδων που δίνουν νόημα στα δεδομένα, ενώ παράλληλα τοποθετεί ένα συμμετοχικό φακό για το πώς τα ευρύτερα συστήματα - συμπεριλαμβανομένων των κανόνων, των αξιών, των παραδοχών- κωδικοποιούνται πολιτισμικά και αναπαράγονται μέσω του σχεδιασμού των κοινωνικοτεχνικών συστημάτων που βασίζονται σε δεδομένα. (Moore κ.ά., 2020) Και η Christin κάνει λόγο για αλγοριθμική εθνογραφία την οποία ορίζει ως την εθνογραφική μελέτη του τρόπου με τον οποίο τα υπολογιστικά συστήματα δομούν τις διαδικτυακές δραστηριότητες και το στηρίζει σε τρεις οπτικές: α) οι μελετητές πρέπει να παρακολουθούν προσεκτικά τα δεδομένα, επειδή οι ψηφιακές πλατφόρμες δεν είναι ουδέτεροι μεσάζοντες β) οι εθνογράφοι πρέπει να εστιάζουν στις λεπτομέρειες της αλγοριθμικής ταξινόμησης, δεδομένου ότι οι πλατφόρμες φιλοξενούν περισσότερο περιεχόμενο από ό,τι προβάλλουν γ) οι εθνογράφοι πρέπει να περιλαμβάνουν μετρήσεις στην επιτόπια έρευνα και να μελετούν τις επιπτώσεις τους στις αλληλεπιδράσεις, τις ιεραρχίες και τις αναπαραστάσεις και καλεί τους ποιοτικούς μελετητές που αναλύουν ψηφιακά δεδομένα να ασχοληθούν επίσης με την αλγοριθμική εθνογραφία (Christin, 2020a).

Από την άλλη πλευρά, επιστήμονες που προέρχονται από το χώρο της πληροφορικής εκφράζουν διαμετρικά αντίθετη άποψη ως προς το πώς ορίζεται η υπολογιστική εθνογραφία. Για παράδειγμα, οι Zheng κ.α. συσχετίζουν την υπολογιστική εθνογραφία με τη συλλογή δεδομένων με αυτόματο και ηλεκτρονικό τρόπο (Zheng κ.ά., 2015, σελ. 114). Παρόμοια προσέγγιση ακολουθούν οι Tallyn κ.α. οι οποίοι εξετάζουν τις προοπτικές και τις προκλήσεις που έχει η συλλογή στοιχείων απουσία εθνογράφου με τη χρήση ενός chatbot (Tallyn κ.ά., 2018). Ενώ οι Arnold & Fuller γράφουν πως η υπολογιστική εθνογραφία αφορά στη χρήση μεγάλων σωμάτων δεδομένων για να παρέχει “insights” στη ρουτίνα των χρηστών και στην πορεία θα χρησιμοποιηθούν στις προσπάθειες σχεδιασμού (Arnold & Fuller, 2018).

Από την εξέταση των ορισμών που παρατέθηκαν διαπιστώνεται η ύπαρξη διαφορετικού σημείου εκκίνησης: σε άλλες περιπτώσεις το σημείο εκκίνησης αφορά στις μεθόδους πρόσκτησης της πληροφορίας ενώ σε άλλες αφορά στις μεθόδους επεξεργασίας της πληροφορίας. Για τους σκοπούς της παρούσας ανάλυσης ο επιθετικός προσδιορισμός «υπολογιστική» στην ανθρωπολογία ή την εθνογραφία αναφέρεται κυρίως στον τρόπο επεξεργασίας ή μοντελοποίησης των δεδομένων. Κατά συνέπεια, δεν

αφορά στο αν ένα ερωτηματολόγιο διακινήθηκε ψηφιακά ή αν έγινε tracking η τοποθεσία ενός ατόμου και στη συνέχεια κάποιος μελέτησε το ψηφιακό αποτύπωμα. Πρωτίστως ο όρος αναφέρεται στο εάν τα δεδομένα θα υποστούν υπολογιστική επεξεργασία ώστε να επιτευχθούν οι στόχοι της ανθρωπολογικής ανάλυσης. Με αυτό το σκεπτικό, το αντικείμενο τοποθετείται στη γενικότερη κατηγορία των υπολογιστικών κοινωνικών επιστημών (computational social science) όπου οι υπολογιστές χρησιμοποιούνται για να μοντελοποιήσουν, προσομοιώσουν, αναλύσουν και αναπαραστήσουν κοινωνικά φαινόμενα. Αντίστοιχα, η μοντελοποίηση της μεθόδου συλλογής δεδομένων λαμβάνοντας υπόψη υπολογιστικές προσεγγίσεις θεωρείται επίσης ως υπολογιστική εθνογραφία (Peronakis κ.ά., 2024, σελ. 88). Αξίζει να σημειωθεί ότι ο Fischer, παρά τη συνεχή ενασχόλησή του τις τελευταίες δεκαετίες με τον τομέα της υπολογιστικής κοινωνικής ανθρωπολογίας, αποφεύγει τη χρήση του όρου «computational anthropology». Ωστόσο, αναφέρεται σε «computational anthropologists» που ασχολούνται με μεθόδους βασιζόμενες στην πληροφορική ως μέρος της ερευνητικής διαδικασίας και όχι απλά για τη δημιουργία ενός εργαλείου το οποίο θα διευκολύνει την ερευνητική διαδικασία (M. D. Fischer, 2006, σελ. 12).

Τέλος, υπάρχει ανάγκη να αποσαφηνιστεί ποιο είναι το ευρύτερο επιστημονικό πεδίο μέσα στο οποίο νοηματοδοτείται η υπολογιστική ανθρωπολογία και εθνογραφία. Παρότι σίγουρα απαιτεί διεπιστημονική προσέγγιση και δεν περιορίζεται απολύτως στα όρια ενός επιστημονικού τομέα (boundaries of disciplinary jurisdiction) επικροτούμε την άποψη του Cioffi ο οποίος σημειώνει ότι «οι υπολογιστικές κοινωνικές επιστήμες είναι κοινωνικές επιστήμες. Η ιδέα είναι απολύτως όμοια με την έννοια ότι η υπολογιστική αστρονομία, η υπολογιστική βιολογία και η υπολογιστική γλωσσολογία είναι τμήματα της αστρονομίας, βιολογίας και γλωσσολογίας αντιστοίχως. Κάθε υπολογιστική Χ-επιστήμη είναι, εξ ορισμού, τμήμα αυτής της Χ-επιστήμης» (Cioffi-Revilla, 2016, σελ. 3).

## 4.2 Πρωτογενή δεδομένα και επεξεργασμένα δεδομένα

Όπως έγινε σαφές στην προηγούμενη υποενότητα, ο επιθετικός προσδιορισμός *υπολογιστική* δεν αναφέρεται στον τρόπο πρόσκτησης των δεδομένων αλλά κυρίως στον τρόπο επεξεργασίας ή μοντελοποίησής τους. Σε αυτή την περίπτωση, ποια είναι τα πρωτογενή δεδομένα (raw data) τα οποία θα μοντελοποιηθούν ή θα υποστούν επεξεργασία; Οι Bartlett et al θέτουν τη διάκριση ανάμεσα σε ακατέργαστα πρωτογενή και επεξεργασμένα δεδομένα (processed data) σε πιο εξεζητημένη βάση και κατηγοριοποιούν τα δεδομένα σε δύο βασικές κατηγορίες: τα 'κατασκευασμένα' (crafted) και τα 'ευρισκόμενα' (found). Τα 'κατασκευασμένα' δεδομένα είναι εκείνα τα οποία έχουν παραχθεί εντός μιας επιστημονικής κοινότητας, η οποία θα τα χρησιμοποιήσει για να στηρίξει τη γνώση που παράγει. Με άλλα λόγια πρόκειται για δεδομένα που παράχθηκαν μέσα στο πλαίσιο που καθορίζουν οι επιστήμονες του εκάστοτε επιστημονικού πεδίου και νοηματοδοτούνται εντός αυτού. Από την άλλη πλευρά, τα 'ευρισκόμενα' (found) δεδομένα υπάρχουν ανεξάρτητα από τους σκοπούς και τους σχεδιασμούς της κοινότητας που κάνει την ανάλυση, όπως το ιστορικό καταγραφής κινήσεων (transaction log files) (Bartlett κ.ά., 2018, σελ. 4). Ακολουθώντας την παραπάνω διάκριση, οδηγούμαστε στο προφανές συμπέρασμα ότι τα

‘κατασκευασμένα’/δευτερογενή δεδομένα ενός επιστημονικού πεδίου είναι δυνατό να αποτελούν ‘ευρισκόμενα’/πρωτογενή για κάποιο άλλο.

Σύμφωνα με άλλη θεώρηση, ο Pool ακολουθεί διαφορετικά κριτήρια κατηγοριοποίησης σημειώνοντας ότι τα εθνογραφικά δεδομένα είναι υβριδικά αφού, σύμφωνα με τον ίδιο, χωρίζονται σε soft και hard. Από τη μία περιέχουν αυτολεξεί μεταγραφές (verbatim transcripts) οι οποίες αποτελούν τα ‘hard data’ αλλά περιέχουν και αναμνήσεις και εντυπώσεις (memories and impressions) οι οποίες αποτελούν τα ‘soft data’. Επίσης τονίζει ότι τα hard data μπορούν να αρχειοθετηθούν και αναλυθούν εκ νέου, ενώ τα soft όχι. Ενώ σημειώνει ότι, παρότι είναι εύκολο η μαλακή πλευρά να γίνει σκληρή μέσω αποτύπωσης σε κείμενα και σημειώσεις, πάντα η ερμηνεία της σκληρής πλευράς θα είναι δύσκολη χωρίς τη μαλακή πλευρά (Pool, 2017).

Καθεμία από τις κατηγορίες δεδομένων είναι πιθανό να είναι περισσότερο πρόσφορη στον έναν ή τον άλλον τρόπο επεξεργασίας, αλλά δεν επιβάλλει κάποιον. Για παράδειγμα οι μύθοι είναι τα δεδομένα: ο Levi-Strauss τους άκουσε/διάβασε για να βγάλει συμπεράσματα ενώ ο Weingart προτίμησε να τρέξει ένα πρόγραμμα και να μετρήσει επαναλήψεις συγκεκριμένων λέξεων ώστε να επιβεβαιώσει εμπειρικά υποθέσεις που είχαν αναπτύξει φεμινίστριες ακαδημαϊκοί (Weingart & Jorgensen, 2013). Παρόμοια προσέγγιση με τον Weingart ακολούθησαν και οι Stubbersfield & Tehrani για να μελετήσουν τον μύθο της “Bloody Mary” (Stubbersfield & Tehrani, 2013). Ενώ οι Santucci et al χρησιμοποιούν το Discrete-Event modeling, ώστε μέσω ενός λογισμικού που επιτρέπει τον ορισμό ενός μύθου ως διασύνδεση μυθμάτων<sup>8</sup>, να δημιουργήσουν γράφους οι οποίοι επιτρέπουν την ανάλυση των μύθων βάσει της δομικής (στρουκτουραλιστικής) ανάλυσης (structural analysis) του Claude Lévi-Strauss και να καταδείξουν την μετασχηματιστική μορφοδυναμική των μυθικών δικτύων (transformational morphodynamics of mythical networks) (Santucci κ.ά., 2020).

### 4.3 Επεξεργασία δεδομένων: στατιστική ανάλυση, μοτίβα (patterns) και επισημειώσεις (tagging)

Η παρούσα υποενότητα αφορά στην επεξεργασία των δεδομένων με σκοπό την ανάλυσή τους για την παραγωγή ανθρωπολογικής γνώσης. Οι μέθοδοι αυτές έχουν έναν υποστηρικτικό (assistive) και επικουρικό χαρακτήρα. Δεν υποκαθιστούν τον ερευνητή αλλά του παρέχουν υποστήριξη στη διαδικασία ανάλυσης, ερμηνείας και εξαγωγής συμπερασμάτων. Τα δεδομένα που πρόκειται να υποστούν υπολογιστική επεξεργασία θα πρέπει προφανώς να βρίσκονται σε κάποιου είδους ψηφιακή μορφή. Αυτό σημαίνει είτε ότι τα δεδομένα έχουν γεννηθεί ψηφιακά, δηλαδή παρήχθησαν πρωτογενώς μέσα στο ψηφιακό περιβάλλον, είτε ότι τα δεδομένα ψηφιοποιήθηκαν, δηλαδή υπήρχαν σε κάποια αναλογική μορφή και μετατράπηκαν σε ψηφιακά. Η δεύτερη περίπτωση μπορεί να περιλαμβάνει την ψηφιοποίηση ενός κειμένου ενώ η πρώτη τις αναρτήσεις στο

<sup>8</sup> Σύμφωνα με τον Lévi-Strauss, μυθήματα (mythemes) είναι τα βασικά δομικά στοιχεία ενός μύθου, δηλαδή τα μικρότερα αφηγηματικά κομμάτια που, όταν συνδυάζονται, δημιουργούν τον πλήρη μύθο.

Facebook. Όπως σημειώνουν οι Παπαηλία και Πετρίδης αρκετοί ανθρωπολόγοι αμύνονται πίσω από θεωρητικές και μεθοδολογικές δικαιολογήσεις για να αποφύγουν τη συστηματική εξέταση τέτοιου υλικού (Παπαηλία & Πετρίδης, 2015). Όμως για την παρούσα διατριβή η ενασχόληση ανθρωπολόγων με αυτό το υλικό θεωρείται επιβεβλημένη για έναν πολύ συγκεκριμένο λόγο: αγνοώντας αυτό το υλικό αγνοούμε ένα πολύ σημαντικό μέρος της ανθρώπινης δραστηριότητας.

Η προσέγγιση που επιχειρείται ίσως δημιουργήσει την εντύπωση σε κάποιους αναγνώστες ότι αναφέρεται αμιγώς σε μεθόδους ποσοτικής ανάλυσης. Οι επιθετικοί προσδιορισμοί ποσοτική και ποιοτική θα χρησιμοποιηθούν στην πορεία της ανάλυσης αφού είναι αρκετά σύνηθες κριτήριο διαχωρισμού, όμως ο συγγραφέας της παρούσας διατριβής είναι αρκετά επιφυλακτικός ως προς αυτόν το διαχωρισμό και τείνει να συμφωνήσει με τον Behrens, ο οποίος σημειώνει τα εξής για τη διάκριση μεταξύ ποιοτικής και ποσοτικής ανάλυσης (qualitative and quantitative analysis): αυτή η διάκριση δεν είναι εγγενής στη φύση των ανθρωπολογικών δεδομένων, αλλά στην πραγματικότητα είναι ένα δημιούργημα προηγούμενων μεθοδολογιών οι οποίες απαιτούσαν την ταξινόμηση των ανθρωπολογικών δεδομένων σε βολικές εννοιολογικές κατηγορίες (Behrens, 1990, σελ. 305).

Ένα τυπικό παράδειγμα της χρήσης των μετρήσεων από τους ανθρωπολόγους είναι το κλασικό paper “Culture as Consensus: A Theory of Culture and Informant Accuracy” (Romney κ.ά., 1986) όπου παρουσιάζεται ένα φορμαλιστικό μαθηματικό μοντέλο βάσει του οποίου ο πολιτισμός ορίζεται ως η πλειοψηφούσα άποψη (consensus) και επομένως μπορεί να μετρηθεί. Οι δημιουργοί της εν λόγω δημοσίευσης προτρέπουν σε μια πιθανοκρατική αντιμετώπιση των απόψεων των πληροφορητών βασιζόμενοι στο ότι δεν προσφέρουν όλοι οι πληροφορητές έγκυρη και αξιόπιστη πληροφορία. Για αυτόν τον λόγο ορίζουν συντελεστές βαρύτητας ως προς την εγκυρότητα κάθε πληροφορητή και θεωρούν ότι με τον τρόπο αυτόν το αποτέλεσμα καθίσταται πιο αξιόπιστο. Σύγχρονα λογισμικά δίνουν τη δυνατότητα διαχείρισης παρόμοιων παραμέτρων (Purzycki & Jamieson-Lane, 2017).

Μια από τις βασικές χρήσεις της πληροφορικής στον τομέα των κοινωνικών, αλλά και των ανθρωπιστικών, επιστημών είναι η αναγνώριση μοτίβων ή αναγνώριση προτύπων (pattern recognition). Η εν λόγω διαδικασία αφορά στην αναζήτηση διακριτών κανονικοτήτων στα δεδομένα μέσω στατιστικής ανάλυσης. Η ανάλυση των μοτίβων και οι μετρήσεις βρίσκονται στο επίκεντρο αυτού που η Nelson ονομάζει “Computational Grounded Theory”, δηλαδή υπολογιστική θεμελιωμένη θεωρία (Nelson, 2017). Η Θεμελιωμένη Θεωρία είναι μια μέθοδος η οποία βασίζεται στον επαγωγικό συλλογισμό (inductive reasoning) και μπορεί να οδηγήσει σε συμπεράσματα μέσα από την ανάλυση των δεδομένων χωρίς να είναι υποχρεωτικό να έχει προηγηθεί ο σαφής ορισμός ενός ερευνητικού ερωτήματος, διότι η διατύπωση του ερωτήματος θα προκύψει μέσα από τη διαδικασία ανάλυσης των δεδομένων.

Μια άλλη σημαντική πτυχή της εθνογραφικής έρευνας πεδίου είναι η καταγραφή και η διαχείριση, με υπολογιστικά μέσα, των σημειώσεων πεδίου. Διάφορα εργαλεία λογισμικού έχουν αναπτυχθεί για να βοηθήσουν στην καταγραφή και το χειρισμό του σχετικού υλικού. Η σωστή κωδικοποίηση και ο εμπλουτισμός των αρχικών δεδομένων

με κατάλληλες ετικέτες, χαρακτηριστικά κ.λπ. θα μπορούσαν να δημιουργήσουν αυτό που οι Albris κ.ά. αποκαλούν «υπολογιστικές σημειώσεις πεδίου» (Albris κ.ά., 2021).

Προφανώς η επεξεργασία μεγάλων όγκων δεδομένων είναι ένας πρόσφορος χώρος για υπολογιστική επεξεργασία. Ενδεικτικό παράδειγμα η ανάλυση που έγινε πάνω στα κείμενα από τα Google Books (Michel κ.ά., 2011). Σε μια παρουσίαση αυτού του έργου (project) στα TEDx Talks εκείνοι που υλοποίησαν το έργο λένε: η γνώση του κόσμου έχει αποτυπωθεί σε εκατομμύρια βιβλία και ο καλύτερος τρόπος για να μάθει κάποιος είναι να τα διαβάσει. Αυτό θα ήταν υπέροχο αλλά πρακτικά αδύνατο. Από την άλλη το να χρησιμοποιήσει κάποιος υπολογιστικές μεθόδους για να τα διαβάσει βέβαια θα ήταν τόσο υπέροχο αλλά είναι σίγουρα πρακτικά εφικτό (TEDx Talks, 2011). Αυτήν τη διαδικασία της μη ανθρώπινης ανάγνωσης ο Moretti, στο πλαίσιο της ανάλυσης στα digital humanities, την ονόμασε “distant reading” στο έργο του με τον ομώνυμο τίτλο, το οποίο κατάφερε σε μια πενταετία να γίνει κλασικό (Moretti, 2013).

Η παραπάνω μέθοδος, δηλαδή το distant reading, χρησιμοποιήθηκε και για εθνογραφικά δεδομένα. Όπως ήδη αναφέρθηκε στην ενότητα 4.2 ο Weingart μέτρησε επαναλήψεις συγκεκριμένων λέξεων ώστε να επιβεβαιώσει εμπειρικά υποθέσεις που είχαν αναπτύξει φεμινίστριες/στες ακαδημαϊκοί (Weingart & Jorgensen, 2013) και παρόμοια προσέγγιση ακολούθησαν οι Stubbersfield & Tehrani για να μελετήσουν τον μύθο της ‘Bloody Mary’ (Stubbersfield & Tehrani, 2013). Από την άλλη πλευρά οι Bakharria & Corrin παρουσιάζουν πιο εξελιγμένες μεθόδους natural language processing (NLP) δίνοντας έμφαση στο πόσο μπορεί αυτό να βοηθήσει στη σημασιολογική αναζήτηση (semantic search) -σε αντίθεση με το απλό ταίριασμα συμβολοσειράς (string matching)- εντός ενός σώματος κειμένων (Bakharria & Corrin, 2019).

Πέρα από την ανάλυση των κειμένων, η αναζήτηση μοτίβων μπορεί να αφορά σε διάφορους τύπους υλικού. Για παράδειγμα η Wilf, στο paper “Toward an Anthropology of Computer-Mediated, Algorithmic Forms of Sociality”, δίνει ένα παράδειγμα για το πώς τα μουσικά στυλ μετατρέπονται σε στατιστικά μοτίβα (statistical patterns) (Wilf, 2013). Ενώ οι Leetaru κ.α. αναλύουν 959.000 επιστημονικές δημοσιεύσεις για την Αφρική ώστε να δείξουν γεωγραφικές κατανομές, να μελετήσουν τους Nuer, να βρουν επιδραστικούς συγγραφείς κτλ. (Leetaru κ.ά., 2014). Επίσης ενδιαφέρον παρουσιάζει η πρόσφατη εργασία των Alvard and Carlson, δημοσιευμένη στο Current Anthropology, ένα από τα πιο έγκριτα περιοδικά ανθρωπολογίας, όπου γίνεται η ανάλυση δεδομένων γεωεντοπισμού (GPS). Στην εργασία τους, αναλύοντας μέσω γεωεντοπισμού τις θέσεις των ψαράδων (Artisanal Fishers) στην Κοινοπολιτεία της Δομινίκας (Commonwealth of Dominica), καταλήγουν πως αυτές οι μέθοδοι μπορούν να μειώσουν το κόστος και να εξοικονομήσουν χρόνο μέσω της αυτοματοποιημένης επεξεργασίας των αυξανόμενων συνόλων big data με σκοπό να απαντήσουν σε ανθρωπολογικά ερωτήματα τα οποία δύσκολα θα μπορούσαν να απαντηθούν ακολουθώντας διαφορετική μέθοδο (Alvard & Carlson, 2020). Και οι Sosna κ.α. διερευνούν ταφικά έθιμα εφαρμόζοντας τη θεωρία των γράφων και συμπεραίνουν πως «το κύριο πλεονέκτημα των φορμαλιστικών υπολογιστικών προσεγγίσεων είναι η ικανότητα να αποκαλύψουν και να αξιολογήσουν μοτίβα τα οποία θα ήταν δύσκολο και πολύ χρονοβόρο να ελεγχθούν μόνο μέσω του [ανθρώπινου] συλλογισμού» (Sosna κ.ά., 2013, σελ. 57).

Για τη χρήση αυτού του είδους των τεχνικών οι Abramson κ.α. σχολιάζουν πως παρότι τέτοιες τεχνικές δεν απαλλάσσουν από τη διαδικασία της Ερμηνευτικής, ωστόσο, επιτρέπουν την ερμηνεία, ενώ ταυτόχρονα περιορίζουν την ασάφεια. Παράλληλα τονίζουν πως η χρήση αυτών των μεθόδων δεν βοηθάει μόνο στην ανάλυση των δεδομένων αλλά και στην οπτικοποίηση της ανάλυσης και δίνουν ως παράδειγμα τη χρήση μοτίβων σε “ethnographic heatmaps (ethnoarrays)” (Abramson κ.ά., 2018). Επίσης έμφαση στο πόσο τα μοτίβα συνδράμουν στην οπτικοποίηση εθνογραφικών δεδομένων, για παράδειγμα μέσω γραφημάτων, δίνουν οι Abramson & Dohan στο έργο τους *Beyond Text: Using Arrays to Represent and Analyze Ethnographic Data* (Abramson & Dohan, 2015). Την ίδια μεθοδολογία ακολουθούν και οι Bernstein and Dohan για την ανάλυση και εμφάνιση (visualization) εθνογραφικών δεδομένων από το χώρο της υγείας (ιατρικά κέντρα (medical centers)) (Bernstein & Dohan, 2020).

Για την επίλυση ενός προβλήματος (σαν αυτά που αναφέρθηκαν παραπάνω), μέσω υπολογιστικής επεξεργασίας, υπάρχουν δύο βασικές προσεγγίσεις: η μία αφορά στη δημιουργία ενός ad-hoc προγράμματος για την επίλυσή του και η δεύτερη τη χρήση κάποιου έτοιμου προς χρήση λογισμικού (ready to use) το οποίο ενδείκνυται για τη σχετική επεξεργασία. Όσο πιο πίσω χρονικά τόσο πιο πιθανόν να μην υπήρχε κάποια ‘έτοιμη’ λύση και να έπρεπε να γραφτεί ένα ειδικό λογισμικό (tailored). Σήμερα υπάρχει πλήθος ‘έτοιμων προς χρήση’ λογισμικών των οποίων ο υψηλός βαθμός παραμετροποίησης τα καθιστά αποτελεσματικά για χρήση σε μεγάλο εύρος προβλημάτων επεξεργασίας

Δεδομένης της όχι ιδιαίτερης προτίμησης των ανθρωπολόγων στις ποσοτικές αναλύσεις μια ομάδα λογισμικού που αφορά σε αυτό που ονομάζεται *Computer Assisted Qualitative Data Analysis Software (CAQDAS)* χρησιμοποιείται ευρέως. Ένας επιπλέον λόγος για αυτή τη σχετικά αυξημένη χρήση είναι αυτό που σημειώνουν οι Dohan and Sánchez πως οι τεχνολογίες CAQDA μπορούν να είναι χρήσιμες για τους ερευνητές που βρίσκονται εκτός της θετικιστικής ερευνητικής παράδοσης (Dohan & Sánchez-Jankowski, 1998, σελ. 478). Αυτός ο τύπος λογισμικού διευκολύνει ιδιαίτερα αναλύσεις που ακολουθούν τη λογική της θεμελιωμένης θεωρίας (grounded theory) (O’Kane, 2020). Στην συγκεκριμένη κατηγορία εντάσσεται μεγάλος αριθμός λογισμικού<sup>9</sup> με παρόμοιες λειτουργικότητες<sup>10</sup>. Πολλά εξ αυτών είναι γενικής χρήσης -όπως τα : Atlas TI, MAXQDA, NVivo- και δεν παρότι δεν απευθύνονται συγκεκριμένα στην ανθρωπολογική κοινότητα χρησιμοποιούνται από αυτήν.

Παράλληλα έχουν αναπτυχθεί λογισμικά παρόμοιας λογικής τα οποία στοχεύουν περισσότερο την ανθρωπολογική κοινότητα. Για παράδειγμα το TAMS (ακρωνύμιο του

---

<sup>9</sup> Βεβαίως τα συστήματα CAQDAS μπορούν να ενσωματώσουν τεχνικές μηχανικής μάθησης (machine learning) κυρίως μέσω επιβλεπόμενης μηχανικής μάθησης (supervised machine learning), που σημαίνει ότι υποδεικνύεται στον αλγόριθμο η επιθυμητή συμπεριφορά μέσω παραδειγμάτων.

<sup>10</sup> Σε μια πρόσφατη μελέτη συγκρίνονται οι λειτουργίες από τέσσερα διαφορετικά λογισμικά (ATLAS.ti, MAXQDA, NVivo, QDA Miner) και οι συγγραφείς εντοπίζουν αρκετές διαφορές μεταξύ τους (Noakes κ.ά., 2023)

Text Analysis Markup System) analyzer (Weinstein, 2006) αναπτύχθηκε για εθνογραφική έρευνα, κυρίως για την καταγραφή και τον εντοπισμό θεμάτων σε σημειώσεις πεδίου και συνεντεύξεις. Ο Warters, στην κριτική που του κάνει, εξαίρει τόσο τις δυνατότητες όσο και την ευχρηστία του (Warters, 2005). Όμως ένα από τα πρώτα λογισμικά ήταν το “THE ETHNOGRAPH” (Seidel & Clark, 1984). Ο Read λέει για αυτό “αποσκοπεί στο να επιτρέψει τη χρήση αρχείων κειμένου ως βάσης δεδομένων αντί να κωδικοποιούνται σε αριθμητική μορφή για περαιτέρω ανάλυση. Το EthnoGraph παρακάμπτει τον περιορισμό είτε της ενσωμάτωσης λέξεων-κλειδιών στο κείμενο είτε του περιορισμού σε υπάρχουσες λέξεις κειμένου ως βάση για την ανάκτηση κειμένου, βάζοντας τον χρήστη να κωδικοποιεί πρώτα χειροκίνητα γραμμές κειμένου με έναν όρο ή μια φράση αναγνώρισης και στη συνέχεια να προσθέτει αυτή την κωδική λέξη και τις γραμμές κειμένου στις οποίες εφαρμόζεται στο αρχείο δεδομένων. Τα τμήματα μπορούν να επικαλύπτονται και η ανάκτηση βασίζεται σε συνδυασμούς Boolean των κωδικών των όρων» (D. W. Read, 1990, σελ. 511). Ο Behrens το ονομάζει “textual database-management system” (Behrens & Read, 1993, σελ. 437). Συνεχίζει να διατίθεται και είναι διαθέσιμο στο <http://www.qualisresearch.com/>.

Αναφερόμενος στην ιστορία των CAQDAS ο Bryda σημειώνει « Η φύση της πρακτικής άσκησης ποιοτικής έρευνας είναι η πολυπαραδειγματικότητα (multiparadigmaticity) της, η οποία νομιμοποιεί τη συνύπαρξη διαφορετικών μεθοδολογιών στον τομέα της ποιοτικής έρευνας και των πρακτικών ανάλυσης δεδομένων » (Bryda, 2020, σελ. 77) και καταλήγει «η χρήση των διαφόρων εργαλείων CAQDAS οδηγεί σε αύξηση της μεθοδολογικής επίγνωσης, η οποία μεταφράζεται σε μεγαλύτερη σαφήνεια και ακρίβεια κατά τη διαδικασία της ποιοτικής ανάλυσης δεδομένων» (Bryda, 2020, σελ. 84).

Σε μια λίγο πιο ποσοτική προσέγγιση κινείται το ANTHROPAC. Πρόκειται για ένα αρκετά παλιό λογισμικό η εξέλιξη του οποίου έχει παραμείνει στην εποχή του DOS. Το λογισμικό αναπτύχθηκε για να «ανάλυση δομημένων ποιοτικών και ποσοτικών δεδομένων, συμπεριλαμβανομένων ελεύθερων λιστών, συστοιχιών (pilesorts), τριάδων, ζευγαρωτών συγκρίσεων (paired comparisons) και αξιολογήσεων (ratings)» (Borgatti, 1996). Πιο σύγχρονο λογισμικό με παρόμοιες δυνατότητες είναι το AnthroTools (Purzycki & Jamieson-Lane, 2017). Πρόκειται για λογισμικό το οποίο κάνει elicitation από Free-list data, προσπαθώντας όμως να μειώσει τις διαδικασίες που σχετίζονται με προεπεξεργασία των δεδομένων. Επίσης παρέχει τη δυνατότητα υπολογισμού του cultural consensus που αναφέρθηκε στην αρχή της παρούσας ενότητας.

Μια άλλη κατηγορία λογισμικών είναι εκείνα που αφορούν Social network analysis. Social network analysis is the process of investigating social structures through the use of networks and graph theory. Εδώ σημαντικό ρόλο έχει το Pajek (το οποίο στην προσπάθειά του να επεξεργαστεί και big data) ανακοίνωσε ότι capability to analyze much larger networks is the second major development. “Pajek can now handle networks with nearly one billion vertices. For even larger networks, PajekXXL and Pajek3XL have been developed, which can handle up to two and ten billion vertices respectively.” Το συγκεκριμένο λογισμικό χρησιμοποίησαν οι Sosna et al to model relationships among burials (Sosna κ.ά., 2013) αλλά και οι Lyon and Magliveras used CSAC Kinship Editor and Pajek για δύο διαφορετικά cases studies που ένα αφορά στις σχέσεις Αρβανιτών και

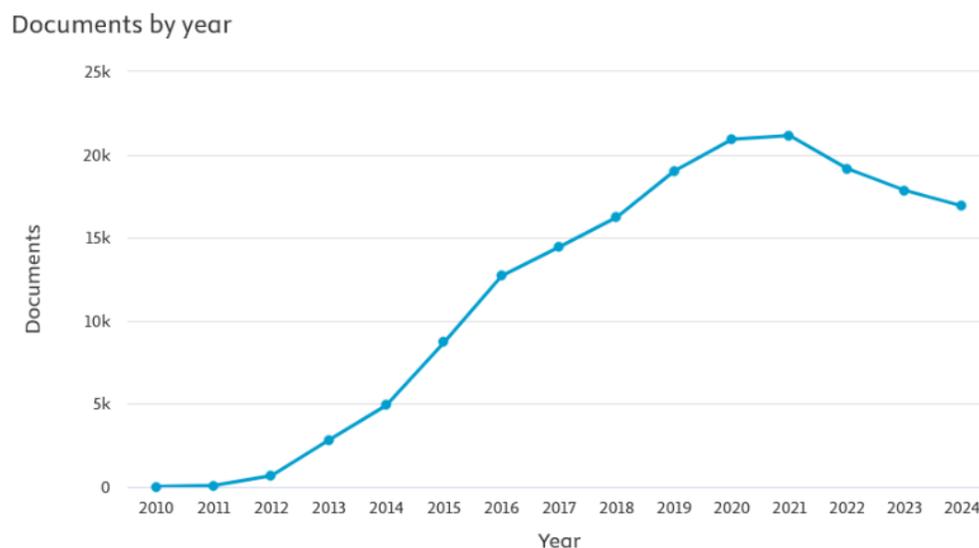
Αλβανών μεταναστών στην Ελλάδα και το άλλο στους Punjab του Pakistan (Lyon & Magliveras, 2006). Επίσης τα ίδια λογισμικά χρησιμοποίησε ο Lyon για την ανάλυση των δικτύων συγγένειας στο Πακιστάν focusing the benefits of formal representations of datasets to both reveal and model certain kinds of sociological phenomena (Lyon, 2013). Επειδή όμως τα δίκτυα συγγένειας αποτελούν τυπικό παράδειγμα όπου μικτές ποιοτικές και ποσοτικές μέθοδοι χρησιμοποιήθηκαν μαζί και σχετικό λογισμικό αναπτύχθηκε με παράλληλη ανάπτυξη σχετικής θεωρίας θα της αφιερώσουμε στη συνέχεια μια ξεχωριστή υποενότητα, την «4.4.1 Συγγένεια και σχετικό λογισμικό».

### 4.3.1 Big data

Στην ψηφιακή εποχή οι άνθρωποι αφήνουν διαρκώς ένα ψηφιακό ίχνος καθώς διαντιδρούν μεταξύ τους ή με το περιβάλλον τους μέσω ψηφιακών συσκευών. Για παράδειγμα ένα smartphone αναφέρει σε τακτά διαστήματα τη θέση του, ένας email server αποθηκεύει όλα τα emails και καταγράφει πότε και από ποια συσκευή διαβάστηκε ή έγινε response κάποιο, ενώ ένα έξυπνο ρολόι (smart watch) ίσως παρακολουθεί το ρυθμό της καρδιάς αυτού που το φοράει. Συγκεντρώνοντας όλες αυτές τις πληροφορίες που μπορεί να προέρχονται από εκατομμύρια ή και δισεκατομμύρια συσκευές δημιουργείται τεράστιος όγκος δεδομένων τα οποία είναι ευρέως γνωστά ως big data. Πρόκειται δηλαδή για υπερβολικά μεγάλα σύνολα δεδομένων (extremely large data sets) των οποίων το μέγεθος και ο βαθμός πολυπλοκότητας καθιστά αδύνατη την επεξεργασία τους από τα παραδοσιακά λογισμικά επεξεργασίας δεδομένων. Τα δεδομένα αυτά είναι ποικιλόμορφα, δομημένα ή αδόμητα, και μπορεί να περιλαμβάνουν κείμενα με τη γνώμη των χρηστών, ηχογραφήσεις ή βίντεο, click streams, website log files κλπ (Chang κ.ά., 2014, σελ. 68). Πολύ συχνά συνδέονται με την ανάλυση της ανθρώπινης συμπεριφοράς και πολλοί ερευνητές τα θεωρούν ως το καλύτερο test bed για προγνωστική αναλυτική (predictive analytics). Έτσι, για να συνεχίσουμε το προηγούμενο παράδειγμα, συνδυάζοντας τις πληροφορίες τού πότε κάποιος διάβασε το ηλεκτρονικό μήνυμα (email) και το πώς η ανάγνωσή του επηρέασε το ρυθμό της καρδιάς του μπορούμε να κρίνουμε αν το μήνυμα του ήταν αδιάφορο ή όχι. Προφανώς εδώ προκύπτουν σοβαρά θέματα ηθικής τόσο σχετικά με την αποθήκευση και διατήρηση αυτών των δεδομένων (Yoon κ.ά., 2025) όσο και γύρω από τη χρήση τους. Μια συζήτηση επί αυτών των θεμάτων γίνεται στη συνέχεια του κειμένου, στην υποενότητα “4.6.2 Επανάχρηση και διαμοιρασμός των δεδομένων”.

Παρότι ο όρος big data είναι σχετικά πρόσφατος, έχει ιδιαίτερο ενδιαφέρον η συμπεριφορά της επιστημονικής κοινότητας γύρω από το θέμα. Σύμφωνα με τη βάση επιστημονικών δημοσιεύσεων Scopus, το έτος 2011 ο όρος εμφανίζεται σε λιγότερες από 100 δημοσιεύσεις, ενώ το 2014 εμφανίζεται σε 4.935 και το 2021 σε 21.121 δημοσιεύσεις και στη συνέχεια να εμφανίσει μια ελαφριά κόπωση, πέφτοντας στις 17.000 το 2024. Η συγκεκριμένη επιστημονική παραγωγή αφορά κυρίως την πληροφορική, αλλά στην ενασχόληση με το θέμα δεν μένουν αμέτοχοι και άλλοι επιστημονικοί τομείς. Σύμφωνα πάντα με τη Scopus, για τα έτη 2010 έως και το έτος 2024 ο συνολικός αριθμός των 175.811 τεκμήρια γύρω από τα big data κατανέμεται ανά επιστημονικό τομέα ως εξής: 31,5% σε Computer Science, 16,2% Engineering, 9,0%

Mathematics, 8.0% Decision Sciences, το **6,4% Social Sciences**, το 4,2% Medicine και το υπόλοιπο κατανέμεται σε άλλους τομείς<sup>11</sup>. Άρα τα big data είναι ένα ζήτημα που δεν απασχολεί μόνο την πληροφορική αλλά πολλούς τομείς της επιστημονικής κοινότητας και οι κοινωνικές επιστήμες διεκδικούν αξιοσημείωτο μερίδιο από την ενασχόληση με το αντικείμενο.

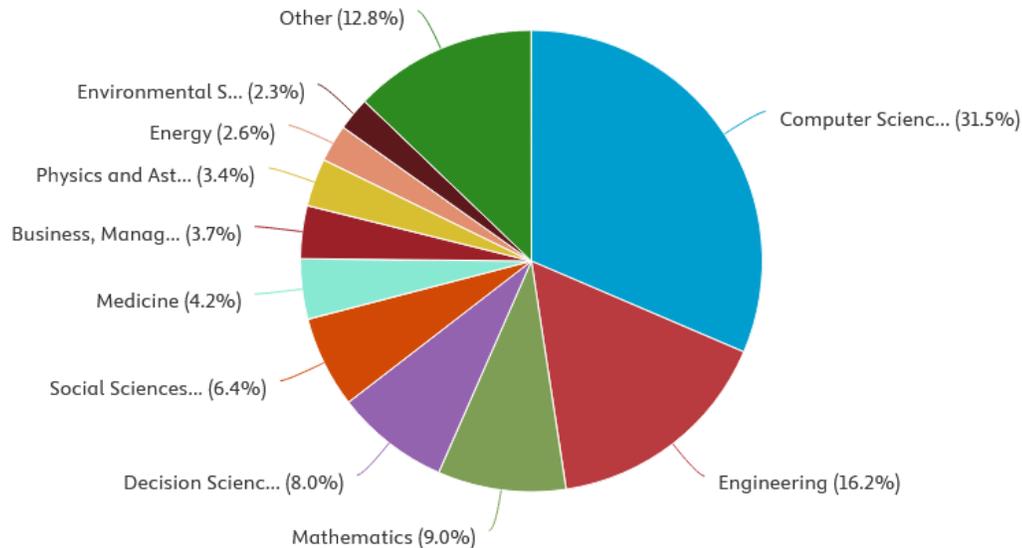


**Γράφημα 1:** η αύξηση των επιστημονικών εργασιών για big data (πηγή Scopus)

---

<sup>11</sup> Η τελευταία αναζήτηση έγινε στη βάση δεδομένων Scopus τον Απρίλιο του 2024 και περιελάμβανε τον όρο big data ως φράση στα πεδία TITLE-ABS-KEY ("big data"). Η κατανομή σε έτη, και η κατ' επέκταση κατανομή σε επιστημονικούς τομείς, έγινε μέσω της διεπαφής της εν λόγω βάσης δεδομένων.

## Documents by subject area



**Γράφημα 2:** κατανομή, ανά επιστημονικό τομέα, του όρου *big data* (πηγή Scopus)

Ο Schroeder ισχυρίζεται ότι τα *big data* έχουν αναζωπυρώσει μια σειρά θεμελιωδών ζητημάτων για τις κοινωνικές επιστήμες, όπως τις ηθικές και κοινωνικές επιπτώσεις από τη χρήση αυτών των δεδομένων, το ρόλο της στατιστικής και της αιτιώδους εξήγησης (*causal explanation*), το πώς τα σύνολα δεδομένων υποστηρίζουν την εγκυρότητα της γνώσης ή ακόμη γενικότερα το πώς η γνώση συγκεντρώνεται. Ωστόσο, παρατηρεί πως έχει δοθεί περιορισμένη προσοχή στο πώς τα *big data* μετασχηματίζουν τη φύση της γνώσης στις κοινωνικές επιστήμες, παρότι αυτά, σε συνδυασμό με άλλα στοιχεία, όπως νέες προσεγγίσεις σχετικά με την αιτιοκρατία, οδηγούν τις κοινωνικές επιστήμες πιο κοντά στο είδος συγκέντρωσης της γνώσης που ταιριάζει στις θετικές επιστήμες (Schroeder, 2019). Οι Cukier & Mayer-Schoenberger ονόμασαν “*datafication*” αυτήν τη διαδικασία της ποσοτικοποίησης προσεγγίσεων μέσω των δεδομένων οι οποίες ποτέ κατά το παρελθόν δεν είχαν ποσοτικοποιηθεί (Cukier & Mayer-Schoenberger, 2013). Σε αυτήν τη συνθήκη εμπίπτει ευρύ φάσμα δεδομένων που εκτείνεται από τα *likes* και *dislikes* στα κοινωνικά δίκτυα μέχρι τα δεδομένα που αφορούν την κυκλοφοριακή συμφόρηση. Μέσα σε αυτό το πλαίσιο, κάποιοι ερευνητές δεν αντιμετωπίζουν τα δεδομένα απλά ως ένα υποπροϊόν της πληροφορικής αλλά σαν μία μηχανή που επιτελεί μια δυναμική αλλαγή της κοινωνίας (Acker, 2015, σελ. 70). Και ειδικά για το περιβάλλον των κοινωνικών επιστημών, αντλώντας από τον χώρο της αναλυτικής κοινωνιολογίας (*analytical sociology*), οι Keuschnigg κ.α. σημειώνουν ότι δεδομένης της τρέχουσας κατάστασης της κοινωνικής θεωρίας θα ήταν σοβαρό λάθος να αγνοήσουμε αυτά τα εργαλεία, για διερευνητικές αναλύσεις (*explorative analyses*) με βάση τα δεδομένα, όταν επιδιώκουμε να κατανοήσουμε καλύτερα την κοινωνική δυναμική (Keuschnigg κ.ά., 2017). Ενώ σύμφωνα με τον Romele αυτές οι προσεγγίσεις δεν αλλάζουν μόνο την

αλληλεπίδραση με τον κόσμο αλλά ολόκληρη την οπτική μας για τον κόσμο (Romele, 2020).

Το ψηφιακό αποτύπωμα που αφήνουν οι άνθρωποι είναι το ιστορικό των εκφρασμένων σκέψεων και δράσεών τους. Από πλευράς των ανθρωπολόγων, από τους πρώτους που κατέδειξαν την ανάγκη για μελέτη αυτού του συνόλου του ψηφιακού αποτυπώματος, ήτοι των big data, ήταν ο Curran στο άρθρο του “Big Data or ‘Big Ethnographic Data’? Positioning Big Data within the ethnographic space” (Curran, 2013). Προφανώς το ότι το αποτύπωμα είναι ήδη ψηφιακό διευκολύνει την υπολογιστική του επεξεργασία. Άρα για όσους δεν συμπλέκουν αδιάρρηκτα τη μέθοδο (συμμετοχική παρατήρηση εν προκειμένω) με το αντικείμενο της έρευνας, ο ισχυρισμός των Charles & Gherman πως «παρότι θεωρούνται ανταγωνιστές, η εθνογραφία και η ανάλυση των big data έχουν από πολλές πλευρές έναν κοινό σκοπό» (Charles & Gherman, 2018) μάλλον δείχνει προς τη σωστή κατεύθυνση. Προς αυτήν συνηγορεί και το άρθρο του Beuving “Ethnography’s future in the big data era” (Beuving, 2020) ο οποίος στηλιτεύει τον τρόπο που τα big data ερμηνεύονται από τους αναλυτές δεδομένων (data analysts).

Είναι κοινώς παραδεκτό στην ανθρωπολογία πως «υπάρχει συστηματική απόκλιση ανάμεσα σε εκείνα που λένε οι άνθρωποι για όσα κάνουν, σε αυτά που πραγματικά κάνουν και σε εκείνα που σκέφτονται» (Kuper, 1994, σελ. 55). Χαρακτηριστικό επί του θέματος είναι το παράδειγμα που παραθέτει η Decoteau για την ασυνέπεια μεταξύ δηλωμένων θρησκευτικών απόψεων και θρησκευτικών πρακτικών, χαρακτηρίζοντας τις πράξεις ως πιο σύνθετες από τις εκφρασμένες απόψεις. Σε αυτό το παράδειγμα ο σύζυγος γυναίκας που νοσούσε δήλωνε ότι η καλύτερη αντιμετώπιση της ασθένειας ήταν μέσω φαρμάκων της δυτικής ιατρικής αποκηρύσσοντας μεθόδους της παραδοσιακής ιατρικής. Ωστόσο, για όσο διάστημα η σύζυγός του εξακολουθούσε να ασθενεί εκείνος επισκεπτόταν έναν παραδοσιακό θεραπευτή (sangoma) (Decoteau, 2017, σελ. 56). Ενώ και οι Jürgens κ.α. επιβεβαιώνουν αυτή την προσέγγιση με εμπειρικά δεδομένα από τα social media σε πρόσφατη μελέτη τους (Jürgens κ.ά., 2019). Τα big data είναι ιδανικός χώρος για να τεκμηριωθεί κατά πόσο αυτά που λένε οι άνθρωποι συμβαδίζουν με αυτά που κάνουν (Lazer & Radford, 2017, σελ. 23), διότι τα big data περιέχουν δεδομένα τα οποία πολλές φορές ανατρέπουν αυτά που τα υποκείμενα θεωρούν ότι ισχύουν.

Όμως απαιτείται ιδιαίτερη προσοχή κατά την ερμηνεία των δεδομένων, διότι, όπως πολύ εύστοχα παρατηρεί ο Acker, ένα απόσπασμα των δεδομένων χωρίς τα συμφραζόμενά του (context) είναι άνευ νοήματος (Acker, 2015, σελ. 70). Πολλές φορές η απολυτότητα των μετρήσεων παραπλανεί, οδηγώντας προς την υποστήριξη απόλυτων ερμηνειών. Ενδεικτικό παράδειγμα για τη σημασία των συμφραζομένων στην ερμηνεία των αριθμών είναι η φωτογραφία του Instagram με τα περισσότερα likes. Εφόσον πρόκειται για τη φωτογραφία με τα περισσότερα likes, κάποιος μπορεί να υποθέσει ότι απεικονίζει κάτι εξαιρετικό. Όμως η φωτογραφία εικονίζει απλά ένα αυγό! Πρόκειται για τη φωτογραφία που δημοσιεύτηκε στο λογαριασμό “world\_record\_egg” και είναι διαθέσιμη στο σύνδεσμο: <https://www.instagram.com/p/BsOGulcndj-/>. Η εν λόγω φωτογραφία είχε ως μόνο στόχο να υπερβεί την πρώτη σε likes φωτογραφία του Instagram, ήτοι μία φωτογραφία της Kylie Jenner η οποία είχε πάνω από 18 εκατομμύρια

likes. Πέτυχε να την ξεπεράσει μέσα σε μερικές μέρες. Σε λιγότερο από δύο εβδομάδες είχε ξεπεράσει τα 35 εκατομμύρια likes και τη στιγμή που πληκτρολογούνται αυτές οι γραμμές τα likes είχαν ξεπεράσει τα 60 εκατομμύρια<sup>12</sup>. Άρα, όπως σημειώνουν οι Dourish & Gómez Cruz, «Τα δεδομένα δεν μιλούν από μόνα τους. Τα δεδομένα πρέπει να αφηγηθούν -να χρησιμοποιηθούν σε συγκεκριμένα συμφραζόμενα, να ενταχθούν σε αφηγήσεις που τους δίνουν σχήμα και νόημα και να κινητοποιηθούν ως μέρος ευρύτερων διαδικασιών ερμηνείας και νοηματοδότησης» (Dourish & Gómez Cruz, 2018, σελ. 1).

#### 4.4 Απευθείας συνεπαγωγή από τα δεδομένα: το παράδειγμα της Prolog

Τα όσα αναφέρθηκαν παραπάνω αφορούν σε μια ημι-αυτόματη διαδικασία. Δηλαδή ο εκάστοτε ερευνητής χρησιμοποιεί το λογισμικό ως εργαλείο για την επιβεβαίωση ή διάψευση υποθέσεων. Με άλλα λόγια ο ανθρώπινος συλλογισμός επικουρείται από την επεξεργασία των δεδομένων μέσω του υπολογιστή. Όμως οι υπολογιστές μπορούν να λειτουργήσουν και ένα πιο αφαιρετικό και αυτόνομο επίπεδο από αυτό. Το πλέον ενδεικτικό παράδειγμα αυτής της δυνατότητας, στον τομέα της ανθρωπολογίας, είναι η χρήση της Prolog.

Η Prolog είναι μια γλώσσα λογικού προγραμματισμού που συνδέεται άρρηκτα με την τεχνητή νοημοσύνη, διότι μπορεί να χρησιμοποιηθεί ως εργαλείο συλλογισμού που οδηγεί σε συμπεράσματα. Το όνομά της προέρχεται από το Programming Logic.

Η Prolog επιτρέπει τη δημιουργία δηλώσεων με την μορφή γεγονότων (facts). Τα γεγονότα στην Prolog περιγράφουν τις ιδιότητες των αντικειμένων ή τις σχέσεις μεταξύ των αντικειμένων. Παράδειγμα ενός γεγονότος είναι η πρόταση «Ο Μέγας Αλέξανδρος ήταν παιδί του Φίλιππου Β'». Ένα σύνολο τέτοιων γεγονότων συνιστά μια βάση γνώσης στην Prolog. Ο σημαντικός μηχανισμός για τη συλλογιστική επί των γεγονότων που αποτελούν τη βάση γνώσης είναι η δημιουργία κανόνων. Οι κανόνες μπορούν να δηλώσουν συνεπαγόμενες (implicit) σχέσεις. Για παράδειγμα μπορεί να δηλωθεί ότι αν το A είναι παιδί για το B, τότε το B είναι γονιός για το A. Άρα οι κανόνες εκφράζουν ένα νέο σύνθετο γεγονός, που ισχύει όταν ισχύουν συγχρόνως ένα ή περισσότερα γεγονότα.

Οι επερωτήσεις (queries) χρησιμοποιούνται για τη λήψη απαντήσεων σχετικά με τις σχέσεις μεταξύ των αντικειμένων βάσει των δηλωθέντων γεγονότων και τους κανόνες που τα διέπουν. Για παράδειγμα, αν ισχύουν τα δύο γεγονότα της προηγούμενης παραγράφου, δηλαδή ο Ο Μέγας Αλέξανδρος ήταν παιδί του Φίλιππου Β' και ο κανόνας αν το A είναι παιδί για το B, τότε το B είναι γονιός για το A, το σύστημα μπορεί να συμπεράνει ότι ο Φίλιππου Β' ήταν γονιός του Μέγα Αλέξανδρου. Παρότι το παράδειγμα είναι απλουστευτικό είναι σαφές ότι το σύστημα παράγει ένα νέο γεγονός (το ότι ο Φίλιππος Β' ήταν γονιός του Μέγα Αλέξανδρου) το οποίο δεν ήταν ρητώς δηλωμένο στη

---

<sup>12</sup> Η Wikipedia φιλοξενεί σχετικό άρθρο όπου περιγράφεται η ιστορία της εν λόγω φωτογραφίας [https://en.wikipedia.org/wiki/Instagram\\_egg](https://en.wikipedia.org/wiki/Instagram_egg).

βάση γνώσης αλλά προέκυψε από λογική συνεπαγωγή η οποία συνδυάζει τα δεδομένων των απλών γεγονότων με εκείνα των κανόνων. Πάνω σε αυτή τη νέα γνώση, που προκύπτει λογικά από τον κανόνα, είναι δυνατόν να 'χτιστεί' ένας πιο περίπλοκος κανόνας του τύπου: ο γονιός που είναι άνδρας είναι πατέρας.

Μια σύντομη εισαγωγή για τη χρήση της Prolog (ειδικά για τη χρήση της στη συγγένεια) έχει γράψει ο γνωστός για την ενασχόλησή του με θέματα υπολογιστής ανθρωπολογίας Michael D. Fischer στο "Representing Anthropological Knowledge: Calculating Kinship" (M. D. Fischer, χ.χ.).

Ακολουθεί ένα μικρό παράδειγμα κώδικα, με λογικές συνεπαγωγές, από το χώρο της συγγένειας. Στο Απόσπασμα κώδικα 1 παρουσιάζεται μια απλουστευμένη περίπτωση συλλογισμού σε Prolog που περιλαμβάνει μερικά μέλη της οικογένειας του συγγραφέα. Τα γεγονότα αποτελούνται από πέντε δηλώσεις όπου αναφέρονται: τα φύλα τριών ατόμων και οι δηλώσεις ποιου/οιας παιδί είναι ο «manolis»<sup>13</sup>. Τα παραπάνω γεγονότα (facts) είναι τα δεδομένα. Ακολουθούν οι κανόνες όπου δηλώνεται ότι α) αν για ο Χ είναι παιδί του Ψ τότε ο Ψ είναι γονιός για τον Χ, β) αν κάποιος γονιός έχει φύλο male τότε είναι πατέρας και γ) αν κάποιος γονιός έχει φύλο female τότε είναι μητέρα. Στις τελευταίες γραμμές του κώδικα, με τη μορφή Prolog σχολίων, υπάρχουν επερωτήσεις που ανήκουν και στις δύο κατηγορίες που αναφέρθηκαν πριν. Η πρώτη αφορά σε απάντηση τύπου αληθές/ψευδές και οι επόμενες επιστρέφουν αποτελέσματα τα οποία δεν είναι ρητώς δηλωμένα στη βάση γνώσης αλλά προκύπτουν από τα όσα είναι ρητώς δηλωμένα, συνδυάζοντας τα απλά γεγονότα με τους κανόνες. Για παράδειγμα, η τελευταία επερώτηση «mother(X, manolis).» υπολογίζει ότι η μητέρα του «manolis» είναι η «eleni» λαμβάνοντας υπόψη τα απλά γεγονότα και τους κανόνες που καταγράφουν ποιοι είναι γονείς και στη συνέχεια ποιοι από τους γονείς είναι μητέρες, για να καταλήξει στη μητέρα του Μανόλη.

---

<sup>13</sup> Τα ονόματα των ατόμων δεν ξεκινάνε με κεφαλαίο γράμμα διότι αυτό θα σήμαινε, βάσει του συντακτικού της Prolog, ότι πρόκειται για μεταβλητές.

### Απόσπασμα κώδικα 1: παράδειγμα συνεπαγωγής σε Prolog

```

% Γεγονότα
male(manolis).
male(alexandros).
female(eleni).
female(xrysa).
child(manolis, alexandros).
child(manolis, eleni).
child(xrysa, alexandros).
child(xrysa, eleni).

% Κανόνες
parent(X, Y) :- child(Y, X).
father(X, Y) :- parent(X, Y), male(X).
mother(X, Y) :- parent(X, Y), female(X).

% Παραδείγματα ερωτήσεων
% male(manolis) -> απάντηση τύπου αληθές/ψευδές (true/false)
% parent(X, Y). -> απάντηση: τα ζευγάρια γονιού-παιδιού
% father(X, Y). -> απάντηση: τα ζευγάρια γονιού-πατέρα
% mother(X, manolis). -> απάντηση: η μητέρα του manolis

```

Είναι σχεδόν αυτονόητο ότι ο φορμαλισμός των τύπων της συγγένειας θα ήταν ένα πρόσφορο πεδίο για την εφαρμογή της Prolog και για αυτό το λόγο χρησιμοποιήθηκε, και χρησιμοποιείται, εκτενώς στο εν λόγω θεματικό πεδίο. Στην επόμενη υπο-ενότητα θα εξετασθεί η συγγένεια ως πεδίο της υπολογιστικής εθνογραφίας και ανθρωπολογίας.

#### 4.4.1 Συγγένεια και σχετικό λογισμικό

Η συγγένεια αποτέλεσε ένα σημαντικό πεδίο έρευνας για τους κοινωνικούς ανθρωπολόγους και αναπτύχθηκαν αρκετές θεωρητικές προσεγγίσεις για τη μελέτη της (Τουνταςάκη, 2008). Στη δεκαετία του 1970 υπήρξε μια εκτενής κριτική στη σημασία που της απέδιδαν και ένα από τα κλασικά κείμενα που αμφισβήτησαν τη συγγένεια ως αναλυτική κατηγορία με καθολική ισχύ ήταν το «Remarks on the analysis of kinship and marriage» που εξέδωσε ο Needham το 1971 (Needham, 2004). Ωστόσο, από τα μέσα της δεκαετίας του 1990, η συγγένεια θα ανακτήσει τη θέση της και ακόμη και κάποιοι από

όσους την επέκριναν ως αναλυτικό εργαλείο, όπως ο Schneider, θα δηλώσουν ότι η συγγένεια «αναγεννήθηκε μέσα από τις στάχτες της» (Schneider, 1995, σελ. 193)<sup>14</sup>.

Ο Fischer σημειώνει ότι οι εφαρμογές λογισμικού που σχετίζονται με τη συγγένεια αποτελούν τις πρώτες προσπάθειες των ανθρωπολόγων να χρησιμοποιήσουν τους υπολογιστές στην έρευνα και ότι οι ανθρωπολόγοι έχουν δείξει σημαντικό ενδιαφέρον για τη χρήση των υπολογιστών για την ανάλυση συγγενικών και γενεαλογικών δεδομένων. Ακριβώς όπως η συγγένεια είναι ένας από τους πιο πολύπλοκους τομείς της κοινωνικής ανθρωπολογίας, οι υπολογιστικοί πόροι που απαιτούνται για να βοηθήσουν στην ανάλυση των δεδομένων συγγένειας είναι σχετικά πολύπλοκοι και ποικίλοι (M. D. Fischer, 1994, σελ. 131). Όμως όπως παρατηρούν οι Lyon και Magliveras οι ανθρωπολόγοι που ασχολούνται με τη συγγένεια έχουν σε μεγάλο βαθμό απογοητευτεί από τα εμπορικά γενεαλογικά λογισμικά, επειδή όλα υποθέτουν ότι η πυρηνική οικογένεια είναι η κεντρική μονάδα. (Lyon & Magliveras, 2006, σελ. 31). Έτσι, οι ανθρωπολόγοι χρειάζονται λογισμικό (και σχετική θεωρία) όχι για τη δημιουργία γενεαλογικών δέντρων αλλά για την ανάλυση της συγγένειας.

Πρωτεργάτες επί του θέματος, με συστηματική δουλειά δεκαετιών, είναι οι Dwight Read και Michael Fischer οι οποίοι συχνά συν-δημοσιεύουν. Ο Read στα κείμενά του για την υπολογιστική επεξεργασία της συγγένειας [ενδεικτικά (D. W. Read & Behrens, 1990) (D. W. Read, 2001), (D. W. Read, 2007), (D. Read κ.ά., 2013)] αναφέρεται διεξοδικά σε ζητήματα φορμαλιστικής ανάλυσης της συγγένειας. Κεντρικό ρόλο διαδραματίζει το Kinship Algebra Expert System (KAES) το οποίο παρέχει ένα γραφικό πλαίσιο για την κατασκευή, αν είναι δυνατόν, ενός γενετικού (generative) αλγεβρικού μοντέλου για τη δομή μιας συγγενικής ορολογίας (kinship terminology = οι όροι που χρησιμοποιούνται για να αναφερθούμε στους συγγενείς) με στόχο να συσχετίσει τις ιδιότητες και τη δομή των συγγενικών ορολογιών (kinship terminologies) με μια υποκείμενη λογική την οποία το πρόγραμμα KAES βοηθά να αποκαλύψει και να μοντελοποιήσει ως γενετική (generative) δομή (D. W. Read, 2006, σελ. 43). Όπως εξηγούν οι δημιουργοί του «το KAES είναι κάτι πολύ περισσότερο από ένα λογισμικό. Στην περίπτωση αυτή, το λογισμικό έχει γίνει το μέσο μέσω του οποίου εκφράζεται μια θεωρία, η οποία γίνεται εμφανής στον χρήστη μέσω της ανάλυσης μιας συγγενικής ορολογίας (kinship terminology)» (Behrens & Read, 1993, σελ. 445). Αυτό μετεξελίχθηκε στο Kinship Algebra Modeler (KAM) που είναι μια σουίτα εργαλείων για την υποστήριξη της ανάδειξης και ανάλυσης των ορολογιών συγγένειας και το οποίο προήλθε από ένα έργο που αναπτύχθηκε αρχικά από τους Dwight Read και Cliff Behrens (1990) ως Kinship Algebra Expert System (KAES) και ξαναγράφηκε από το 1999 από τους Dwight Read και Michael Fischer, οδηγώντας στο KAM. Το KAM, το οποίο ενσωματώνει το KAES, είναι μια σουίτα εφαρμογών σε java, για την υποστήριξη της άντλησης (elicitation) και ανάλυσης των ορολογιών συγγένειας (D. Read κ.ά., 2013, σελ. 16).

---

<sup>14</sup> «First, until recently, kinship had ceased being a major popular subject in anthropology. Papers on kinship clearly fell off. They became fairly rare. Now, of course, phoenix-like, it's risen from its ashes».

Μελετητές επισημαίνουν ότι ο φορμαλισμός παίζει κεντρικό ρόλο στην έκφραση των ιδιοτήτων και των σχέσεων που προκύπτουν για μια ορολογία συγγένειας μέσω της διαδικασίας άντλησης (elicitation). Οι προϋποθέσεις για μια ορολογία συγγένειας είναι οι συγγενικοί όροι που αντιστοιχούν στις πολιτισμικά αναγνωρισμένες πρωτογενείς καταστάσεις συγγένειας. Το νόημα των συγγενικών όρων που συνθέτουν μια ορολογία συγγένειας καθορίζεται στη συνέχεια μέσω των διασυνδέσεων που αφορούν τους πρωτογενείς συγγενικούς όρους και όχι από γενεαλογικά δηλωτικά (genealogical denotata). Έτσι, οι ορολογίες συγγένειας μοιάζουν με γεωμετρικές και όχι με σύνολα λέξεων, δεν είναι συλλογές λέξεων αλλά συστήματα ιδεών. Και ο καλύτερος τρόπος για να συλλάβει κανείς την εννοιολογική δομή τέτοιων συστημάτων, χωρίς να χρησιμοποιεί τις δικές του πολιτισμικές αντιλήψεις ως επισκιάζον φίλτρο, είναι να ζητήσει διαγράμματα και όχι λίστες. Έτσι, για παράδειγμα, μια γραφική απεικόνιση ανάμεσα στο αγγλικό σύστημα συγγένειας και εκείνο των Punjabi δείχνει ότι τα σχήματα δεν είναι τα ίδια. Τα αγγλικά μοιάζουν με χριστουγεννιάτικο δέντρο και τα Punjabi με πεταλούδα. (D. Read κ.ά., 2013). Όμως, δυστυχώς, όπως σημειώνουν οι Carlson and Anderson, οι ανθρωπολόγοι χρησιμοποιούν πολύ λίγο τα διαγράμματα για να μοντελοποιήσουν τις κοινωνικές διαδικασίες που μελετούν (Carlson & Anderson, 2007, σελ. 642).

Βασικό συστατικό για την υλοποίηση λογισμικού που μοντελοποιεί τη συγγένεια είναι η γλώσσα λογικού Prolog στην οποία αναφερθήκαμε πριν και η οποία θεωρείται, από πολλούς ερευνητές, ως το καταλληλότερο περιβάλλον για την ανάλυση συγγένειας. (ενδεικτικά: (M. D. Fischer, 1994, Κεφάλαια 6–7), (D. W. Read & Behrens, 1990) (D. W. Read, 2006)).

Κλείνοντας, πρέπει να τονισθεί πώς οι φορμαλιστικές πτυχές των ορολογιών συγγένειας, ιδίως οι αλγεβρικές τους ιδιότητες, λειτουργούν ως περιορισμός για τα είδη των σχέσεων που εκφράζονται, αλλά δεν αποτελούν επαρκές εργαλείο πρόβλεψης για τον τρόπο με τον οποίο χρησιμοποιούνται ή διατηρούνται αυτές οι σχέσεις (Lyon, 2013, σελ. 47).

## 4.5 Αναζήτηση πληροφορίας

Η ενότητα αυτή αναφέρεται σε επεξεργασμένα δεδομένα, άρα σε δομημένη πληροφορία και μάλιστα με τη λογική των crafted data (βλέπε ενότητα 4.2), τα οποία όμως έχουν κατασκευαστεί για συγκεκριμένη χρήση. Όμως παρότι τα crafted data, σε αντίθεση με τα found, έχουν ήδη μια συγκεκριμένη δομή που εξυπηρετεί τους σκοπούς δημιουργίας τους πολλές φορές η δομή αυτή καθιστά δυνατή και τη χρήση τους σε άλλα περιβάλλοντα. Για παράδειγμα οι Novak and Harris χρησιμοποιούν δεδομένα που δημιουργήθηκαν για την ιατρική παρακολούθηση ασθενών εντός του συστήματος υγείας, ως εμπειρικά δεδομένα για ανθρωπολογική χρήση (Novak & Harris, 2020). Άρα, παρότι, ο αρχικός σκοπός δημιουργίας αυτός των δεδομένων είναι συγκεκριμένος δεν μπορούν να προεξοφληθούν όλες οι πιθανές χρήσεις τους.

Μολονότι η επόμενη ενότητα [4.6] εστιάζει στη μοντελοποίηση των δεδομένων, εδώ θα γίνει αναφορά στην ήδη, από τη δεκαετία του 1980, καθιερωμένη μοντελοποίηση των βάσεων δεδομένων ως την επιτομή αυτού που ονομάσαμε ως

crafted data. Η κυριαρχία του Μοντέλου Οντοτήτων-συσχετίσεων (Entity-Relationship Model) στο σχεδιασμό Σχισιακών Βάσεων Δεδομένων (Relational Databases) και η προαγωγή σε πρότυπο (standard) της Structured Query Language (SQL) οδήγησε στην εκρηκτική άνοδο των Relational Database Management System (RDBMS) με τεράστιο όγκο δεδομένων, από πάρα πολλούς τομείς και για πάρα πολλές χρήσεις. Τα όσα θα ακολουθήσουν δεν περιορίζονται μόνο στις σχεσιακές βάσεις δεδομένων, διότι κάποιες από τις προσπάθειες αφορούν και άλλου είδους υλοποιήσεις οι οποίες περιλαμβάνουν γεωχωρικές απεικονίσεις ή άλλου τύπου δεδομένα και μοντελοποιήσεις. Παρόλα αυτά τη μερίδα του λέοντος κατέχουν οι υλοποιήσεις σε σχεσιακές βάσεις δεδομένων. Προφανώς η εκάστοτε μοντελοποίηση των δεδομένων θα καθορίσει κατά πόσον μπορούν να είναι αποτελεσματικά για συγκεκριμένου τύπου αναζητήσεις και συγκεκριμένου βαθμού εξειδικεύσεις.

#### 4.5.1 Βάσεις δεδομένων για Εθνογραφία και Ανθρωπολογία

Ζητήματα που σχετίζονται με τη μοντελοποίηση της καταγραφής, της αναζήτησης και της σύγκρισης εθνογραφικών δεδομένων έχουν τεθεί από τα μέσα του προηγούμενου αιώνα. Στα τέλη του ο καταλληλότερος μηχανισμός που είχε αναπτυχθεί ήταν οι βάσεις δεδομένων. Οι βάσεις δεδομένων, γενικά, θεωρούνται πρόσφορες για διαπολιτισμική (cross-cultural) ή συγκριτική ανθρωπολογία (Schmidt & Schweizer, 1990). Περιπτώσεις όπου υπάρχουν λίστες -σε απλή μορφή κειμένου- που καταγράφουν σύνολα δεδομένων και δεν ενσωματώνουν κάποιον μηχανισμό αναζήτησης, δεν περιλαμβάνονται παρακάτω<sup>15</sup>.

Η πιο εμβληματική βάση δεδομένων για την Ανθρωπολογία είναι το Human Relations Area Files (HRAF) (<https://hraf.yale.edu/>). Ξεκίνησε στο τέλος της δεκαετίας του 1940 ως ένα διεπιστημονικό ερευνητικό εγχείρημα με σκοπό την καταγραφή πολιτισμικών δεδομένων από το Πανεπιστήμιο του Yale. Ο βασικός του στόχος του ήταν (και εξακολουθεί να είναι) η εφαρμοσμένη συγκριτική ανθρωπολογία. Προφανώς η πρώτη του εκδοχή, το μακρινό 1949, ουδεμία σχέση είχε με υπολογιστές. Ξεκίνησε με την αποδελτίωση (σε φυσικά δελτία τα οποία προσομοιάζαν με εκείνα των δελτιοκαταλόγων των βιβλιοθηκών) καταγράφοντας την πληροφορία που αφορούσε συγκεκριμένου τύπου θέματα και παραπέμποντας σε συγκεκριμένες σελίδες εντός των αντίστοιχων τεκμηρίων. Την δεκαετία του 1990 αυτοματοποιήθηκε πλήρως και τον 21<sup>ο</sup> αιώνα επεκτάθηκε με το eHRAF Archaeology το οποίο εστιάζει σε αρχαιολογικά δεδομένα. Στο τέλος του 2018, οι δύο βάσεις δεδομένων του HRAF περιέχουν σχεδόν τρία εκατομμύρια εγγραφές σε επίπεδο παραγράφου, οι οποίες προέρχονται από

<sup>15</sup> Παράδειγμα τέτοιας λίστας είναι το «Μητρώο ανθρωπολογικών σημειώσεων και χειρογράφων σε αρχαιακά αποθετήρια» (Registry of Anthropological Data) που είναι ένας διαρκώς ανανεούμενος πόρος που περιγράφει τη θέση περισσότερων από 850 συλλογών εθνογραφικών σημειώσεων πεδίου και χειρογράφων σε αρχεία που φιλοξενούνται σε διάφορους φορείς. Η λίστα, η οποία περιγράφει περισσότερες από 650 συλλογές, διατίθεται στο <http://copar.org/fieldnotes.htm> και αλλού. Άλλο παράδειγμα είναι το <http://www.digitalhimalaya.com/> το οποίο σχεδιάστηκε για την αρχειοθέτηση και την διάθεση εθνογραφικού υλικού που σχετίζεται με την περιοχή των Ιμαλαίων.

περισσότερα από 8.000 τεκμήρια τα οποία περιγράφουν πάνω από 400 κοινωνίες και αρχαιολογικές παραδόσεις (M. D. Fischer & Ember, 2018, σελ. 323). Ο Tobin, σε μια δημοσίευσή του στο περιοδικό *Cultural Anthropology* εξαίρει τη σημασία του και το χαρακτηρίζει ως ριζοσπαστικό (radical text) κείμενο (Tobin, 1990). Η βάση δεδομένων είναι προσβάσιμη στο διαδίκτυο αλλά απαιτείται συνδρομή.

Η έτερη ιστορική προσπάθεια έγινε από τον George Peter Murdock -ο οποίος ήταν και ο πρωτεργάτης της δημιουργίας του HRAF- την δεκαετία του 1960 οπότε και δημιουργήθηκε το *Ethnographic Atlas* (Murdock, 1967) (Murdock & White, 1969). Κάθε εγγραφή του *Ethnographic Atlas* αντιστοιχεί σε μία κοινωνία, σε μία συγκεκριμένη χρονική στιγμή, με πάνω από 90 μεταβλητές που περιγράφουν κοινωνικές και πολιτισμικές πρακτικές της. Ο σκοπός ήταν να προσφέρει μια τυποποιημένη βάση δεδομένων πολιτισμικών χαρακτηριστικών (π.χ. οικονομία, θρησκεία, κοινωνική οργάνωση) η οποία μπορεί να χρησιμοποιηθεί για στατιστική και συγκριτική ανάλυση. Και εδώ έγκειται η βασική του διαφορά από το HRAF. Το *Ethnographic Atlas* περιέχει μια σειρά από τυποποιημένες μεταβλητές οι οποίες διευκολύνουν στατιστικές αναλύσεις, ενώ το HRAF παραθέτει αποσπάσματα από τεκμήρια. Επίσης υπάρχει μια ακόμη σχέση μεταξύ τους, αφού το HRAF επιχειρεί μια αντιστοίχιση (mapping) με το *Standard Cross-Cultural Sample* (SCCS) του *Ethnographic Atlas* (Ember, 2007). Παρότι ερευνητές θέτουν ζητήματα μεροληψίας ποιότητας δεδομένων (Rohner κ.ά., 1982) το *Ethnographic Atlas* χρησιμοποιήθηκε πολύ και χρησιμοποιείται ακόμη αφού ενσωματώθηκε στο D-Place.

Το D-PLACE (<https://d-place.org/>) είναι μια βάση δεδομένων που συνδυάζει πληροφορίες σχετικά με γεωχωρικά σημεία, γλώσσες, πολιτισμούς και το περιβάλλον τους (Kirby κ.ά., 2016). Δημιουργήθηκε για να επιτρέψει τη συσχέτιση και τη συγκριτική μελέτη των πολιτισμικών, γλωσσικών και οικολογικών δεδομένων. Από την αρχή αξιοποίησε τα δεδομένα από τον *Ethnographic Atlas* του Murdock. Αποτελεί ένα εργαλείο για την αναλυτική σύγκριση πολιτισμών και τη διερεύνηση σχέσεων μεταξύ πολιτιστικών, γλωσσικών και οικολογικών χαρακτηριστικών σε παγκόσμιο επίπεδο. Ενημερώνεται συνεχώς και προσφέροντας νέα δεδομένα. Μάλιστα υπάρχει συνεργασία και με τη βάση του HRAF <https://hraf.yale.edu/hraf-collaboration-with-d-place/>.

Βασιζόμενοι στα δεδομένα των παραπάνω βάσεων αρκετοί ερευνητές έχουν δημοσιεύσει διάφορες μελέτες. Ενδεικτικά αναφέρονται δύο σχετικά πρόσφατες και ιδιαίτερα επιδραστικές δημοσιεύσεις. Και οι δύο δημοσιεύτηκαν σε πολύ έγκριτα επιστημονικά περιοδικά. Η μία στο *Current Anthropology* που, όπως λέει και ο τίτλος του, εστιάζει στην ανθρωπολογία και η άλλη στο *PLOS ONE* που είναι ένα περιοδικό που δημοσιεύει μελέτες για πολλά επιστημονικά πεδία. Η πρώτη είναι το άρθρο «*Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies*» (Curry κ.ά., 2019), που βασίστηκε στα δεδομένα HRAF, και έχει μεγάλη απήχηση τόσο εντός της ακαδημαϊκής κοινότητας (με 354 αναφορές στη Scopus, μπαίνει στο 99th percentile)<sup>16</sup> όσο και εκτός αυτής, όπως δείχνουν τα στοιχεία του Altmetric <https://www.altmetric.com/details/55202657>). Με παρόμοιες επιδόσεις (στοιχεία του

---

<sup>16</sup> Τελευταία επικαιροποίηση των στοιχείων Απρίλιος 2025.

Altmetric στο <https://www.altmetric.com/details/150646309>) είναι και το «*The Myth of Man the Hunter: Women's contribution to the hunt across ethnographic contexts*» (Anderson κ.ά., 2023) που βασίζεται σε δεδομένα του D-place.

Τα τρία έργα που παρουσιάστηκαν παραπάνω είναι ίσως τα πιο εμβληματικά του χώρου αλλά σίγουρα δεν είναι τα μόνα. Από τη μία πλευρά υπάρχουν προσπάθειες που μπορεί να μην αναφέρονται στον πυρήνα της ανθρωπολογικής έρευνας αλλά παρόλα αυτά να έχουν ανθρωπολογικό ενδιαφέρον. Μια τέτοια προσπάθεια είναι η Ethnologue <https://www.ethnologue.com/> που περιέχει εθνογραφικά στοιχεία για τους ομιλητές της κάθε γλώσσας που περιλαμβάνει και μπορεί να χρησιμοποιηθεί σε ανθρωπολογικές μελέτες που συνδυάζουν γλώσσα και πολιτισμό. Από την άλλη υπάρχουν προσπάθειες που έχουν έναν πιο επικουρικό χαρακτήρα. Για παράδειγμα η Weber αναπτύσσει μια βάση όπου προσπαθεί να δώσει έμφαση στις σχέσεις ανάμεσα στις δημοσιεύσεις και στα δεδομένα από τα οποία προήλθαν (Weber, 2017) ενώ άλλοι να δημιουργήσουν ελεγχόμενα λεξιλόγια τιμών τα οποία θα διευκολύνουν την αναζήτηση των ανθρωπολογικών συνόλων δεδομένων (Emmelhainz, 2014).

Επίσης καταγράφονται προσπάθειες όπως το Mukurtu <https://mukurtu.org/> που είναι μια δωρεάν, ανοιχτού κώδικα πλατφόρμα διαχείρισης ψηφιακής πολιτιστικής κληρονομιάς. Το Mukurtu είναι ένα έργο που στοχεύει στην ενδυνάμωση των κοινοτήτων ώστε να διαχειρίζονται, να μοιράζονται, να αφηγούνται και να ανταλλάσσουν την ψηφιακή τους κληρονομιά με πολιτισμικά σχετικούς και ηθικά ορθούς τρόπους. Πρόκειται για λογισμικό CMS (Content Management System) που έχει υιοθετηθεί από περισσότερες από 600 ιθαγενείς ομάδες σε όλο τον κόσμο, οι οποίες το χρησιμοποιούν για να καταγράψουν και να μοιραστούν την πολιτιστική τους κληρονομιά.

Τέλος, θα πρέπει να επισημανθεί ότι κάποιες άλλες προσπάθειες, όπως το <http://www.worldcultures.org/>, δεν κατάφεραν να επιβιώσουν και πλέον είναι μερικώς προσβάσιμες μέσω του Way Back Machine του Internet Archive<sup>17</sup>. Στην εποχή της χρήσης τους, ερευνητές επισημαίνουν την αξία τους ως εργαλεία για την συγκριτική ανθρωπολογία (White, 1986) (Schmidt & Schweizer, 1990). Όμως στο ψηφιακό περιβάλλον όταν μια υπηρεσία (service) πάψει να λειτουργεί τότε τα δεδομένα που διαθέτετε μπορεί όχι απλά να πάψουν να είναι εύκολα προσβάσιμα αλλά να χαθούν για πάντα. Για αυτό είναι ιδιαίτερα κρίσιμο να μπορούν τα δεδομένα να απεμπλακούν από τα προγράμματα που τα διαχειρίζονται και αυτό είναι ένα από τα θέματα τα οποία διαπραγματεύεται η επόμενη ενότητα.

---

<sup>17</sup> Οι σύνδεσμοι που ακολουθούν είναι διαθέσιμοι μόνο μέσω του Internet Archive: <http://eclectic.ss.uci.edu/~drwhite/courses/SCCCodes.htm>, <http://eclectic.ss.uci.edu/~drwhite/links2pdf.htm>, <http://eclectic.ss.uci.edu/~drwhite/>

## 4.6 Η σημασία της μοντελοποίησης των δεδομένων

Στην ενότητα αυτή εξετάζεται ο τρόπος με τον οποίο τα δεδομένα κωδικοποιούνται, εστιάζοντας στις δομές των δεδομένων και την εκφραστικότητα αυτών των δομών (τι επιτρέπουν να εκφραστεί), καθώς και στις δυνατότητες επαναχρησιμοποίησης των δεδομένων δίνοντας έμφαση στη σχέση της επαναχρησιμοποίησης (re-use) με τη διαλειτουργικότητα (interoperability). Δηλαδή, ενώ στην προηγούμενη ενότητα το κέντρο βάρους ήταν στο πώς μπορεί να χρησιμοποιηθεί η πληροφορική για την ανάλυση των δεδομένων, στην ενότητα αυτή το κέντρο βάρους μετατοπίζεται στη μοντελοποίηση των δεδομένων είτε για να διευκολυνθεί η ανάλυση είτε για να αναπαρασταθεί/περιγραφεί αποτελεσματικά το αποτέλεσμα της ανάλυσης, ήτοι τα συμπεράσματα της έρευνας. Η μοντελοποίηση των δεδομένων θα καθορίσει, σε μεγάλο βαθμό, κατά πόσον μπορούν να είναι αποτελεσματικά για σύνθετη ανάλυση. Επίσης θα επιτρέψει να τεθούν οι βάσεις για το δρόμο προς την ανοικτή επιστήμη (open science).

Ο Geertz, στο κλασικό έργο του «Η Ερμηνεία των Πολιτισμών», θεωρεί ότι ουσιαστικός προορισμός της ανθρωπολογίας είναι η δημιουργία ενός προσβάσιμου αρχείου των όσων έχει πει ο άνθρωπος (Geertz, 2003, σελ. 41) και ο Marcus αναρωτιέται κατά πόσο η γνώση που παράγουν οι ανθρωπολόγοι μπορεί να έχει τον αθροιστικό χαρακτήρα ενός αρχείου (Marcus, 1998). Η ενότητα αυτή εξετάζει δομές οι οποίες θα επιτρέψουν την περιγραφή αυτού του «consultable record» κατά Geertz ή «archive» κατά Marcus, ώστε, μέσω της μοντελοποίησης, να είναι εφικτή η επεξεργασία του αλλά και η επαναχρησιμοποίησή του στο ψηφιακό περιβάλλον.

Μέσα από μια ψηφιακή περιγραφή είναι δυνατόν να αυξηθεί η επίδραση της ανθρωπολογικής γνώσης τόσο σε διεπιστημονικές προσεγγίσεις όσο και στο ευρύ κοινό. Είναι γνωστό ότι η κοινωνική ανθρωπολογία υποστηρίζει πως δεν επιβεβαιώνονται από τα δεδομένα του πραγματικού κόσμου θέσεις και απόψεις που θεωρούνται τόσο από το ευρύ κοινό αλλά και από επιστημονικούς τομείς ως θέσφατα. Ένα τυπικό παράδειγμα είναι το έργο της McKinnon, η οποία αμφισβητεί βασικές θέσεις της εξελικτικής ψυχολογίας δίνοντας πλήθος παραδειγμάτων που βασίζονται σε εθνογραφικά δεδομένα (McKinnon, 2009). Αυτού του είδους τα συμπεράσματα μπορούν, μέσω της μοντελοποίησης μίας φορμαλιστικής και ακριβούς περιγραφής, να διαδοθούν αλλά και να επικοινωνήσουν/διαλειτουργήσουν με άλλα ώστε να κριθούν για τον ορθότητά τους. Παραφράζοντας το «ethnography is Anthropology, or it is nothing» του Sahllins (2002, σελ. 12) εδώ θα ισχυριστούμε πως είτε η γνώση που παράγεται μπορεί να σταθεί έξω από τον εθνογράφο και να επικοινωνήσει με άλλη γνώση ή, μέχρι να το κάνει, δεν είναι τίποτα. Αν ο ερευνητής δεν μπορεί να μοιραστεί αυτό που είδε, για την επιστήμη, είναι σαν να μην παρατηρήθηκε ποτέ.

Συνεπώς, ο αποτελεσματικός διαμοιρασμός της πληροφορίας απαιτεί μοντελοποίηση. Διότι, όπως παρατηρεί ο Van Der Leeuw, τα μοντέλα επιτρέπουν στους ερευνητές να περιγράψουν χωρίς πλεονασμούς ένα ευρύ φάσμα σχέσεων με ένα βαθμό ακρίβειας που δεν απαντάται στη φυσική γλώσσα (Van Der Leeuw, 2004, σελ. 122). Με δεδομένο ότι η παρούσα μελέτη αναφέρεται στο ψηφιακό περιβάλλον, οι φορμαλισμοί τους οποίους προκρίνει προέρχονται από αυτό και συντάσσονται με τις βασικές αρχές του FAIR, δηλαδή τα δεδομένα πρέπει να είναι ευρέσιμα, προσβάσιμα, διαλειτουργικά

και επαναχρησιμοποιούμενα (findable, accessible, interoperable and reusable). Από το κλασικό κείμενο, με τίτλο «The FAIR Guiding Principles for scientific data management and stewardship» το οποίο συνυπογράφουν δεκάδες συγγραφείς και δημοσιεύτηκε το 2016 (Wilkinson κ.ά., 2016) μέχρι και σήμερα το FAIRness των δεδομένων είναι διακύβευμα της επιστημονικής κοινότητας που τίθεται, μετ' επιτάσεως, σε πολλά επίπεδα. Οι αρχές του FAIR αποτελούν θεμέλιο λίθο της προσπάθειας εδραίωσης ενός περιβάλλοντος ανοικτής επιστήμης (open science), η οποία προωθεί την ελεύθερη διακίνηση γνώσης, την προσβασιμότητα στα επιστημονικά δεδομένα, τη συνεργασία και τη διαφάνεια στην έρευνα<sup>18</sup>. Επιμέρους ερευνητές, αλλά και φορείς, προσπαθούν πείσουν την επιστημονική κοινότητα για την αξία τους και να καταστήσουν ελκυστική τη χρήση τους (Balakireva & Balakirev, 2022). Ειδικά στο πλαίσιο της Ευρωπαϊκής Ένωσης κρίνεται ιδιαίτερα σημαντική η εναρμόνιση με τις εν λόγω αρχές και για αυτό σε πολλά από τα χρηματοδοτούμενα έργα το πλάνο διαχείρισης των δεδομένων (data management plan) θα πρέπει να εναρμονίζεται με τις FAIR αρχές.

Πριν ακολουθήσουν ειδικά ζητήματα που σχετίζονται με την μοντελοποίηση, η οποία κρίνεται ιδιαίτερης σημασίας για την ανθρωπολογική έρευνα, θεωρείται σκόπιμο να διασαφηνιστεί ότι η μοντελοποίηση της πληροφορίας επιτελεί έναν ακόμη κρίσιμο σκοπό: συμβάλει στον διαχωρισμό και την ανεξάρτηση των δεδομένων από το εκάστοτε λογισμικό που διαχειρίζεται αυτά τα δεδομένα. Εξαιτίας της ταχύτατης αλλαγής που συντελείται στο χώρο των τεχνολογικών υποδομών για τη διακίνηση της πληροφορίας (ICT infrastructure) τα προγράμματα είναι καταδικασμένα σε πολύ σύντομη διάρκεια. Για παράδειγμα το site ενός γνωστού προγράμματος καταπολέμησης ιών υπολογιστών, συγκεκριμένα πρόκειται για το Avast, έδινε το εξής μήνυμα για έναν φυλλομετρητή ο οποίος έτρεχε σε λειτουργικό που δεν επέτρεπε, λόγω παλαιότητας, αναβάθμιση για πάνω από μια τριετία: «We support browsers, not dinosaurs. Please update your browser if you want to see the content of this webpage correctly». Δηλαδή αρκεί μια πενταετία για να καταταχθεί ένα λογισμικό στην εποχή των δεινοσαύρων! Άρα, για τη μακροπρόθεσμη διατήρηση και τη βιωσιμότητα (sustainability) των δεδομένων είναι ζωτικής σημασίας να ελαχιστοποιηθεί η εξάρτησή τους από συγκεκριμένα λογισμικά. Και αυτό μπορεί να επιτευχθεί δίνοντας έμφαση στη μοντελοποίηση των δεδομένων ανεξάρτητα από το πρόγραμμα που τα φιλοξενεί.

Τα παραπάνω οδηγούν σε δύο βασικά ζητήματα: το θέμα της διαλειτουργικότητας και εκείνο της επανάχρησης των δεδομένων (data re-use). Με δεδομένο ότι στο σύγχρονο ψηφιακό περιβάλλον η διαλειτουργικότητα είναι προαπαιτούμενο για την αποτελεσματική επαναχρησιμοποίηση των δεδομένων θα ξεκινήσουμε από αυτήν.

---

<sup>18</sup> Στο ελληνικό οικοσύστημα σημαντικό ρόλο διαδραματίζει η [Εθνική Πρωτοβουλία για την Ανοικτή Επιστήμη](#) (ΕΠΑΕ) η οποία έχει εκδώσει και ένα Εθνικό σχέδιο για την ανοικτή επιστήμη (open science) (Athanasίου κ.ά., 2020)

### 4.6.1 Διαλειτουργικότητα

Ένας όρος κλειδί που συνδέεται άμεσα με τη μοντελοποίηση των δεδομένων είναι η διαλειτουργικότητα (interoperability). Ως διαλειτουργικότητα ορίζεται η δυνατότητα ενός συστήματος να επικοινωνεί αποτελεσματικά με άλλα συστήματα. Συνήθως διακρίνεται σε δύο βασικά επίπεδα: το συντακτικό (syntactic) και το σημασιολογικό (semantic)<sup>19</sup>. Το συντακτικό επίπεδο αφορά στην κωδικοποίηση των δεδομένων για τη μεταφορά της πληροφορίας και μπορεί να επιτευχθεί αν δύο συστήματα χρησιμοποιούν κοινά μορφότυπα δεδομένων (data formats) και πρωτόκολλα επικοινωνίας (communication protocols). Από την άλλη πλευρά η σημασιολογική διαλειτουργικότητα είναι η δυνατότητα των υπολογιστικών συστημάτων να ανταλλάσσουν δεδομένα με ξεκάθαρο (unambiguous) κοινό νόημα και είναι μια από τις σημαντικότερες προκλήσεις για την ολοκλήρωση (integration) των δεδομένων.

Η διαλειτουργικότητα συνεισφέρει σημαντικά στην απεξάρτηση των δεδομένων από το λογισμικό που τα διαχειρίζεται (software agnostic), και αυτή η απεξάρτηση είναι ιδιαίτερα κρίσιμη για τη διατήρηση των δεδομένων (preservation) αφού η ζωή των δεδομένων δεν συνδέεται άρρηκτα με το λογισμικό που τα διαχειρίζεται. Προφανές όφελος είναι ότι τα δεδομένα που χρησιμοποιούνται από ένα λογισμικό μπορούν να χρησιμοποιηθούν και από ένα άλλο, άρα αυξάνεται η πιθανότητα της επαναχρησιμοποίησης δεδομένων, εφόσον αυτά δεν είναι «εγκλωβισμένα» για χρήση από ένα συγκεκριμένο λογισμικό. Όμως ένα πιο έμμεσο αλλά πολύ σημαντικό όφελος είναι το ότι η διαλειτουργικότητα βοηθάει στη μοντελοποίηση των δεδομένων, διότι απαιτεί τη δημιουργία κατηγοριών και κανόνων που έχουν καθολική εφαρμογή. Άρα συμβάλλει σημαντικά στην δυνατότητα αφαίρεσης (abstraction) χωρίς όμως να καταργεί την πρακτική εφαρμογή, δηλαδή πρόκειται για εφαρμοσμένη θεωρία.

Η εφαρμογή της διαλειτουργικότητας στην πράξη παρουσιάζει σημαντικές δυσκολίες και βασικοί λόγοι για αυτό είναι η ετερογένεια των δεδομένων και η μεταβολή τους τόσο μέσα στο χρόνο όσο και εντός των διαφορετικών πολιτιστικών κλάδων της γνώσης (M. D. Fischer & Ember, 2018, σελ. 330). Μια ενδιαφέρουσα ένδειξη για τη δυσκολία στην εφαρμογή της διαλειτουργικότητας προκύπτει από το “Special Issue on Computing in Anthropology” (Vol. 31 Issue 1, February 2013) του έγκριτου περιοδικού Social Science Computer Review όπου η λέξη Interoperability εμφανίζεται πέντε (5) φορές στο εισαγωγικό άρθρο “Harmonizing Diversity: Tuning Anthropological Research to Complexity” (M. D. Fischer κ.ά., 2013) αλλά δεν υπάρχει πουθενά στα υπόλοιπα επτά

---

<sup>19</sup> Κάποιοι ερευνητές θέτουν ένα αφαιρετικότερο επίπεδο που αφορά στην “pragmatic interoperability”. Το επίπεδο αυτό προϋποθέτει τα δύο προαναφερόμενα και έχει να κάνει κυρίως με τη σκοπιμότητα (intended effect) κατά την επικοινωνία των συστημάτων (Asuncion & van Sinderen, 2010; Pokraev κ.ά., 2005). Άρα το επίπεδο αυτό είναι αρκετά γενικό και αφορά στην ερμηνεία (interpretation) των μοντέλων -και των δεδομένων που αυτά περιγράφουν- μέσα στο πλαίσιο της ίδιας της υπολογιστικής επικοινωνίας αλλά είναι έξω από αυτήν. Με την ίδια λογική ο συγγραφέας της παρούσας διατριβής δεν συμμερίζεται την άποψη ότι όλες τις γενικεύσεις που αφορούν σε αντίστοιχα επίπεδα (π.χ. θεσμική ή οργανωσιακή διαλειτουργικότητα,) είναι όντως ζητήματα διαλειτουργικότητας.

(7) άρθρα. Από την άλλη, στο προηγούμενο ειδικό τεύχος σχετικά με Computing in Anthropology (Vol.24, Issue 1, 2006) δεν υπάρχει αναφορά σε αυτή την έννοια.

Ένας βασικός παράγοντας για την ανάπτυξη προδιαγραφών διαλειτουργικότητας έχει να κάνει με το πόσο ώριμος είναι ο τομέας. Για παράδειγμα, η έμφαση στη συστηματοποίηση της γενεαλογίας οδήγησε από πολύ νωρίς στην προτυποποίηση των μορφοτύπων (standardization of formats). Το GEDCOM (ακρωνύμιο για το Genealogical Data Communication) είναι απλό plain txt format το οποίο ξεκίνησε να αναπτύσσεται από το 1984 και αποτελεί έναν de facto προσδιορισμό για την ανταλλαγή γενεαλογικών δεδομένων μεταξύ διαφορετικών λογισμικών που προορίζονται για διαχείριση γενεαλογικών δεδομένων <https://en.wikipedia.org/wiki/GEDCOM>. Από την άλλη, απαιτείται προσοχή διότι η προτυποποίηση σε ένα κοινό σχήμα μπορεί να γίνει αντιπαραγωγική αφού παγιώνει το state-of-the-art και δεν επιτρέπει ειδικές περιπτώσεις και εξαιρέσεις (Doerr & Iorizzo, 2008, σελ. 3). Κατά συνέπεια, η προτυποποίηση που απαιτεί η διαλειτουργικότητα μπορεί να αποτελέσει τροχοπέδη για την εξέλιξη ενός επιστημονικού τομέα αφού αποδίδει κατηγορηματικά τόσο τις έννοιες που χρησιμοποιεί ο τομέας όσο και το νόημα αυτών των εννοιών. Αυτό όμως απλά σημαίνει ότι όσο εξελίσσονται τα όποια μοντέλα και πρότυπα τόσο θα πρέπει να εξελίσσεται και η διαλειτουργικότητα μεταξύ τους. Άρα η διαλειτουργικότητα δεν είναι κάτι πεπερασμένο αλλά διαρκής διαδικασία (διαρκής επανα-εννοιολόγηση) η οποία είναι απαραίτητη προϋπόθεση για το αποτελεσματικό re-use των δεδομένων ανάμεσα σε διαφορετικά συστήματα.

#### 4.6.2 Επανάχρηση και διαμοιρασμός των δεδομένων

Η διαλειτουργικότητα μπορεί να είναι σημαντική προϋπόθεση για το αποτελεσματικό data re-use αλλά η επίτευξη της δεν είναι το μόνο πρόβλημα που πρέπει να αντιμετωπιστεί. Γενικά οι κοινωνικοί επιστήμονες είναι επιφυλακτικοί στο data sharing (Kim & Adler, 2015), και όσοι ανήκουν στο qualitative social science community ακόμη περισσότερο (Cliggett, 2013). Σίγουρα όχι ο μοναδικός, αλλά ένας σημαντικός λόγος αυτής της επιφυλακτικότητας είναι ότι αρκετές φορές υπάρχουν «ευαίσθητα» δεδομένα τα οποία έχουν ιδιαίτερους περιορισμούς στο να κοινοποιηθούν. Τον τελευταίο καιρό εντοπίζεται μια γενικευμένη τάση να ζητείται η συναίνεση του ατόμου ως προς τη χρήση προσωπικών δεδομένων από τρίτους, αλλά και η εφαρμογή κανονισμών, όπως το General Data Protection Regulation (GDPR) της ΕΕ, ή άλλων πρακτικών, όπως η απάλειψη των ονομάτων (ανωνυμοποίηση) και η χρήση ψευδωνύμων, για την προστασία των ατόμων.

Από την άλλη, πολλοί ερευνητές θεωρούν ότι ούτως ή άλλως τα πρωτογενή δεδομένα είναι μόνο για προσωπική χρήση διότι ένας τρίτος θα τα ερμηνεύσει εκτός των συμφραζομένων και νομοτελειακά θα καταλήξει σε λαθεμένα συμπεράσματα. Οι Koning et al παρουσιάζουν μερικές βασικές αρχές για το data sharing των πρωτογενών ανθρωπολογικών δεδομένων οι οποίες μάλιστα έχουν επικυρωθεί από τη ολλανδική ανθρωπολογική ομοσπονδία. Παρόλα αυτά οι ίδιοι δηλώνουν ότι το ερευνητικό υλικό των ανθρωπολόγων δεν μπορεί να θεωρείται disembodied and transferable 'data' (de Koning κ.ά., 2019, σελ. 171). Ως εκ τούτου για τη χρήση ποιοτικής έρευνας τρίτων η

πρόσβαση στα συμφραζόμενα είναι ιδιαίτερης σημασίας αλλά, όπως σημειώνει η Jo-Anne Kelder «όταν χρησιμοποιείς τα δεδομένα κάποιου άλλου είναι δύσκολο να έχεις πρόσβαση στα συμφραζόμενα» (Kelder, 2005). Μια ευρύτερη συζήτηση για τους προβληματισμούς των ανθρωπολόγων γύρω από το θέμα αναπτύσσεται στο επιστημονικό forum “Data management in anthropology: the next phase in ethics governance?” (Pels κ.ά., 2018) ενώ πάνω σε ζητήματα ηθικής για τη διαχείριση των δεδομένων (ειδικά των big data) στο άρθρο “Future Anthropology Ethics and Datafication: Temporality and Responsibility in Research” (Pink & Lanzeni, 2018).

Σε αντίθεση με τους παραπάνω, η Reyes, η οποία μάλιστα στέκεται κριτικά απέναντι στην καθιερωμένη πρακτική της χρήσης ψευδωνύμων, ισχυρίζεται πως το κέντρο βάρους πρέπει να μετακυληθεί από το πού θα αποθηκευτούν τα δεδομένα και σε ποιους θα είναι προσβάσιμα, προς το πώς τα εθνογραφικά δεδομένα θα είναι διαφανή (transparent). Αυτός ο όρος αναφέρεται στη σαφήνεια τού τι θεωρούμε ως δεδομένα ή πληροφορία και ποια είναι η χρονική περίοδος κατά την οποία συλλέχθηκαν (Reyes, 2017). Στο ίδιο μήκος κύματος ο Murillo συνδέει τα ανοικτά δεδομένα (open data) με αυτό που ονομάζει “collaborative ethnography” και αναγνωρίζει την ανάγκη για την ενσωμάτωση των ανθρωπολογικών προβληματισμών στο σχεδιασμό και την υλοποίηση των ψηφιακών τεχνολογιών (Murillo, 2018, σελ. 577) ενώ τονίζει τη σημασία που έχει η αναφορά στις συνθήκες παραγωγής των δεδομένων διότι, χωρίς τη σαφή περιγραφή τους, τα δεδομένα δεν μπορεί να είναι χρήσιμα για διαμοιρασμό σε άλλους ερευνητές (Murillo, 2018, σελ. 579–580).

Με βάση τα παραπάνω, η διαφάνεια και ο διαμοιρασμός των δεδομένων φαίνεται πως έχουν σχέση αλληλεξάρτησης, διότι η διαφάνεια συμβάλει στο διαμοιρασμό και αντίστροφα. Εντούτοις θα πρέπει να καταστεί σαφές ότι η μοντελοποίηση των δεδομένων με σκοπό την επαναχρησιμοποίησή τους δεν σημαίνει, κατ’ ανάγκη, το διαμοιρασμό των δεδομένων. Το re-use μπορεί να γίνει και από τον ίδιο ερευνητή με σκοπό τη μελέτη διαφορετικού ερευνητικού ερωτήματος, διότι τα δεδομένα δεν ερμηνεύονται άπαξ αλλά επιδέχονται πολλαπλών αναλύσεων και ερμηνειών. Άρα, η αξία της επαναχρησιμοποίησης των εθνογραφικών δεδομένων είναι υψηλή, παρότι είναι δύσκολο να επιτευχθεί (M. D. Fischer & Ember, 2018, σελ. 329). Ένας ακόμη βασικός λόγος για αυτό είναι ότι οι περισσότεροι ανθρωπολόγοι δεν κωδικοποιούν τα δεδομένα τους με τέτοια μορφή που θα καθιστούσε εφικτή την κυκλοφορία τους. Ως εκ τούτου, η ανθρωπολογία ανήκει στις επιστήμες που αν εφαρμοστεί η επαναχρησιμοποίηση των δεδομένων, θα μπορούσε να την οδηγήσει σε ριζική επιστημολογική αλλαγή (Carlson & Anderson, 2007, σελ. 642).

Ήδη διαμορφώνονται δειλά δειλά και κάποιες πλατφόρμες για το διαμοιρασμό τέτοιων πληροφοριών και τη συνεργασία μεταξύ των ανθρωπολόγων όπως το Platform for Experimental, Collaborative Ethnography (PECE) <https://worldpece.org/about>. Το PECE βρίσκεται στο επίκεντρο ενός ερευνητικού έργου που διερευνά πώς μπορεί να σχεδιαστεί η ψηφιακή υποδομή για να υποστηρίξει τη συνεργατική έρευνα και ερμηνεία. Το PECE, σύμφωνα με τους δημιουργούς του, παρέχει ένα μέρος για την αρχειοθέτηση και την κοινή χρήση πρωτογενών δεδομένων που παράγονται από μελετητές των

εμπειρικών ανθρωπιστικών και κοινωνικών επιστημών, διευκολύνοντας την αναλυτική συνεργασία και ενθαρρύνοντας τον πειραματισμό με ποικίλους τρόπους δημοσίευσης.

Θα πρέπει να τονισθεί ότι για όλους τους τύπους δεδομένων, δηλαδή πρωτογενή, δευτερογενή, ή μετα-αναλύσεις, «ένας παράγοντας κλειδί για την επαναχρησιμοποίηση των δεδομένων είναι η ορθή δόμησή τους. Όσο πιο κανονικοποιημένα και ορθά ορισμένα είναι τα δεδομένα τόσο ευκολότερα μπορούν να αναπτυχθούν εργαλεία επεξεργασίας για αξιόπιστη επαναχρησιμοποίησή τους» (Heath & Bizer, 2011, σελ. 2). Το σύγχρονο περιβάλλον για την κωδικοποίηση των δεδομένων είναι τα διασυνδεδεμένα δεδομένα (linked data) -ανοιχτά ή όχι-, που είναι μέρος του σημασιολογικού ιστού, και στον οποίο είναι εφικτή τόσο η επαναχρησιμοποίηση των πρωτογενών δεδομένων όσο και των συμπερασμάτων που προκύπτουν από την επεξεργασία τους.

### 4.6.3 Προσπάθειες αναπαράστασης ανθρωπολογικής γνώσης

Παρότι όπως έδειξε η συστηματική και εκτεταμένη επισκόπηση της βιβλιογραφίας ελάχιστα πράγματα έχουν γίνει γύρω από θέματα αναπαράστασης ανθρωπολογικής γνώσης. Όμως δεν μπορούμε να μην αναφερθούμε σε μια σειρά από προσπάθειες αναπαράστασης γνώσης που έχουν λάβει χώρα και οι οποίες αφορούν, σε μεγαλύτερο ή μικρότερο βαθμό, την ανθρωπολογία. Θα πρέπει να τονιστεί ότι σε ελάχιστες από αυτές είχαν ενεργή συμμετοχή ανθρωπολόγοι. Από όσες αναφέρονται στη συνέχεια μόνο η τελευταία εκπονήθηκε από ανθρωπολόγους.

Κάποιοι, όπως οι Rhefo et al, φιλοδοξούν να περιγράψουν όλους τους πολιτισμούς σε μια οντολογία με αποτέλεσμα να καταλήγουν σε πολύ απλουστευτικές προσεγγίσεις (Rhefo κ.ά., 2015). Άλλοι προσπαθούν να περιορίσουν το αντικείμενο της περιγραφής και εστιάζουν σε θέματα ιθαγενούς γνώσης. Επ' αυτού μια σχετική δουλειά είναι των Haron & Hamiz (Haron & Hamiz, 2014a) (Haron & Hamiz, 2014b) η οποία, παρότι δεν δρέπει δάφνες ούτε από την ανθρωπολογική ούτε από την πλευρά της αναπαράστασης γνώσης, η καινοτομία της την καθιστά μια ενδιαφέρουσα προσπάθεια οντολογικής αναπαράστασης. Οι εν λόγω ερευνητές χρησιμοποιούν τη γλώσσα OWL και το λογισμικό Protégé για την ανάπτυξη μιας οντολογίας αλλά δεν δικαιολογούν επαρκώς ούτε την επιλογή των κλάσεων ούτε των ιδιοτήτων που επέλεξαν. Επίσης σε μια προσπάθεια αναπαράστασης της ιθαγενούς γνώσης, συγκεκριμένα παραδοσιακής ιατρικής, οι Ayimdjji et al παρουσιάζουν τη λογική η οποία θα μπορούσε να χρησιμοποιηθεί για την ανάπτυξη οντολογίας και προτείνουν τη description logic (DL) για την ανάπτυξη μιας οντολογίας για την παραδοσιακή αφρικανική ιατρική. Όμως δίνουν μεγαλύτερη σημασία στον φορμαλισμό που απαιτείται για την υλοποίηση της DL και δεν προτείνουν συγκεκριμένες κλάσεις ή ιδιότητες. Ενδιαφέρον έχει το συμπέρασμα στο οποίο καταλήγουν: «πριν την ολοκλήρωση μιας οντολογίας για την παραδοσιακή αφρικανική ιατρική είναι απαραίτητη η δημιουργία μιας πολύ-επιστημονικής ομάδας» (Ayimdjji κ.ά., 2011, σελ. 250). Οι Chi et al συνέλεξαν ανοικτά σύνολα δεδομένων (open datasets) σχετικά με τους Ταϊβανέζικους ιθαγενείς πληθυσμούς, μετέτρεψαν σε RDF όσα από αυτά δεν ήταν ήδη σε RDF και ανέπτυξαν SPARQL endpoints για να διασυνδέσουν διαφορετικές πηγές (Chi κ.ά., 2020). Παρότι δεν αναπτύσσουν μια ενιαία οντολογία οι

τεχνολογίες που χρησιμοποιούν θα μπορούσαν να οδηγήσουν προς αυτή την κατεύθυνση. Άλλες προσπάθειες εστιάζουν σε ακόμη πιο περιορισμένα θέματα, όπως η προσπάθεια αναπαράστασης χορών (Chantas κ.ά., 2018) η οποία εντάσσεται στη διατήρηση της άυλης πολιτισμικής κληρονομιάς (preservation of Intangible Cultural Heritage). Ενώ μια πολύ πρόσφατη προσπάθεια επιχειρεί να δημιουργήσει γράφους γνώσης για τις μειονότητες που κατοικούν στην περιοχή του ποταμού Μεκόνγκ (Chansanam κ.ά., 2025). Τέλος, θα ήταν παράλειψη να μην αναφερθεί μια προσπάθεια αναπαράστασης γνώσης με όρους πολύ κοντινούς στο τι ορίζεται σήμερα ως αναπαράσταση γνώσης, και στην οποία συμμετείχαν ανθρωπολόγοι. Πρόκειται για το έργο των Fischer & Finkelstein με τίτλο “Social knowledge representation: A case study” (M. D. Fischer & Finkelstein, 1991). Το συγκεκριμένο έργο αντί για την OWL, η οποία δεν υπήρχε τότε, βασίζεται στη Modal Action Logic, η οποία είναι μια γλώσσα, αποτελούμενη από ένα σύνολο αξιωμάτων και δηλώσεων, που χρησιμοποιείται για την περιγραφή καταστάσεων οι οποίες σχετίζονται τόσο με δομικές σχέσεις όσο και με τυχόν επιδράσεις σε αυτές τις σχέσεις όπως προκύπτουν από ενέργειες υποκειμένων. Αυτήν τη γλώσσα χρησιμοποίησαν ως μέσο για να αναπτύξουν και εφαρμόσουν μια φορμαλιστική μέθοδο για την περιγραφή της κοινωνικής συμπεριφοράς -αλλά και θεωριών για την κοινωνική συμπεριφορά- σε μία εθνογραφική μελέτη σχετικά με τους γάμους μιας κοινότητας Punjabi στο Πακιστάν, έχοντας ως απώτερο στόχο να προσδιορίσουν τις αναγκαίες πληροφορίες ώστε να επιτευχθεί λογική συνέπεια για όσα υποστήριζαν στις μαρτυρίες τους τα υποκείμενα.

#### 4.7 Συμπέρασμα κεφαλαίου

Ο προβληματισμός για την χρήση των υπολογιστών στην ανθρωπολογία είναι πολύ παλιός. Έχει τεθεί από τη δεκαετία του 1960 και το 1965 έγινε μια συνάντηση εργασίας (workshop) επί του θέματος, στην οποία συμμετείχαν ανθρωπολόγοι, γλωσσολόγοι και πληροφορικοί (Hymes, 1965). Παρότι το θέμα επανέρχεται συχνά: π.χ. ο Hakken συγγραφέας έγραφε στις αρχές της δεκαετίας του 1990, αναφερόμενος στην δεκαετία του 1980, ότι οι ανθρωπολόγοι έχουν πλέον αρκετούς λόγους να ασχοληθούν με την πληροφορική (Hakken, 1993) μάλλον λίγοι πείσθηκαν να το πράξουν. Όμως αυτή η μικρή ομάδα ανθρώπων, κατάφερε να δημιουργήσει μια αξιομνημόνευτη παρακαταθήκη. Η έμφαση δόθηκε κυρίως σε θέματα συγγένειας, αλλά όπως έδειξε το κείμενο που προηγήθηκε δεν περιορίστηκε σε αυτά.

Ένα θεμελιώδες ζήτημα που θέτει η ενασχόληση της Ανθρωπολογίας με την χρήση υπολογιστών αφορά στη σχέση της ανθρωπολογίας με τις μεθόδους που εκείνη χρησιμοποιεί. Οι ανθρωπολόγοι έχουν συνυφάνει την έρευνά τους με την επιτόπια συμμετοχική παρατήρηση. Όμως ένας επιστημονικός τομέας δεν είναι δυνατόν να περιορίζεται από μια συγκεκριμένη μεθοδολογία. Ο επιστημονικός κλάδος ορίζει το αντικείμενο της μελέτης και όχι το πώς αυτό θα μελετηθεί. Δεν είναι λίγες οι φορές που οι ανθρωπολόγοι θεωρούν το «πώς» αναπόσπαστο μέρος της μελέτης, εξίσου σημαντικό με το «τι». Οι υπολογιστικές τεχνολογίες προσθέτουν ένα ακόμη μεθοδολογικό εργαλείο στην φαρέτρα της Κοινωνικής Ανθρωπολογίας. Και αυτό το μεθοδολογικό εργαλείο μπορεί να χρησιμοποιηθεί είτε επικουρικά σε πιο συμβατικές μεθόδους συλλογής και

οργάνωσης δεδομένων (π.χ. για την οργάνωση των σημειώσεων του πεδίου (fieldnotes)) είτε για καινοτόμες αναλύσεις που δεν ήταν εφικτές πριν. Άλλωστε ο σύγχρονος άνθρωπος αφήνει διαρκώς πίσω του ένα τεράστιο ψηφιακό αποτύπωμα το οποίο περιέχει αδιανόητα μεγάλο όγκο πληροφορίας. Αυτό το αποτύπωμα, τόσο λόγω του τύπου του -αφού είναι ψηφιακό- τόσο και λόγω του μεγέθους του είναι πρακτικά αδύνατον να μελετηθεί χωρίς τη χρήση υπολογιστικής τεχνολογίας.

Όπως προκύπτει από την ανάλυση της βιβλιογραφίας, ένας σημαντικός παράγοντας που οδηγεί στη χρήση υπολογιστικών μεθόδων, ειδικά όταν υπάρχει μεγάλος όγκος δεδομένων, είναι και η ευελιξία οπτικοποίησης των αποτελεσμάτων. Οι ποικίλοι τρόποι οπτικοποίησης, τόσο των ανεπεξέργαστων δεδομένων όσο και των διαφόρων αναλύσεων αυτών, προσφέρουν τόσο έναν πιο φιλικό τρόπο παρουσίασης δεδομένων όσο και μερικές φορές έναν πιο ουδέτερο -χωρίς πολιτισμικές προκαταλήψεις- τρόπο μετάδοσης νοήματος. Οι Read κ.ά. σημειώνουν ότι ο καλύτερος τρόπος για να αποτυπωθεί η εννοιολογική δομή των συστημάτων συγγένειας, χωρίς να χρησιμοποιούνται οι πολιτισμικές αντιλήψεις κάποιου ως συσκοτιστικό φίλτρο, είναι συνήθως να ζητούνται διαγράμματα και όχι λίστες όρων. Όπως αναφέρθηκε και προηγουμένως, συγκρίνοντας το αγγλικό σύστημα συγγένειας με αυτό του Punjabi, ο Read κ.α. αναφέρουν ότι το πιο σημαντικό είναι το σημείο της πιο εμφανής αντίθεσης. Τα σχήματα δεν είναι τα ίδια. Τα αγγλικά συστήματα μοιάζουν με χριστουγεννιάτικο δέντρο και τα Punjabi με πεταλούδα. Ως εκ τούτου, θεωρούν ότι, στην ουσία, οι ορολογίες συγγένειας δεν μοιάζουν με σύνολα ετικετών αλλά με γεωμετρίες. Δεν είναι συλλογές λέξεων αλλά συστήματα ιδεών (D. Read κ.ά., 2013). Στο πλαίσιο της ανθρωπολογικής έρευνας, ένα άλλο παράδειγμα οφελών από την οπτικοποίηση δεδομένων δίνεται από τους Abramson κ.ά. με τη χρήση προτύπων για αυτό που ονομάζουν «ethnoarrays» (Abramson κ.ά., 2018). Η έμφαση που δίνεται στο ρόλο των μοτίβων για την οπτικοποίηση εθνογραφικών δεδομένων, για παράδειγμα μέσω διαγραμμάτων, αναπτύσσεται στο έργο τους *Beyond Text: Using Arrays to Represent and Analyze Ethnographic Data* (Abramson & Dohan, 2015).

Η ανθρωπολογία βασίζεται εξ ορισμού στη σύγκριση, η οποία αποτελεί θεμελιώδη μεθοδολογικό πυλώνα, καθώς επιτρέπει την κατανόηση της πολιτισμικής ποικιλομορφίας. Η σύγκριση δεν έχει στόχο την κατάταξη των πολιτισμών και τη δημιουργία στερεοτύπων, αλλά τη βαθύτερη κατανόηση των τρόπων με τους οποίους οι άνθρωποι νοηματοδοτούν τον κόσμο τους. Οι βάσεις δεδομένων, ως εργαλεία που επιτρέπουν τις συγκρίσεις, θεωρούνται κρίσιμης σημασίας για τη συγκριτική ανθρωπολογία (Schmidt & Schweizer, 1990). Διότι οι βάσεις δεδομένων επιτρέπουν την εφαρμογή ποικίλων, ανάλογα με τον τρόπο σχεδιασμού τους, κριτηρίων σε πολύ μεγάλους όγκους πληροφορίας. Όμως οι βάσεις δεδομένων -αφού είναι crafted data- δημιουργούνται με συγκεκριμένο σκοπό. Στην παρούσα διατριβή η εστίαση είναι στο επιστημονικό πεδίο της ανθρωπολογίας αλλά υπάρχουν και προσπάθειες που εντάσσουν την κωδικοποίηση στο πλαίσιο της δημιουργίας αυτοματοποιημένων μηχανισμών για την ανάλυση των πολιτισμικών ομοιοτήτων και διαφορών που θα μπορούσαν να χρησιμοποιηθούν ως βοηθήματα για τη διαπολιτισμική επικοινωνία (Chung & Lieberman, 2007, σελ. 65). Και είναι εντυπωσιακό ότι οι Chung & Lieberman

δημιούργησαν -είκοσι χρόνια πριν- μια εφαρμογή που θα μπορούσε να χρησιμοποιηθεί από τουρίστες για να τους βοηθάει στις αλληλεπιδράσεις τους με τον πολιτισμό που επισκέπτονται δίνοντας τους συμβουλές σε απλά καθημερινά πράγματα όπως για παράδειγμα το αν ένα άτομο πρέπει ή όχι να αφήσει φιλοδώρημα στο εστιατόριο του Τόκιο (Chung, 2006). Το αποτέλεσμα που παρήγαγαν μπορεί να μην εγκαταστάθηκε σε πλήθος συσκευών αλλά η θεωρητική του προσέγγιση διαθέτει ακόμη μεγάλο ενδιαφέρον και για αυτό και αναφέρονται σε διάφορα σημεία της παρούσας μελέτης.

Αφήνοντας τα apps για τους τουρίστες και επιστρέφοντας στον πυρήνα των θεμάτων που απασχολούν την διατριβή, η συζήτηση για τις βάσεις δεδομένων οδήγησε στην ανάγκη μοντελοποίησης των δεδομένων. Καταδείχθηκε το πώς η διαλειτουργικότητα αποτελεί θεμέλιο λίθο της μοντελοποίησης αφού επιτρέπει τόσο την απεξάρτηση των δεδομένων από το λογισμικό το οποίο τα διαχειρίζεται όσο και τη δυνατότητα να επικοινωνήσει ο ένας επιστημονικός τομέας με τους υπόλοιπους. Αυξάνοντας με τον τρόπο αυτό τόσο την επανάχρηση των δεδομένων, όσο και την διεύρυνση του περιβάλλοντος μέσα στο οποίο μπορούν να επεξεργαστούν, τείνοντας το χέρι στην διεπιστημονική προσέγγιση. Επίσης, παρουσιάστηκαν οι αντιρρήσεις και οι ενδοιασμοί που εκφράζονται από πλήθος ανθρωπολόγων τόσο για το διαμοιρασμό όσο και την επανάχρηση των δεδομένων.

Τέλος να σημειωθεί ότι η ανάλυση της σχετικής βιβλιογραφίας έδειξε ότι δεν υπάρχει σχεδόν τίποτα, που να έχει παραχθεί από ανθρωπολόγους, και να σχετίζεται με την αναπαράσταση της ανθρωπολογικής γνώσης στοχεύοντας το οικοσύστημα του σημασιολογικού ιστού.

## 5 Αναπαράσταση και οργάνωση της γνώσης: φυσικές και τεχνητές γλώσσες

Η επισκόπηση του πεδίου, όπως αποτυπώθηκε στο προηγούμενο κεφάλαιο, κατέδειξε ότι πολύ μικρή μερίδα ερευνητών έχει ασχοληθεί συστηματικά με θέματα οργάνωσης και κυρίως αναπαράστασης της ανθρωπολογικής γνώσης, πέραν της κειμενικής αποτύπωσης. Στο κεφάλαιο αυτά εξετάζονται πιθανές προσεγγίσεις που θα μπορούσαν να φανούν αποτελεσματικές για την οργάνωση και την αναπαράσταση της ανθρωπολογικής γνώσης. Για το λόγο αυτό διερευνώνται αντιπροσωπευτικά συστήματα οργάνωσης και αναπαράστασης γνώσης με σκοπό να αναδειχθούν τα καταλληλότερα είδη αυτών.

Αρχικά αντιπαραβάλλεται η φυσική γλώσσα ως μέσο/φορέας αποτύπωσης της γνώσης με τεχνητές φορμαλιστικές γλώσσες αποτύπωσης της γνώσης. Παρότι υπάρχει διαφωνία κατά πόσο η γνώση είναι δυνατόν να αποτυπωθεί -ο αντίλογος λέει ότι η γνώση βρίσκεται μόνο μέσα στον ανθρώπινο νου και πως ό,τι αποτυπώνεται σε κείμενα, εικόνες κλπ είναι πληροφορία- εδώ, επειδή δεν αναφερόμαστε σε απαρίθμηση πληροφοριών αλλά σε υψηλότερο επίπεδο αφαίρεσης, ο όρος «γνώση» χρησιμοποιείται για αυτή την εξελιγμένη (σοφιστική/advanced) αποτύπωση πληροφορίας. Άλλωστε ο όρος είναι γενικά αποδεκτός και οριοθετημένος στο πεδίο της επιστήμης της πληροφορίας. Η ευρέως αποδεκτή διαβάθμιση είναι: τα δεδομένα αφορούν σχετικά αδόμητες μορφές, τα δεδομένα μετατρέπονται σε πληροφορία όταν επεξεργαστούν και οργανωθούν περισσότερο και γίνονται γνώση όταν οργανωθούν με τέτοιο τρόπο ώστε να μπορούν να οδηγήσουν σε άμεσα συμπεράσματα. Άρα, αυτή η προσέγγιση διαθέτει ένα ακόμη επίπεδο ανάλυσης και έτσι το ανθρώπινο χαρακτηριστικό δεν είναι η γνώση αλλά, η σοφία (wisdom) (Frické, 2019). Με αυτή τη λογική είναι δυνατή η καταχώρηση «γνώσης» -αλλά όχι σοφίας- εντός ενός υπολογιστικού συστήματος. Η προηγούμενη προσέγγιση είναι ιδιαίτερα γνωστή στο χώρο της επιστήμης της πληροφορίας ως η *πυραμίδα DIKW*, από τα αρχικά των αντίστοιχων αγγλικών όρων Data, Information, Knowledge και Wisdom.

## 5.1 Η φυσική γλώσσα ως φορέας νοήματος

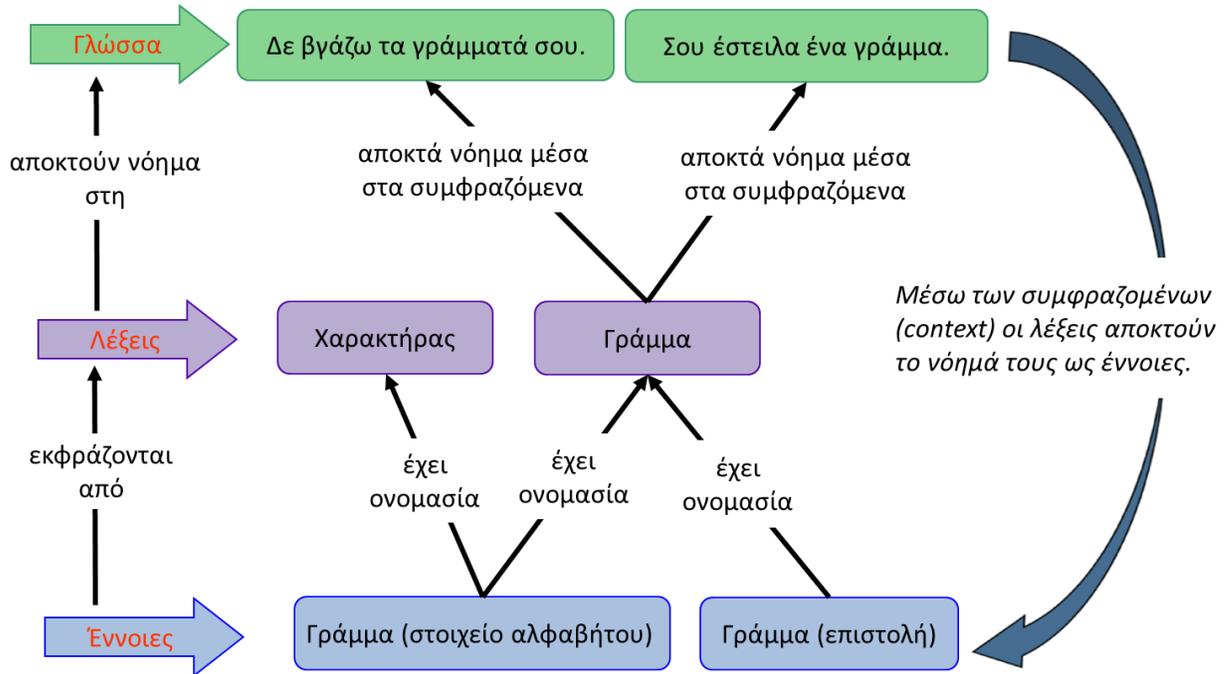
Οι λέξεις είναι εξαιρετικά ουσιώδεις για την επικοινωνία και για τη σκέψη, επειδή ακριβώς ανταποκρίνονται σε έννοιες (Παναγιωτίδης, 2014, σελ. 18). Σύμφωνα με την εγκυκλοπαίδεια φιλοσοφίας του Στάνφορντ «οι έννοιες είναι τα συστατικά της σκέψης. Ως εκ τούτου είναι κρίσιμες για τις διαδικασίες της κατηγοριοποίησης, της εξαγωγής συμπερασμάτων, της μνήμης, της μάθησης και της λήψης αποφάσεων» (Margolis & Laurence, 2011, παρ. 1). Στο κλασικό έργο του «Women, Fire, and Dangerous Things: What Categories Reveal about the Mind», ο Lakoff τονίζει πως χωρίς την ικανότητά μας να κατηγοριοποιούμε, δεν μπορούμε να λειτουργήσουμε ούτε στον φυσικό κόσμο ούτε στην κοινωνική ζωή (Lakoff, 1990, σελ. 6). Ενώ ο Rips σημειώνει ότι οι έννοιες είναι διανοητικές αναπαραστάσεις οι οποίες αναφέρονται σε κατηγορίες (Rips κ.ά., 2013, σελ. 177). Άρα, στο πλαίσιο της νοητικής διαδικασίας, οι λέξεις ανάγονται σε έννοιες και οι έννοιες σε κατηγορίες.

Όμως η σχέση ανάμεσα στη λέξη και στην έννοια δεν είναι αμφίδρομα μονοσήμαντη. Για παράδειγμα, μία λέξη μπορεί να αναφέρεται σε πολλές έννοιες (πολυσημία) ενώ μία έννοια να έχει πολλές λέξεις για να περιγραφεί (συνωνυμία). Προφανώς η συνωνυμία και η πολυσημία δεν είναι οι μόνοι παράγοντες που δημιουργούν ασάφεια στη φυσική γλώσσα. Ένα ακόμη παράδειγμα ασάφειας σε ένα λεκτικό μήνυμα είναι η μεταφορά, παρότι ο ανθρώπινος εγκέφαλος είναι αρκετά εκπαιδευμένος στην αναγνώρισή της (Lakoff & Johnson, 2003). Αυτή η εγγενής ασάφεια της φυσικής γλώσσας απασχόλησε και απασχολεί τη σύγχρονη διανόηση. Ένας από τους πρώτους διανοητές που ασχολήθηκε συστηματικά με το θέμα ήταν ο Ludwig Wittgenstein. Το συμπέρασμα της πρώιμης ενασχόλησής του ήταν η ανάγκη για συσχέτιση μίας λέξης με μία έννοια. Στην πορεία αναθεώρησε και τόνισε την ανάγκη των συμφραζομένων (context) για την αποσαφήνιση του νοήματος (Blair, 2006).

Το Σχήμα 1 –παραλλαγή σχήματος που εμφανίζεται στην εργασία «Expressiveness and machine processability of Knowledge Organization Systems (KOS): an analysis of concepts and relations» (Peponakis κ.ά., 2019)– παρουσιάζει τον τρόπο με τον οποίο οι έννοιες αποκτούν το νόημά τους μέσα στη φυσική γλώσσα. Η βασική λογική πάνω στην οποία στηρίζεται η γραφική απεικόνιση είναι αυτή της ύστερης φιλοσοφικής θεώρησης του Wittgenstein. Όπως είναι προφανές, μια λέξη εντός ενός κειμένου, συχνά απαιτεί την αναγωγή της στα συμφραζόμενα για να αποσαφηνιστεί σε ποια έννοια αντιστοιχεί<sup>20</sup>. Πολλές φορές όμως, ακόμη και η αναγωγή στα άμεσα συμφραζόμενα, δεν είναι αρκετή για την αποσαφήνιση του νοήματος. Παραμένοντας στο παράδειγμα του σχήματος, η φράση «Είναι μια μαργαρίτα πάνω στο τραπέζι» δεν παρέχει αρκετές πληροφορίες για την αποσαφήνιση της έννοιας «μαργαρίτα» αφού μπορεί να αφορά τόσο στο ποτό μαργαρίτα όσο και στο λουλούδι. Και αυτό μπορεί να συμβαίνει όχι μόνο στην καθημερινή επικοινωνία αλλά και γραπτά που παράγονται από ειδικούς και έχουν ως αποκλειστικό στόχο την σαφή διατύπωση των νοημάτων. Σε μια σχετικά πρόσφατη

<sup>20</sup> Οι διφορούμενες λέξεις (ακόμη και αν είναι διαφορετικές αλλά μοιάζουν ηχητικά π.χ. ζει μόνη, ζυμώνει) αποτελούν το βασικό μηχανισμό για τη δημιουργία των λογοπαιγνίων.

μελέτη για νομικά κείμενα, ερευνητές από το MIT και το Πανεπιστήμιο του Εδιμβούργου κατέδειξαν τις ασάφειες που περιέχει η φυσική γλώσσα όταν αποτυπώνεται σε νομικές πράξεις όπως τα συμβόλαια (Martínez κ.ά., 2022).



Σχήμα 1: λέξεις και έννοιες βάσει της λογικής (της δεύτερης φάσης) του Wittgenstein

Ωστόσο τα συμφραζόμενα δεν έχουν να κάνουν μόνο με την ίδια τη γλώσσα αλλά και με το γενικότερο πολιτισμικό πλαίσιο, αφού κατά τη γλωσσική επικοινωνία υπάρχει ένα άρρητο υπόβαθρο πολιτισμικής πληροφορίας. Όπως πολύ γλαφυρά σημειώνει ο Pinker για τα συμφραζόμενα “consider how much knowledge of human behavior must be interpolated to understand what *he* means in a simple dialogue like this:

Woman: I’m leaving you.

Man: Who is he?”

(Pinker, 2007, σελ. 227)

Συνεπώς, γίνεται αντιληπτό πως στην γλωσσική επικοινωνία παρεμβαίνει πολιτισμική πληροφορία. Όμως αυτή η άρρητη γνώση, η οποία ενυπάρχει στα υποκείμενα, πολλές φορές δεν επιτρέπει την ανάγνωση του μηνύματος με τους όρους που επιθυμούσε ο δημιουργός του. Και εδώ δεν αναφερόμαστε σε μια απλή παρανόηση αλλά στην υποκειμενικότητα της ερμηνείας, η οποία στη μεταμοντέρνα εποχή γιγαντώθηκε, κυρίως μέσα από τη λογοτεχνική κριτική, και ιδίως μετά τον Jacques Derrida. Συνοπτικά, εν λόγω κίνημα, ισχυρίζεται ότι είναι αδύνατη η αντικειμενική επικοινωνία και ποτέ δεν είναι δυνατόν ο αποδέκτης του μηνύματος να γνωρίζει τι ακριβώς εννοούσε ο δημιουργός του. Μια σημαντική κριτική σε αυτή τη θεώρηση ασκεί ο Abrams, παραδεχόμενος ότι «[δεν] μπορούμε να εγγυηθούμε τη βεβαιότητα και την αλήθεια μιας ερμηνείας αναφερόμενοι σε ένα απόλυτο και διαχρονικό κριτήριο που

βρίσκεται πέρα από τις κανονικότητες των κοινών γλωσσικών μας πρακτικών. Ως εκ τούτου, η βεβαιότητά μας για μια ερμηνεία και την αλήθεια της δεν είναι ποτέ απόλυτη, δεδομένου ότι είναι πάντα πιθανόν (αν και σε ορισμένες περιπτώσεις η πιθανότητα είναι απειροελάχιστη) να κάνουμε λάθος» (Abrams, 2014, σελ. 73). Ωστόσο, στηλιτεύει την υπερβολή των θεωρητικών του εν λόγω κινήματος σημειώνοντας ότι «ο θεωρητικός τους κόσμος [...] είναι ακατοίκτης από τους ίδιους τους θεωρητικούς. Στην καθημερινή ζωή, όταν προκύπτει κάποιο θέμα [...] εγκαταλείπουν τη θεωρητική τους γλώσσα και υιοθετούν την κανονική ανθρωποκεντρική γλώσσα που αναφέρεται σε πρόσωπα που έχουν προθέσεις, σε ό,τι λένε και ό,τι εννοούν, στην πνευματική και ηθική ευθύνη που έχουν για τα λεγόμενά τους » (Abrams, 2014, σελ. 69).

Ένα ακόμη θέμα που λαμβάνεται υπόψη είναι ο συνεχής μετασχηματισμός της φυσικής γλώσσας και των εννοιών, είτε με την προσθήκη νέων λέξεων είτε με την αλλαγή του νοήματος που αποδίδεται στις υπάρχουσες. Αυτός είναι άλλωστε και ο λόγος που συχνά αποδίδεται στη γλώσσα ο χαρακτηρισμός *ζωντανή*. Αυτό κάνει τον θεωρητικό της ιστορίας των επιστημών Feyerabend να τονίσει πως χωρίς τη συνεχή εσφαλμένη χρήση (misuse) της γλώσσας δεν είναι δυνατό να συντελεστεί καμία ανακάλυψη, καμιά πρόοδος (Feyerabend, 2009, σελ. 18). Αυτή η διαδικασία είναι γνωστή με την ονομασία σημασιολογική μετατόπιση (semantic drift ή semantic change ή semantic shift). Ενδιαφέρον παράδειγμα από το χώρο της ανθρωπολογίας είναι το πώς οι νέες τεχνολογίες ή οι διακρατικές/ διαφυλετικές υιοθεσίες μετασχηματίζουν το περιεχόμενο των όρων της συγγένειας (Τουντασάκη, 2015, 2022).

Αντίστοιχος μηχανισμός υπάρχει και στη διαδικασία παραγωγής ρετρονύμων (retronym), δηλαδή στους νεολογισμούς όπου παρέχεται ένα νέο όνομα σε κάτι ώστε να διαφοροποιήσει το πρωτότυπο από μια πιο πρόσφατη μορφή ή εκδοχή. Για παράδειγμα ο όρος «τυπωμένο βιβλίο» προέκυψε αφού πρώτα υπήρξε ο όρος «ηλεκτρονικό βιβλίο». Έτσι η συνεχής αλλαγή της γλώσσας μάς αναγκάζει να αλλάζουμε τους όρους που παραπέμπουν στις παλιές έννοιες. Βέβαια η παράδοση χρήση τού να αποδίδεται μια νέα έννοια σε μία ήδη γνωστή (αλλά με άλλη έννοια) λέξη είναι πρωτίστως μηχανισμός του ανθρώπινου εγκεφάλου και δευτερευόντως της γλώσσας. Χαρακτηριστικό παράδειγμα αποτελεί το εικονίδιο που αναπαριστά τα emails, το οποίο κατά κανόνα παραπέμπει σε έναν παραδοσιακό φάκελο ταχυδρομείου. Αντίστοιχα, πολύ συχνά το εικονίδιο που ενεργοποιεί την τηλεφωνική κλήση σε ένα smart phone απεικονίζει το ακουστικό μιας τηλεφωνικής συσκευής της δεκαετίας του 1980<sup>21</sup>.

Μια λύση για την αντιμετώπιση των φαινομένων της πολυσημίας, αμφισημίας, συνωνυμίας κτλ είναι η χρήση φορμαλιστικών γλωσσών, αφού όπως σημειώνει ο Tobias Kuhn, ειδικός σε θέματα Ελεγχόμενων Φυσικών Γλωσσών (Controlled Natural Languages), η φυσική γλώσσα είναι πολύ ανακριβής υπό την έννοια ότι απαιτείται πλήθος συμφραζομένων ώστε να κατανοηθεί το νόημα τυπικών προτάσεων. Οι τυπικές

---

<sup>21</sup> Η τάση σχεδιασμού ενός αντικειμένου με τρόπο που να θυμίζει αντικείμενο άλλου υλικού ή άλλου σχεδίου ονομάζεται σκευομορφισμός (skeuomorphism). Για παράδειγμα ένας ηλεκτρικός λαμπτήρας που έχει το σχήμα ενός κεριού.

λογικές γλώσσες, από την άλλη πλευρά, μπορούν να έχουν τη μέγιστη δυνατή ακρίβεια, επειδή το νόημά τους ορίζεται αυστηρά μόνο με βάση τις πιθανές ακολουθίες των γλωσσικών συμβόλων τους (Τ. Κιηη, 2014, σελ. 128). Έτσι η χρήση μιας εννοιοκεντρικής φορμαλιστικής γλώσσας είναι καθοριστικής σημασίας στην αποτύπωση του περιεχομένου διότι οι έννοιες -και όχι οι λέξεις- είναι κυρίαρχο συστατικό του περιεχομένου και του νοήματος.

## 5.2 Φυσική γλώσσα, γραφή και ανθρωπολογική γνώση

Πριν εξεταστούν τεχνητές γλώσσες, επιχειρείται μια αποτύπωση της σημασίας που έχει, ειδικά για την ανθρωπολογία, η φυσική γλώσσα ως μέσο για την παραγωγή -και κυρίως- για την αποτύπωση της ανθρωπολογικής γνώσης. Ο Van Der Leeuw τονίζει πως ένα από τα θεμελιώδη προβλήματα των κοινωνικών και ανθρωπιστικών επιστημών είναι το γεγονός ότι, επειδή μελετούν κοινωνικά φαινόμενα, οι περιγραφές τους είναι στενά συνδεδεμένες με συγκεκριμένες φυσικές γλώσσες (Van Der Leeuw, 2004, σελ. 123). Η κοινωνική ανθρωπολογία δεν αποτελεί εξαίρεση. Η ετυμολογία της λέξης ανθρωπολογία (τόσο στα ελληνικά όσο και στα αγγλικά) παραπέμπει στο «λόγος». Ενώ η ετυμολογία της λέξης εθνογραφία, τόσο στα ελληνικά όσο και στα αγγλικά, παραπέμπει στη «γραφή»<sup>22</sup>. Οι ανθρωπολόγοι όχι μόνο δεν αμφισβήτησαν την ηγεμονία της φυσικής γλώσσας, τόσο κατά τη διαδικασία παραγωγής όσο και στην αποτύπωση του έργου τους, αλλά μάλλον ενδυνάμωσαν τη θέση της. Ο Mueller τονίζει πως η «εθνογραφία, τόσο ως διαδικασία όσο και ως προϊόν, παραμένει αποφασιστικά λεκτική και γραφοκεντρική (scriptocentric)» (Mueller, 2016, σελ. 99). Η ανθρωπολόγος Orin Starn σημειώνει «Μέχρι τώρα, το γράψιμο είναι αναπόφευκτα μέρος του εαυτού μου -στην πραγματικότητα, δεν ξέρω πώς να κάνω κάτι άλλο. Ο παππούς μου ήταν μηχανικός- έφτιαχνε αυτοκίνητα. Εγώ είμαι ανθρωπολόγος- γράφω» (Starn, 2022). Μέσα σε αυτό το πλαίσιο δεν είναι τυχαίο ότι ένα από τα επιδραστικότερα βιβλία της ανθρωπολογίας, με χιλιάδες αναφορές, είναι το *Writing culture* (Clifford & Marcus, 1986).

Παρότι αρκετοί αναφέρονται στα κείμενα αυτού του συλλογικού έργου ως το σημείο αναφοράς για την κρίση της αναπαράστασης (the crisis of representation), η έμφαση στην κειμενική αναπαράσταση, σαφώς επηρεασμένη από τις θέσεις του Jacques Derrida, πυροδότησε μια στροφή της ανθρωπολογίας προς τις ανθρωπιστικές επιστήμες, όπως αυτές εκφράζονται στα πλαίσια της μεταμοντέρνας θεώρησης, οδηγώντας σε σχετική απομάκρυνση από τις κοινωνικές επιστήμες. Στον επίλογο του *Writing Culture*, στο κείμενο που υπογράφει ο ίδιος ο Marcus, σημειώνει: η «κειμενοποίηση είναι στην καρδιά του εθνογραφικού εγχειρήματος, τόσο στο πεδίο όσο και στο πανεπιστήμιο» (Clifford & Marcus, 1986, σελ. 264). Ενώ ο James Clifford, στον ίδιο τόμο, στο άρθρο του με τίτλο *On Ethnographic Allegory* σημειώνει για την εθνογραφική πρακτική ότι «αποτελεσματικά κειμενοποιεί τον άλλο» (Clifford & Marcus, 1986, σελ. 111) για να

<sup>22</sup> Η ετυμολογία βεβαίως και δείχνει την ιστορία ενός όρου αλλά δεν προεξοφλεί την πορεία του επιστημονικού πεδίου στο οποίο αναφέρεται. Για παράδειγμα και η βιολογία ετυμολογικά παραπέμπει στο λόγο αλλά τόσο η έρευνα όσο και η αναπαράστασή της δεν είναι κειμενοκεντρικές.

επαναλάβει τη θέση του δύο σελίδες παρακάτω αναφέροντας ότι η «πρακτική της εθνογραφίας είναι μια διαδικασία γραφής, συγκεκριμένα κειμενοποίησης». Και στην αποτίμηση της επίδρασης του εν λόγω έργου διαπιστώνεται, από τον ίδιο τον Marcus, είκοσι χρόνια μετά, ότι δεν μετατοπίζεται καθόλου το σημείο εστίασης από την κειμενοκεντρική αναπαράσταση της εθνογραφίας (Marcus, 2007). Ακόμη ένας συλλογικός τόμος που πραγματεύεται την ιστορία της κειμενοποίησης στην ανθρωπολογία είναι το «*Modernist anthropology: From fieldwork to text*». Στο εν λόγω έργο αφιερώνεται μια ολόκληρη ενότητα για την *Εθνογραφία ως Λόγος (Ethnography as Discourse)*, η οποία μας εισάγει στην εποχή της μονογραφίας (Manganaro, 1990).

Κατά συνέπεια, όχι το σύνολο της φυσικής γλώσσας αλλά συγκεκριμένα η κειμενοποίησή της γίνεται ο βασικός φορέας για την παραγωγή και αποτύπωση της ανθρωπολογικής γνώσης. Είναι όμως αυτό τόσο αυτονόητο και φυσικό επακόλουθο για την ανθρωπολογική έρευνα; Στη συνέχεια θα σταθούμε κριτικά απέναντι στην άποψη ότι αυτή η προσέγγιση είναι μονόδρομος, εξετάζοντας κατά πόσο οι εκφραστικές δυνατότητες της γραφής αποτελούν το μοναδικό μέσο για την αναπαράσταση της φυσικής γλώσσας, σε αναλογία με την πληροφορική ως γλώσσα αναπαράστασης της πληροφορίας.

### 5.2.1 Εκφραστικότητα: η γραφή ως αναπαράσταση της φυσικής γλώσσας και η πληροφορική ως γλώσσα αναπαράστασης πληροφορίας

Η γραφή -σε αντίθεση με τη φυσική γλώσσα- είναι μια συνειδητή ανακάλυψη του ανθρώπου και όχι μια βιολογική εξέλιξη. Έστω, λοιπόν, ένας κόσμος όπου δεν υπάρχει γραφή. Ως εκ τούτου απουσιάζει ο βασικός μηχανισμός για τη δημιουργία κειμένου. Σε αυτόν τον κόσμο γίνεται προσπάθεια να εισαχθεί ένα σύστημα γραφής αντίστοιχο με αυτά που χρησιμοποιούνται σήμερα στο δυτικό κόσμο. Τώρα ας επιχειρήσουμε να τεκμηριώσουμε πάνω στο γιατί η εκφραστικότητα αυτού του μοντέλου<sup>23</sup> αναπαράστασης της γλώσσας δεν είναι επαρκώς εκλεπτυσμένη, εστιάζοντας στις αδυναμίες που αυτή έχει σε σχέση με τους ήχους της φυσικής γλώσσας. Καταρχάς, αδυνατεί να αναπαραστήσει δύο πολύ βασικά χαρακτηριστικά της ομιλούμενης γλώσσας: την ένταση του ήχου (decibel) και την ένταση στη φωνή, ως ψυχολογική ένταση (tension). Επιπλέον δεν διαθέτει μηχανισμό για να αποδώσει ζητήματα που σχετίζονται με τη χροιά, την προφορά ή τον τόνο της φωνής και πολλά ακόμη χαρακτηριστικά που διαθέτει ο ήχος και έχει τη δυνατότητα να διακρίνει η ανθρώπινη ακοή. Από την άλλη διαθέτει κάποια, πολύ απλοϊκά, σύμβολα, όπως το θαυμαστικό και

---

<sup>23</sup> Στη γλωσσολογία η γραφή θεωρείται σύστημα και όχι μοντέλο. Μια απλουστευμένη διάκριση ανάμεσα στο σύστημα και το μοντέλο είναι ότι το σύστημα είναι αυτό που υπάρχει και το μοντέλο ο τρόπος που βλέπουμε αυτό που υπάρχει. Όμως η γραφή παρότι έχει τα χαρακτηριστικά του συστήματος, μπορεί ταυτόχρονα να θεωρηθεί ως ένα μοντέλο για την αποτύπωση των ήχων ή των εννοιών. Με άλλα λόγια, κάποιες φορές ένα σύστημα μπορεί να ιδωθεί ως μοντέλο -και αντίστροφα- αφού, στην περίπτωση μας, η γραφή ενώ είναι μοντέλο 'αντικειμενοποιείται' από τη γλωσσολογία και εξετάζεται ως σύστημα.

το ερωτηματικό, τα οποία όμως ούτε καν την πιθανή ειρωνεία μιας ερώτησης δεν μπορούν να αποτυπώσουν. Επίσης δεν υπάρχει τρόπος να αποδοθούν οι παύσεις μεταξύ των λέξεων ή η διάρκεια των συλλαβών κλπ<sup>24</sup>. Αναμενόμενη, και μάλλον αιτιολογημένη, αντίδραση θα ήταν ο προβληματισμός γιατί να χρησιμοποιήσουμε μια αναπαράσταση που μπορεί να κάνει πολύ λιγότερα (δηλαδή έχει πολύ μικρότερη εκφραστική ικανότητα) από αυτήν που ήδη εκφράζει ο προφορικός λόγος;

Στα παραπάνω επιχειρήματα ενάντια στη γραφή θα μπορούσε να προστεθεί η προσέγγιση του Galeano, ο οποίος στο κείμενό του με τίτλο «Δεν χρειαζόμαστε τη γραφή» σημειώνει: «Περίπου πέντε χιλιάδες χρόνια πριν από τον Σαμπολιόν, ο θεός Θωτ ταξίδεψε στις Θήβες και πρόσφερε στο βασιλιά Θάμφθυ της Αιγύπτου την τέχνη της γραφής. Του εξήγησε τα ιερογλυφικά, και είπε πως η γραφή ήταν το καλύτερο γιατρικό για τη μνήμη, και πως έδινε γνώση. Ο βασιλιάς δεν δέχτηκε το δώρο. 'Μνήμη; Γνώση; Η ανακάλυψη αυτή θα φέρει τη λήθη. Η γνώση βρίσκεται στην αλήθεια, όχι στην εξωτερική της εμφάνιση. Δεν μπορείς να θυμηθείς με ξένη μνήμη. Οι άνθρωποι θα καταγράψουν, αλλά δεν θα θυμούνται. Θα επαναλαμβάνουν, αλλά δεν θα ζουν. Θα πληροφορούνται πολλά πράγματα, αλλά δεν θα γνωρίζουν τίποτα.'» (Galeano, 2008, σελ. 16).

Αφήνοντας πίσω τον υποθετικό κόσμο και τον κόσμο της μυθολογίας, εύκολα αντιλαμβάνεται κανείς ότι στο σύγχρονο (modern) δυτικό κόσμο επετεύχθη η πλήρης αντιστροφή των παραπάνω επιχειρημάτων. Η θέση της γραφής από την Αναγέννηση και μετά είναι τόσο θεμελιώδης για τον άνθρωπο της Δύσης που ο Φουκώ γράφει: «Εφεξής πρώτη φύση της γλώσσας είναι ο γραπτός της χαρακτήρας. Οι ήχοι της μεταγραφής αποτελούν την μεταβατική και πρόσκαιρη μετάβασή της»<sup>25</sup> (Foucault, 1986, σελ. 73).

Ο συγγραφέας της ανά χειράς διατριβής εκτιμά ότι, τηρουμένων των αναλογιών, το ίδιο ισχύει και με την πληροφορική. Η αναπαράσταση αλλά και η επεξεργασία των δεδομένων μέσω των υπολογιστών έχει πολλούς περιορισμούς σε σχέση με τα όσα μεταφέρουμε με τη φωνή ή γράφουμε στα χαρτιά μας. Όμως αποτελεί μια νέα 'γλώσσα'

<sup>24</sup> Αυτή η μειωμένη εκφραστικότητα της γραφής οδήγησε τη σύγχρονη γλωσσολογία, και ειδικά τον τομέα της φωνολογίας, στην ανάπτυξη εξειδικευμένων «αλφαβήτων» τα οποία επιτρέπουν την αποτελεσματικότερη καταγραφή των φωνημάτων.

<sup>25</sup> Ας εφαρμόσουμε το κύρος και τη σοβαρότητα των θεωριών του Φουκώ στα ευτράπελα του σύγχρονου ελληνικού διαδικτύου. Τα τελευταία χρόνια εμφανίζεται συχνά στα ελληνικά κοινωνικά δίκτυα, συνήθως ως αστειό, η φράση «ΟΝΟ ΟΝΟ ΗΤΟ». Και το αστειό βρίσκεται στο ότι πρόκειται για μια σχετικά πιστή ηχητική μεταγραφή της φράσης «ο νοών νοείτω». Όμως δεν είναι απλά γραμμένη ανορθόγραφα αλλά δεν συνδέει καν τα φωνήματα με τις λέξεις στις οποίες αυτά αντιστοιχούν. Και εδώ θα ήθελα να σημειώσω, με ένα παράδειγμα από προσωπικό βίωμα, πώς η αναγνώριση των λέξεων, εντός μιας πρότασης, δεν είναι εγγενής ικανότητα ενός φυσικού ομιλητή μιας γλώσσας. Η γιαγιά μου ήταν Μικρασιάτισσα και μιλούσε δύο γλώσσες: Ελληνικά και Τούρκικα. Ταυτόχρονα ήταν παντελώς αναλφάβητη, αλλά αρκετά ευφυής, για τις προσλαμβάνουσές της, άνθρωπος. Όντας μαθητής δημοτικού, μου ζήτησε να την βοηθήσω να μάθει γράμματα! Εκείνη είχε περάσει ήδη τα ογδόντα. Θυμάμαι τον οχτάχρονο εαυτό μου να απορεί με την αδυναμία της όχι μόνο να ξεχωρίσει τους επιμέρους φθόγγους εντός μιας λέξης αλλά, κάποιες φορές, και τις λέξεις μεταξύ τους. Όμως στην καθημερινή της επικοινωνία ουδείς μπορούσε να διακρίνει κάποιο πρόβλημα, η αδυναμία της αυτή ουδόλως επηρέαζε την ικανότητα παραγωγής, γλωσσικά έγκυρου, λόγου. Και «όνο όνο ήτο»...!

και ανοίγει έναν νέο κόσμο επεξεργασίας και αποτύπωσης ο οποίος πιθανότατα θα επιφέρει μεγαλύτερες αλλαγές από όσες έφερε η ίδια η γραφή.

Λαμβάνοντας υπόψη τους ίδιους προβληματισμούς με τη γραφή, η ανάπτυξη των σύγχρονων πληροφοριακών συστημάτων οφείλει να λάβει υπόψη τα επιμέρους δεδομένα και κατά πόσο τα εν λόγω συστήματα θα τα υπηρετούν, ώστε να προκύψουν διαφορετικά «αλφάβητα» που θα καλύπτουν ανάλογες ανάγκες, με την ίδια λογική που δεν έχει προκύψει κοινό αλφάβητο για όλες τις γλώσσες, αλλά διαφορετικά, ανάλογα με τις ανάγκες των φωνημάτων κάθε γλώσσας. Αν όμως προχωρήσουμε το συλλογισμό σχετικά με τη γραφή, θα δούμε ότι δεν αφορά μόνο σε φυσικές γλώσσες. Όταν, σε κάποιες περιπτώσεις, διαπιστώθηκε ότι η φυσική γλώσσα δεν μπορεί να διαθέσει αρκετά υψηλό φορμαλισμό ώστε να αναπαρασταθούν με ακρίβεια τα ευρήματα ή οι υποθέσεις ενός πεδίου, τότε διάφοροι τομείς ανέπτυξαν δικές τους γλώσσες, π.χ. μαθηματικά ή γλώσσες πληροφορικής -είτε γλώσσες προγραμματισμού (programming languages) είτε γλώσσες επισήμανσης (markup languages)- οι οποίες διαθέτουν δικές τους γραφές και συντακτικό. Αντίστοιχα θα πρέπει οι επιμέρους τομείς κάθε επιστήμης να αναπτύξουν τις δικές τους “γραφές”, ώστε να μπορέσουν να αξιοποιήσουν σε μεγαλύτερο βαθμό τις δυνατότητες της πληροφορικής. Η αύξηση θα είναι εκθετική όταν οι επιμέρους επιστήμες κατανοήσουν τις βασικές αρχές της μοντελοποίησης που απαιτεί η πληροφορική και αρχίσουν, όχι απλά να τη χρησιμοποιούν, αλλά να την επεκτείνουν για να καλύψουν τις δικές τους ανάγκες. Δηλαδή το θέμα δεν είναι μόνο τι μπορεί να κάνει η πληροφορική για την ανθρωπολογία αλλά και τι μπορεί να προσφέρει η ανθρωπολογία στην πληροφορική, ώστε να υπάρξει μια αμφίδρομη επικοινωνιακή σχέση.

Επομένως, το ζητούμενο δεν είναι η χρήση ενός λογισμικού επεξεργασίας κειμένου για τη δημιουργία του παρόντος κειμένου ή ενός στατιστικού πακέτου για την ποσοτική ανάλυση ενός όγκου δεδομένων. Το ζητούμενο είναι πώς τα ίδια τα ανθρωπολογικά δεδομένα θα γίνουν αντικείμενο υπολογιστικής επεξεργασίας. Ένας πρόσφορος τομέας για να επιτευχθεί αυτό είναι η αναπαράσταση γνώσης όπως νοηματοδοτείται και υλοποιείται στο περιβάλλον των γράφων γνώσης και του σημασιολογικού ιστού.

### 5.3 Σημασιολογικός ιστός και διασυνδεδεμένα δεδομένα: ως φορείς νοήματος

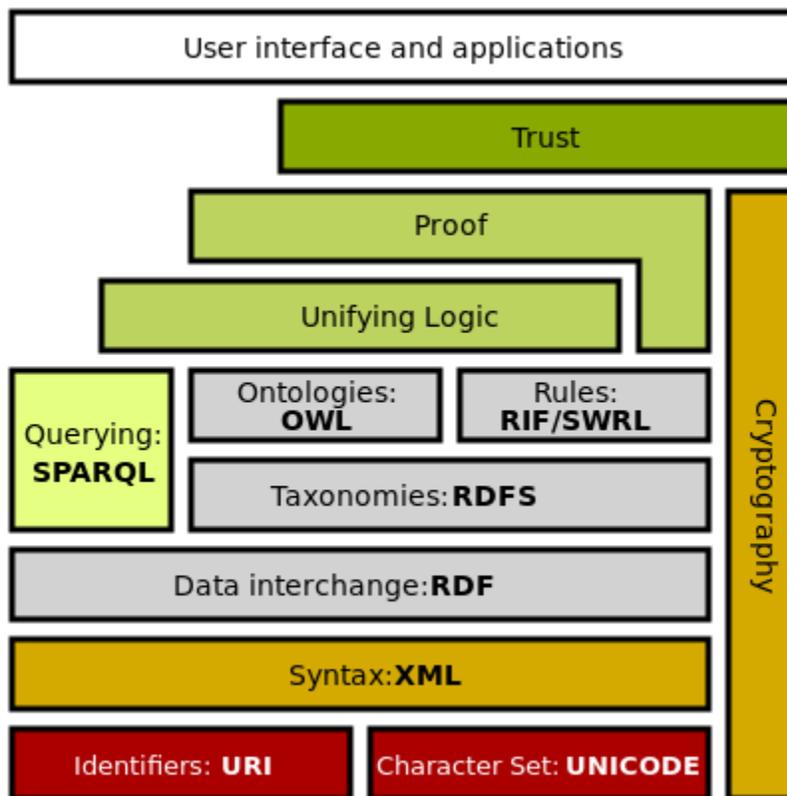
Στις αρχές του 21<sup>ου</sup> αιώνα υπάρχει αυξανόμενο ενδιαφέρον για τη μοντελοποίηση ενώ η τεχνητή νοημοσύνη κάνει την εμφάνισή της όχι μόνο ως θεωρητική προοπτική αλλά και ως υλοποίηση<sup>26</sup>. Οι Antoniou & Van Harmelen θεωρούν ότι υπάρχουν δύο βασικές προσεγγίσεις για την επίτευξη της τεχνητής νοημοσύνης. Η πρώτη λύση είναι η ανάπτυξη πολύ εξελιγμένων τεχνικών για την ανάλυση των δεδομένων. Η εναλλακτική λύση

---

<sup>26</sup> Σε αυτό το κεφάλαιο παρουσιάζεται μια συγκεκριμένη οπτική σχετικά με την τεχνητή νοημοσύνη. Η σχέση μεταξύ διαφορετικών προσεγγίσεων και μοντέλων με τεχνητής νοημοσύνης παρουσιάζεται στην ενότητα [Generative AI, Large Language Models, Deep Learning και μορφές Τεχνητής Νοημοσύνης](#).

εστιάζει στη μορφή των δεδομένων και προκρίνει τη δόμησή τους με τρόπο που να διευκολύνει την επεξεργασία, άρα η δομή των δεδομένων είναι εκείνη που καθιστά ικανή την εξελιγμένη επεξεργασία. Η δεύτερη οπτική είναι εκείνη που οδηγεί στο σημασιολογικό ιστό (Antoniou & Van Harmelen, 2008, σελ. 3).

Ο σημασιολογικός ιστός μπορεί να ιδωθεί ως μια τεχνητή γλώσσα για την απόδοση νοημάτων. Ο βασικός λόγος της ανάπτυξης μιας τεχνητής γλώσσας είναι το γεγονός ότι η φυσική γλώσσα δεν διαθέτει υψηλό βαθμό ακρίβειας αφού απαιτείται μεγάλος αριθμός συμφραζομένων ώστε να αποκτήσει αδιαμφησβήτητο νόημα μια πρόταση. Σε αντίθεση, οι φορμαλιστικές λογικές γλώσσες (formal logic languages) διαθέτουν μέγιστη ακρίβεια, διότι το νόημά τους ορίζεται αυστηρά αποκλειστικά στη βάση των πιθανών ακολουθιών των συμβόλων τους (T. Kuhn, 2014, σελ. 128).



Σχήμα 2: semantic web stack

Ο Σημασιολογικός Ιστός (Semantic Web) ή ιστός των δεδομένων (web of data) ή web 3.0<sup>27</sup> είναι μια επέκταση του σημερινού διαδικτύου η οποία στοχεύει στη δομή των

<sup>27</sup> Ο όρος web 1 αφορά στην πρώτη φάση εξάπλωσης του διαδικτύου όπου η δυνατότητα αλληλεπίδρασης των τελικών χρηστών με το περιεχόμενο ήταν περιορισμένη έως και ανύπαρκτη. Το web 2 αφορά στην επόμενη εκδοχή του διαδικτύου όπου, κυρίως μέσω τεχνολογιών που επέτρεψαν την

δεδομένων. Στόχος είναι να δοθεί έμφαση σε δομές δεδομένων οι οποίες θα είναι αποτελεσματικότερα επεξεργάσιμες από τους υπολογιστές, σε αντιδιαστολή με το σημερινό κειμενοκεντρικό διαδίκτυο. Στη συνέντευξή στον Brian Runciman ο 'εφευρέτης' του web όπως το γνωρίζουμε σήμερα, οραματιστής του σημασιολογικού ιστού και δημιουργός του σχήματος 2 λέει «The Web is now philosophical engineering». (Berners-Lee, 2006). Διότι, όπως είναι εμφανές στο Σχήμα 2 να μεν στην κορυφή του σημασιολογικού ιστού βρίσκονται οι διεπαφές του χρήστη και τα λογισμικά αλλά όλα αυτά εδράζονται στην εμπιστοσύνη, η οποία βασίζεται στην απόδειξη, η οποία με τη σειρά της βασίζεται σε μια ενοποιημένη λογική. Ενώ το «engineering» τμήμα βρίσκεται από εκεί και κάτω. Και η λογική του είναι τέτοια που μπορεί να υποστηρίξει τα παραπάνω.

Πολλοί συγγραφείς χρησιμοποιούν ως συνώνυμα τους όρους σημασιολογικός ιστός, ιστός των δεδομένων (web of data) και διασυνδεδεμένα δεδομένα (linked data). Από την άλλη οι Machado, Souza, & Simões αφιερώνουν μια ολόκληρη δημοσίευση για τις διαφορές τους (καθώς και τη διολίσθηση του εκάστοτε ορισμού μέσα στο χρόνο) και συνοψίζουν, για τα διάφορα ονόματα, πως ο στόχος είναι να γίνει το διαδίκτυο μια παγκόσμια βάση δεδομένων γνωστή ως ιστός των δεδομένων (web of data). Οι προδιαγραφές που αναπτύχθηκαν ή συσχετίστηκαν με αυτόν το σκοπό και το σύνολο διαδικασιών και κανόνων προς την επίτευξη του στόχου είναι γνωστά ως συνδεδεμένα δεδομένα (linked data). Τα διασυνδεδεμένα δεδομένα καθιστούν εφικτή τη δημοσίευση των δεδομένων με τρόπο που επιτρέπει τη σύνδεση προς άλλα δεδομένα από διαφορετικές πηγές και κωδικοποιημένα σε διαφορετικά μορφότυπα (Machado κ.ά., 2019, σελ. 712).

Άρα τα διασυνδεδεμένα δεδομένα (linked data) είναι η μέθοδος σύνδεσης των δομημένων δεδομένων, και παρέχουν ένα ανεκτίμητο εργαλείο για να αναπτύσσεται από κοινού, όχι κείμενο, αλλά δομημένη πληροφορία. Η αναπαράσταση γνώσης και οι τεχνολογίες σημασιολογίας παρέχουν τα μέσα για την ανάδειξη των διασυνδεδεμένων δεδομένων σε ουσιαστικές δηλώσεις αφού επικοινωνούν την επιθυμητή σημασία η οποία είναι αναγκαία για την κατανόηση αυτών των δηλώσεων και των συσχετίσεών τους, αναφορικά όχι μόνο με την περιγραφή, αλλά και με τα συμφραζόμενα και την προέλευση (Oldman κ.ά., 2016, σελ. 267).

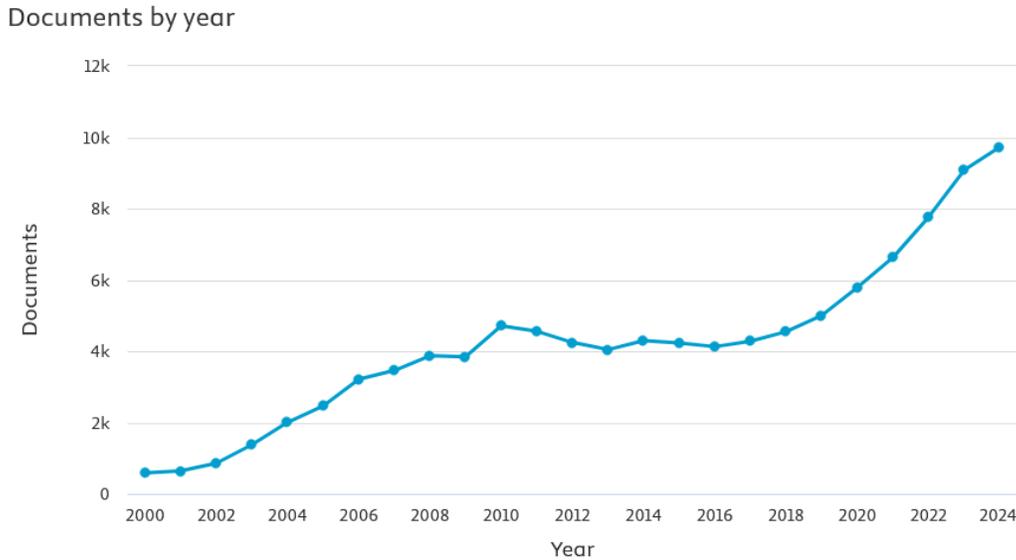
Αναλύοντας τις επιστημονικές δημοσιεύσεις του 21<sup>ου</sup> αιώνα, γύρω από τα παραπάνω θέματα, είναι σαφές ότι υπάρχει ένα αυξανόμενο επιστημονικό ενδιαφέρον. Όπως δείχνει το Γράφημα 3, οι μόλις 586 δημοσιεύσεις για το έτος 2000 ανεβαίνουν σχεδόν στις 10.000 το έτος 2024. Το σύνολο των επιστημονικών δημοσιεύσεων που έχει ευρετηριάσει η βιβλιογραφική βάση δεδομένων Scopus και αφορά στον 21<sup>ο</sup> αιώνα<sup>28</sup> είναι λίγο πάνω από 105 χιλιάδες εγγραφές.

---

άνθηση των social media και ανάπτυξη πλατφορμών όπως το YouTube και το Facebook, οι τελικοί χρήστες μπορούν εύκολα να δημιουργήσουν και να διαθέσουν οι ίδιοι το περιεχόμενο που επιθυμούν.

<sup>28</sup> Το γράφημα προέκυψε από αναζήτηση στη Scopus το Μάιο του 2025. Η αναζήτηση επέστρεψε 105.377 εγγραφές. Το ερώτημα ήταν ( TITLE-ABS-KEY ( "knowledge graph\*" ) OR

Το Γράφημα 4<sup>29</sup> δείχνει το δίκτυο συσχετίσεων που δημιουργούν οι 50 πιο διαδεδομένοι όροι που περιέχονται στις 105 χιλιάδες εγγραφές του συνόλου δεδομένων. Όπως είναι προφανές δημιουργούνται δύο βασικές συστάδες (clusters) όπου η μία αφορά στις τεχνολογίες και η άλλη στους ανθρώπους. Σημειώνεται ότι το μέγεθος κάθε κόμβου προκύπτει ανάλογα με τον αριθμό των εγγραφών που αντιστοιχούν σε αυτόν και το πάχος κάθε γραμμής στον αριθμό των συσχετίσεων.



**Γράφημα 3:** επιστημονικές δημοσιεύσεις -ανά έτος- για *linked data*, *semantic web*, *ontologies* και *Knowledge representation* (πηγή Scopus)

---

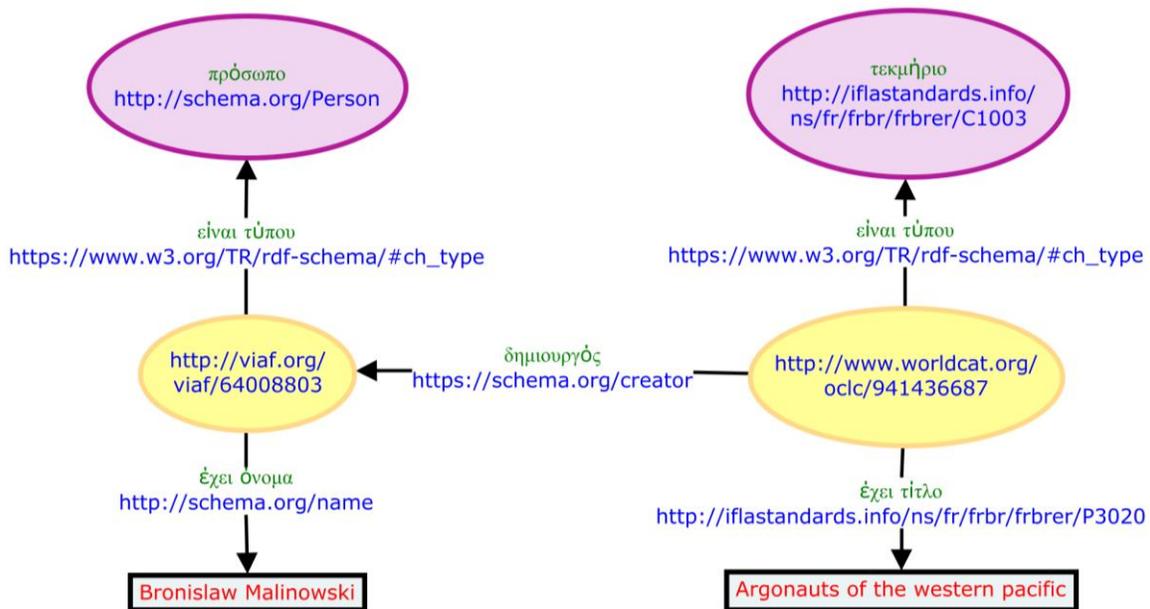
```
TITLE-ABS-KEY ( "semantic web" ) OR TITLE-ABS-KEY ( "linked data" ) OR
TITLE-ABS-KEY ( "Knowledge representation" ) ) AND PUBYEAR > 1999
AND PUBYEAR < 2025.
```

<sup>29</sup> Το γράφημα δημιουργήθηκε μέσω το λογισμικού VOSviewer το οποίο επιτρέπει την κατασκευή και οπτικοποίηση βιβλιογραφικών δικτύων. Τονίζεται ότι το δίκτυο συσχετίσεων δεν αφορά στους όρους εντός των κειμένων των δημοσιεύσεων αλλά στα μεταδεδομένα αυτών και συγκεκριμένα στα πεδία τίτλος, περίληψη και λέξεις κλειδιά.



παραλληλόγραμμα αφορούν τιμές κειμένου οι οποίες δεν μπορούν να αποτελέσουν υποκείμενο για νέες σχέσεις, δηλαδή αφορούν στα ονόματα των οντοτήτων.

Με μια πρώτη ανάγνωση, στο Σχήμα 3 δηλώνεται ότι δημιουργός του δεξιού κίτρινου κόμβου είναι ο αριστερός κίτρινος κόμβος, δηλαδή το <http://viaf.org/viaf/64008803> δημιούργησε το <http://www.worldcat.org/oclc/941436687>. Όμως δεν απεικονίζεται αυτό στο σχήμα. Το σχήμα δείχνει (χωρίς δυνατότητα παρερμηνείας) ότι αυτό που υπάρχει πίσω από την πρώτη διεύθυνση (δηλαδή ο Μαλινόφσκι) δημιούργησε αυτό που υπάρχει πίσω από τη δεύτερη (δηλαδή τους Αργοναύτες του Ειρηνικού). Ο λόγος που ισχύει αυτό παρουσιάζεται πολύ εύληπτα από τον Sowa, έναν από τους πρωτοπόρους της τεχνητής νοημοσύνης. Όπως επισημαίνει, η αναπαράσταση γνώσης είναι από τη φύση της ένα υποκατάστατο. Φυσικά αντικείμενα, γεγονότα ή σχέσεις, βεβαίως και δεν μπορούν να αποθηκευτούν απευθείας μέσα σε έναν υπολογιστή. Όμως αυτά μπορούν να αναπαρασταθούν από σύμβολα και αυτά τα σύμβολα να λειτουργήσουν ως υποκατάστατα για τα εξωτερικά πράγματα (external things). Διαχειριζόμενο αυτά τα εσωτερικά υποκατάστατα (internal surrogates) το λογισμικό μπορεί να προσομοιώσει το εξωτερικό σύστημα ή να συμπεράνει (reasoning) για αυτό (Sowa, 2009, σελ. 134). Άρα επανερχόμενοι στο Σχήμα 3, έχουμε από μία διεύθυνση στο διαδίκτυο να λειτουργεί ως υποκατάστατο είτε μιας κατηγορίας (κλάσης), είτε μιας οντότητας (στιγμιότυπο κάποιας κλάσης), είτε ως σχέση (μεταξύ κλάσεων ή μεταξύ στιγμιότυπων).



Σχήμα 3: παράδειγμα δόμησης δεδομένων σε RDF (γραφική απεικόνιση).

**Σημείωση:** Τα πράσινα γράμματα δεν αποτελούν μέρος της κωδικοποίησης των δεδομένων, καταχωρήθηκαν για τη διευκόλυνση της ανάγνωσης του γραφήματος<sup>31</sup>.

Ακολουθεί -στο Απόσπασμα κώδικα 2- ο RDF/XML κώδικας που κωδικοποιεί τις πέντε τριπλέτες που εμφανίζονται στο Σχήμα 3.

Μια μεταφορά των πέντε τριπλετών από το Σχήμα 3 σε κείμενο θα ήταν: ο Μαλινόφσκι, που είναι πρόσωπο, δημιούργησε ένα τεκμήριο που έχει τίτλο «*Αργοναύτες του Ειρηνικού*». Αυτό που έχει ιδιαίτερη σημασία είναι ότι οι λέξεις «πρόσωπο», «τεκμήριο», «δημιούργησε», «έχει τίτλο» κτλ (όλα όσα εμφανίζονται με πράσινη γραμματοσειρά) είναι σαφώς ορισμένες από αντίστοιχα πρότυπα στα σχετικά URIs<sup>32</sup>. Οι δύο κόμβοι με κίτρινο χρώμα δείχνουν τους δύο βασικούς πόρους που διαθέτει ο γράφος, ήτοι την εγγραφή για τον Μαλινόφσκι (<http://viaf.org/viaf/64008803>) από το *Virtual International Authority File* και την εγγραφή για το βιβλίο *Argonauts of the western pacific* (<http://www.worldcat.org/oclc/941436687>) από τον διεθνή συλλογικό κατάλογο WorldCat. Οι κατηγορίες (οι κόμβοι που εμφανίζονται με μωβ χρώμα στο σχήμα) δηλώνουν κατηγορίες οντοτήτων και όχι μεμονωμένες οντότητες. Σε κάποιες γλώσσες (όπως η OWL) αυτή η διάκριση είναι πάντα άμεσα δηλωμένη.

---

<sup>31</sup> Ηθελμένα δεν επιλέχθηκαν ιδιότητες και κλάσεις από ένα συγκεκριμένο σχήμα μεταδεδομένων για την περιγραφή. Ο στόχος του παραδείγματος είναι να δείξει την ευελιξία στην επιλογή ιδιοτήτων κατά τη δημιουργία γράφων γνώσης και όχι στο να παρουσιάσει μια συνεκτική περιγραφή συμβατή με μια οντολογία ή κάποιο συγκεκριμένο σχήμα μεταδεδομένων.

<sup>32</sup> Το «πρόσωπο» είναι αυτό που ορίζει το <http://schema.org/Person>, το «τεκμήριο» είναι αυτό που ορίζεται από το <http://iflastandards.info/ns/fr/frbr/frbrer/C1003>, το «δημιούργησε», είναι αυτό που ορίζεται από το <http://purl.org/dc/terms/creator>, το «έχει τίτλο» είναι αυτό που ορίζεται από το <http://iflastandards.info/ns/fr/frbr/frbrer/P3020>

## Απόσπασμα κώδικα 2: το Σχήμα 3 σε RDF/XML κώδικα

```

<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:schema="https://schema.org/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:frbrer="http://iflastandards.info/ns/fr/frbr/frbrer#">
<rdf:Description rdf:about="http://www.worldcat.org/oclc/1129737708">
<frbrer:P3020>Argonauts of the western pacific </frbrer:P3020>
<rdf:type
rdf:resource="https://www.iflastandards.info/fr/frbr/frbrer#C1003"/>
<dcterms:creator rdf:resource="http://viaf.org/viaf/64008803"/>
</rdf:Description>
<rdf:Description rdf:about="http://viaf.org/viaf/64008803">
<schema:name>Bronislaw Malinowski</schema:name>
<rdf:type rdf:resource="http://schema.org/Person"/>
</rdf:Description>
</rdf:RDF>

```

Άρα, με κίνδυνο υπεραπλούστευσης, μπορεί να ειπωθεί ότι με την RDF και την RDFs είναι δυνατή η αποτύπωση όσων επιθυμούμε να εκφράσουμε σε μια μορφή η οποία είναι μηχανικά επεξεργάσιμη και ταυτόχρονα πολύ κοντά σε όσα μπορεί να εκφράσει η φυσική γλώσσα. Δηλαδή πρόκειται για ένα συντακτικό το οποίο επιτρέπει τη δόμηση της πληροφορίας, αποτυπώνοντας σχέσεις μεταξύ των οντοτήτων και των κατηγοριών (κλάσεων) στις οποίες αυτές ανήκουν. Το σύγχρονο τεχνολογικό περιβάλλον όμως παρέχει και ένα υψηλότερο επίπεδο αφαίρεσης από αυτό. Πρόκειται για την αποτύπωση των κανόνων οι οποίοι θα διέπουν τη δομή της πληροφορίας σε επίσης μηχαναναγνώσιμη μορφή. Σε αυτό το πιο αφαιρετικό επίπεδο είναι δυνατή η αποτύπωση λογικών κανόνων, και στη συνέχεια ο έλεγχος των δεδομένων ως προς τη συνέπειά τους προς αυτούς τους λογικούς κανόνες.

Για το διαχωρισμό της συντακτικής από τη λογική ορθότητα ακολουθεί ένα παράδειγμα. Έστω η πρόταση «Η αγελάδα μου γέννησε ένα μοσχάρι» και η πρόταση «Η καρέκλα μου γέννησε ένα καρεκλάκι». Και οι δύο προτάσεις είναι συντακτικά σωστές. Εάν όμως έχει οριστεί ο κανόνας ότι μόνο τα έμβια όντα μπορούν να γεννήσουν, και σε αυτή την κατηγορία έχουν ενταχθεί οι αγελάδες αλλά έχουν αποκλειστεί από αυτή την κατηγορία οι καρέκλες, τότε η δεύτερη πρόταση παραμένει συντακτικά ορθή αλλά είναι λογικά ψευδής. Και αυτό μπορεί να το εντοπίσει η διαδικασία του συλλογισμού. Οι πόροι όπου κωδικοποιούνται αυτοί οι λογικοί κανόνες, οι οποίοι είναι σε θέση να συμπεριλάβουν την κωδικοποιημένη γνώση που απαιτείται για τη διαδικασία του συμπερασμού, ονομάζονται οντολογίες.

### 5.3.2 Οντολογίες και OWL

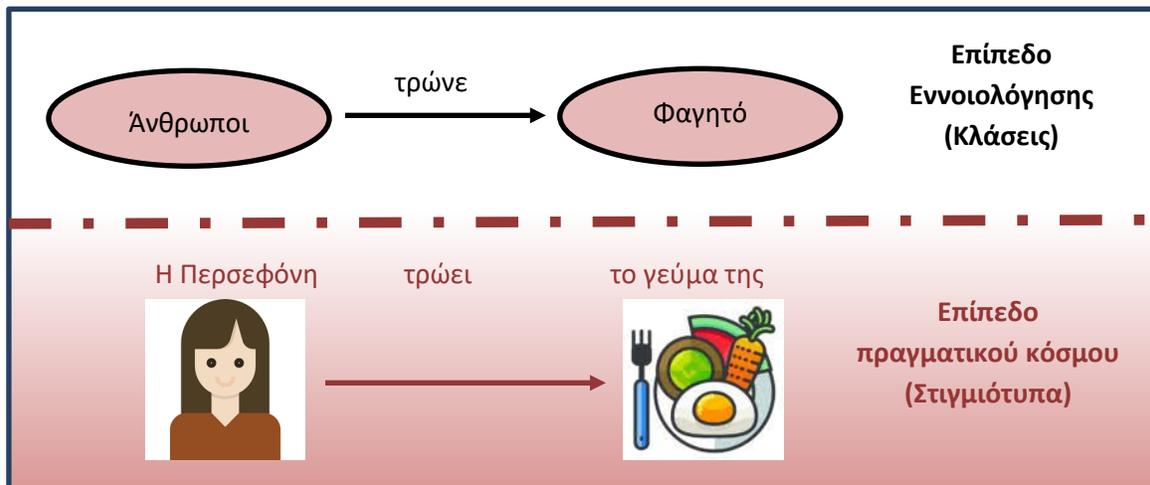
Ένα από τα σημαντικά συστατικά του σημασιολογικού ιστού είναι η αναπαράσταση γνώσης μέσω οντολογιών. Η αντίληψη τού τι είναι οντολογία «αφορά δύο οπτικές: η μία προέρχεται παραδοσιακά από τη φιλοσοφία, ενώ η άλλη, πιο πρόσφατη, από την πληροφορική. Η φιλοσοφική οπτική της οντολογίας εστιάζει στην κατηγορική ανάλυση. [...] Εκ πρώτης όψευς, η πρόθεση της κατηγορικής ανάλυσης είναι να καταγράψει την πραγματικότητα. Η οπτική της οντολογίας που προέρχεται από την πληροφορική, δηλαδή η οντολογία ως τεχνολογία, εστιάζει στις ίδιες ακριβώς ερωτήσεις αλλά η πρόθεση είναι διαφορετική: η δημιουργία υπολογιστικά επεξεργάσιμων μοντέλων της πραγματικότητας, τεχνουργήματα τα οποία μπορούν να χρησιμοποιηθούν από λογισμικό, και ίσως μπορούν να είναι άμεσα ερμηνεύσιμα αλλά και να χρησιμεύσουν για την επίτευξη συλλογισμών με τη χρήση ειδικού λογισμικού που καλείται 'μηχανές συμπερασμού', ώστε να διαποτίσουν το λογισμικό με σημασιολογία ανθρώπινου επιπέδου» (Poli & Obrst, 2010, σελ. 1). Ένας από τους πιο διαδεδομένους ορισμούς, στο πεδίο της πληροφορικής, για το τι είναι οντολογία είναι πώς πρόκειται για «μία ρητή δήλωση μιας εννοιολόγησης. Ο όρος αποτελεί δάνειο από την φιλοσοφία, όπου Οντολογία είναι η συστηματική καταγραφή της Ύπαρξης. Για τα συστήματα τεχνητής νοημοσύνης 'υπάρχει' αυτό που μπορεί να αναπαρασταθεί» (Gruber, 1995, σελ. 908). Από τα παραπάνω είναι προφανές ότι αυτός που ασχολείται με τις οντολογίες «απομακρύνεται από ερωτήματα σχετικά με την ουσία της πραγματικότητας και στρέφεται σε ερωτήματα που αφορούν στην αναπαράσταση της πραγματικότητας» (Kohne, 2014, σελ. 88). Αυτή η αναπαράσταση λειτουργεί ως ένα υποκατάστατο εντός του λογισμικού που επιτελεί το συλλογισμό (reasoning) για τα πράγματα που υπάρχουν στον κόσμο (Davis κ.ά., 1993, σελ. 18) και συχνά καλείται συμβολικός υπολογισμός (symbolic computing).

Η αναπαράσταση γνώσης διαθέτει 4 βασικά συστατικά: α) τις οντότητες ή στιγμιότυπα (αντικείμενα, έννοιες, άνθρωποι κτλ), β) τις κατηγορίες στις οποίες αυτά ανήκουν, γ) τις σχέσεις τους και δ) τα ονόματά τους<sup>33</sup>. Στο Σχήμα 3 οι οντότητες απεικονίζονται με τους κίτρινους κόμβους (ελλείψεις), τα ονόματα απεικονίζονται με τα παραλληλόγραμμα που περιέχουν κόκκινα γράμματα, οι σχέσεις με τις συνδέσεις (ακμές) και οι κατηγορίες (τάξεις πραγμάτων) με τους μωβ κόμβους (ελλείψεις).

Όπως σημειώνουν οι Ribes & Bowker στο εξαιρετικό άρθρο τους «Between meaning and machine»: «οι οντολογίες μπορούν να θεωρηθούν ως ένα είδος τεχνολογικής σημασιολογικής πύλης που επιτρέπουν την επικοινωνία και τη συνέργεια μεταξύ ετερογενών επιστημών, γλωσσών, κατηγοριών ή σχημάτων βάσεων δεδομένων. Αυτό είναι ένας βασικός στόχος των κυβερνοϋποδομών (cyber infrastructures) αφού επιτρέπει στα δεδομένα και τους πόρους να κινούνται ελεύθερα μεταξύ των θεσμών της επιστήμης» (Ribes & Bowker, 2009, σελ. 201).

<sup>33</sup> Μπορούν να δηλωθούν επιπλέον χαρακτηριστικά (attributes) αλλά η λεπτομερής περιγραφή των δυνατοτήτων των εν λόγω γλωσσών είναι πέραν των σκοπών του παρόντος κειμένου.

Στο Σχήμα 4<sup>34</sup> απεικονίζονται τα δύο διαφορετικά επίπεδα τα οποία μπορεί να αναπαραστήσει μια οντολογία. Το ένα είναι το αφαιρετικό επίπεδο των κατηγοριών (κλάσεων σύμφωνα με την καθιερωμένη ορολογία του χώρου) και το άλλο είναι το επίπεδο του πραγματικού κόσμου, δηλαδή εκείνο των ατόμων/στιγμιότυπων.



Σχήμα 4: πραγματικός κόσμος και εννοιολόγηση κλάσεων

Για τον πρωτοπόρο της οντολογικής αναπαράστασης Gruber, οι οντολογίες είναι η τεχνολογία που καθιστά δυνατή την ύπαρξη του σημασιολογικού ιστού αφού είναι ένα μέσο για να διασαφηνίσουν οι άνθρωποι τι εννοούν με τους όρους που χρησιμοποιούν στα δεδομένα τους (Gruber, 2007). Έτσι οι οντολογίες συμβάλουν στο να απαλλαγούμε από ένα σημαντικό πρόβλημα που παρατηρούσε ο Fischer στο τέλος του προηγούμενου αιώνα ότι «δυστυχώς τα υπολογιστικά εργαλεία για την ανάλυση περίπλοκων καταστάσεων είναι γενικά hand-crafted για την κάθε εφαρμογή» (M. D. Fischer, 1994, σελ. 10) διότι πλέον με τις οντολογίες και τη διαλειτουργικότητα που αυτές εξασφαλίζουν, το πρόβλημα δεν έχει να κάνει με τις εκάστοτε εφαρμογές αλλά τίθεται σε ένα αφαιρετικό επίπεδο.

Η πιο διαδεδομένη γλώσσα για την κωδικοποίηση οντολογιών είναι η Web Ontology Language (OWL) (W3C, 2004). Η εν λόγω γλώσσα, είναι συμβατή με όλα όσα αναφέρθηκαν παραπάνω. Εν συντομία, ορίζει κλάσεις, άτομα και ιδιότητες των ατόμων. Οι ιδιότητες αποτελούνται από τα datatype properties (ιδιότητες οι οποίες λαμβάνουν μια λεκτική ή αριθμητική τιμή (data literal) και οι object properties (ιδιότητες οι οποίες συνδέουν μεταξύ τους τα στιγμιότυπα). Για παράδειγμα, οι ιδιότητες του πρώτου τύπου θα μπορούσαν να χρησιμοποιηθούν για να δηλώσουν το ύψος ενός ανθρώπου ενώ εκείνες του δεύτερου τύπου για δηλώσουν τη σχέση ενός ατόμου με ένα άλλο άτομο.

<sup>34</sup> Παραλλαγή του σχήματος εμφανίζεται στο έργο «Τεκμηρίωση μουσειακών συλλογών: Μεταδεδομένα για αντικείμενα πολιτισμικής κληρονομιάς» (Καπιδάκης κ.ά., 2024), στο κεφάλαιο «Σημασιολογικά δίκτυα και οντολογίες».

Για κάθε σχέση μπορεί να οριστεί πεδίο ορισμού (domain) και πεδίο τιμών (range). Υπάρχουν ακόμη και οι annotation properties. Αυτές χρησιμοποιούνται για να προστεθούν σχόλια σε κλάσεις, ιδιότητες ή άτομα αλλά δεν επηρεάζουν το συλλογισμό (reasoning). Επίσης, η OWL επιτρέπει τη χρήση περιορισμών (restrictions), όπως ξένες μεταξύ τους κλάσεις (disjoint). Ξένες κλάσεις μεταξύ τους είναι εκείνες που δεν μπορούν να έχουν κοινά μέλη.

Ένας πολύ βασικός μηχανισμός της OWL είναι η ύπαρξη της σχέσης «is-a». Η εν λόγω σχέση είναι εκείνη που δημιουργεί μια ιεραρχία συνδέοντας τις υποκλάσεις με την κλάση στην οποία ανήκουν. Η ιεραρχία αυτή βασίζεται στο ότι κάθε υποκλάση κληρονομεί όλα τα χαρακτηριστικά των υπερκλάσεων αυτής. Άρα αν ένα στιγμιότυπο ανήκει σε πάνω από μία κλάσεις, τότε κληρονομεί όλα τα χαρακτηριστικά αυτών. Η σχέση «is-a» μπορεί να χρησιμοποιηθεί και για την δημιουργία μιας αντίστοιχης ταξινόμιας των ιδιοτήτων.

Ένα σημαντικό ζήτημα που αφορά την οντολογική αναπαράσταση είναι η διάκριση μεταξύ του ανοικτού κόσμου και του κλειστού κόσμου. Η παραδοχή του κλειστού κόσμου (Closed World Assumption) θεωρεί ότι οτιδήποτε είναι αληθές γνωρίζουμε ότι είναι αληθές, ή με άλλα λόγια ό,τι δεν γνωρίζουμε ότι είναι αληθές τότε είναι ψευδές. Αντίθετα, η παραδοχή του ανοικτού κόσμου (Open World Assumption) θεωρεί πως για ό,τι δεν γνωρίζουμε αν είναι αληθές τότε η κατάστασή του είναι άγνωστη. Η OWL είναι συμβατή με την παραδοχή του ανοικτού κόσμου, ενώ οι βάσεις δεδομένων (αλλά και η Prolog, στην οποία αναφερθήκαμε πριν) είναι συμβατές με την παραδοχή του κλειστού κόσμου.

Τέλος θα πρέπει να σημειωθεί ότι η OWL είναι πλήρως συμβατή με την RDF και κάθε OWL οντολογία μπορεί να αναπαρασταθεί ως ένας RDF γράφος. Για να γίνει αυτό κάθε OWL ιδιότητα μπορεί να γίνει υποκείμενο μιας RDF τριπλέτας. Στο Απόσπασμα κώδικα 3 παρουσιάζεται μια ιδιότητα η οποία εκφράζει τη σχέση ανάμεσα σε αδέρφια. Συγκεκριμένα, η OWL ιδιότητα, όπως έχει οριστεί, περιορίζει τη σχέση ανάμεσα σε ανθρώπους και της αποδίδει το χαρακτηριστικό της συμμετρικής ιδιότητας. Αυτό σημαίνει ότι αν το A άτομο συνδέεται με το B άτομο με αυτή τη σχέση τότε και το B άτομο συνδέεται με το A με την ίδια σχέση. Επίσης και τα δύο άτομα είναι υποχρεωτικά άνθρωποι. Άρα, έτσι όπως έχει οριστεί η ιδιότητα, δεν μπορεί να χρησιμοποιηθεί για να δηλώσει ότι δύο γατάκια είναι αδέρφια.

### Απόσπασμα κώδικα 3: μια OWL ιδιότητα (σε RDF/XML) με τα χαρακτηριστικά της

```
<owl:ObjectProperty rdf:about="http://MyPhDExamples#has_sibling">
  <rdf:type
    rdf:resource="http://www.w3.org/2002/07/owl#SymmetricProperty"/>
  <rdfs:domain rdf:resource="http://MyPhDExamples#Human"/>
  <rdfs:range rdf:resource="http://MyPhDExamples#Human"/>
</owl:ObjectProperty>
```

Ένα ζήτημα που προκύπτει συχνά κατά την αναπαράσταση γνώσης έχει να κάνει με τον μη σαφή διαχωρισμό των δύο επιπέδων που παρουσιάζει το «Σχήμα 4: *πραγματικός κόσμος και εννοιολόγηση κλάσεων*». Όπως αναφέρει ο Doerr συχνά καταγράφεται μια σχέση ανάμεσα στη *factual knowledge* (δηλαδή αυτή των ατόμων/στιγμιότυπων) και σε εκείνη του επιπέδου των κλάσεων. Ονομάζει αυτού του είδους τη γνώση “*Cross-categorical knowledge*”. Ένα από τα παραδείγματα που δίνει είναι το «η γάτα μου έφαγε ένα ψάρι» (Doerr, 2005). Όπου «η γάτα μου» είναι ένα συγκεκριμένο άτομο αλλά το «ένα ψάρι» σε αναφέρεται σε ένα συγκεκριμένο στιγμιότυπο. Η μοντελοποίηση του παραπάνω στην OWL-DL είναι δυνατή μόνο μέσω της χρήσης περιορισμών (*restrictions*). Στο απόσπασμα κώδικα που ακολουθεί κωδικοποιείται μια τέτοια προσέγγιση.

**Απόσπασμα κώδικα 4: σύνδεση στιγμιότυπου με κλάση αντί με στιγμιότυπο**

```

<owl:NamedIndividual rdf:about="http://MyPhDExamples#H_Γάτα_Μου">
  <!--Η Γάτα μου έφαγε ένα ψάρι-->
  <rdf:type>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://MyPhDExamples#έφαγε"/>
      <owl:someValuesFrom rdf:resource="http://MyPhDExamples#Ψάρι"/>
    </owl:Restriction>
  </rdf:type>
</owl:NamedIndividual>

```

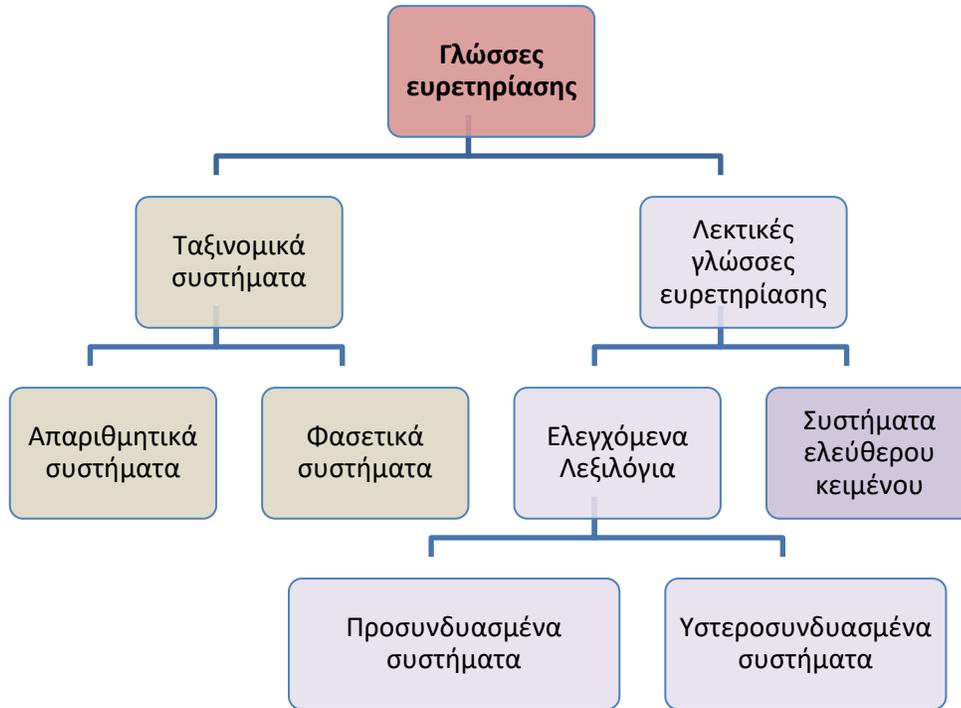
### 5.3.2.1 Διαδικασία ανάπτυξης μιας οντολογίας

Η διαδικασία αποτύπωσης της γνώσης του ερευνητή σε όρους οντολογίας, δηλαδή η ανάπτυξη μιας οντολογίας, απαιτεί μια συγκεκριμένη διαδικασία μοντελοποίησης. Στην έρευνα που οδήγησε στο άρθρο «*Between meaning and machine*», οι Ribes & Bowker χρησιμοποιούν εθνογραφικές μεθόδους για να παρακολουθήσουν την ανάπτυξη μιας οντολογίας, συγκεκριμένα της GEON που αφορά γεωγραφικά δεδομένα. Και σημειώνουν: επειδή η αναπαράσταση γνώσης έχει μια δική της γλώσσα (συχνά καλείται «*λογική πρώτου βαθμού*» (*first order logic*), και η οποία στην ουσία αποτυπώνεται μέσω της OWL) το επιστημονικό πεδίο πρέπει να μάθει να κωδικοποιεί τη γνώση του σε αυτήν τη γλώσσα (Ribes & Bowker, 2009, σελ. 207). Αυτή η γλώσσα προϋποθέτει ότι τα μέλη της επιστημονικής κοινότητας είναι σε θέση να εκφράσουν τη γνώση τους μεταφράζοντάς τη σε οντότητες, τελεστές και λογικές σχέσεις (Ribes & Bowker, 2009, σελ. 206). Όμως αυτές οι κατηγορίες σκέψης, ή οντολογικές κατασκευές, δεν είναι άμεσα προφανείς ακόμη και στον ειδικό και απαιτούν την εξαγωγή γνώσης από πλήθος ειδικών και τη συμπύκνωσή της σε φορμαλιστική λογική, μια διαδικασία η οποία απαιτεί αξιοσημείωτη δεξιότητα και διεπιστημονική γνώση (Doerr & Iorizzo, 2008, σελ. 3). Το ότι οι απαιτούμενες κλάσεις για τη μοντελοποίηση είναι κάθε άλλο παρά προφανείς φαίνεται και από το ότι πολλές φορές στις οντολογίες υπάρχουν έννοιες οι οποίες δεν

υπάρχουν στα λεξικά που περιέχουν την ορολογία του πεδίου (L'Homme & Bernier-Colborne, 2012).

## 5.4 Τα Συστήματα Οργάνωσης της Γνώσης (ΣΟΓ) ως τεχνητές 'γλώσσες' και ως φορείς νοήματος

Όλοι οι τύποι Συστημάτων Οργάνωσης της Γνώσης (ΣΟΓ), λεκτικοί ή μη, βασίζονται στην ικανότητα του ανθρώπινου μυαλού να δημιουργεί και να διαχειρίζεται κατηγορίες. Σαφώς, κάθε κατηγοριοποίηση δεν συνιστά ταξινόμηση (Jacob, 2004), ωστόσο, η κατηγοριοποίηση αποτελεί προϋπόθεση για την ταξινόμηση. Η θεματική πρόσβαση αποτελεί θεμελιώδες ζήτημα για τους παρόχους περιεχομένου, και έχουν προσπαθήσει σκληρά να αναπτύξουν μια σειρά από ΣΟΓ για να εξασφαλίσουν και να βελτιστοποιήσουν την θεματική πρόσβαση στις συλλογές τους. Τα ΣΟΓ, κατά τα πρώτα στάδια ανάπτυξής τους τον 19ο αιώνα, χρησίμευαν κυρίως ως «συσκευές συμπαράθεσης» (collocating devices), έχοντας ως κύριο στόχο να ομαδοποιήσουν στο φυσικό χώρο τα σχετικά τεκμήρια (R. P. Smiraglia, 2014, Κεφάλαιο 4.2.4) ενώ αργότερα χρησιμοποιήθηκαν ως εργαλεία ευρετηρίασης. Ο θεωρητικός της οργάνωσης της γνώσης Birger Hjørland κατηγοριοποιεί τις παραδοσιακές γλώσσες θεματικής ευρετηρίασης σε δύο βασικές κατηγορίες (βλέπε Σχήμα 5). Η πρώτη περιλαμβάνει τα συστήματα ταξινόμησης που υποδιαιρούνται σε απαριθμητικά και σε φασετικά (faceted), και η δεύτερη περιλαμβάνει τις γλώσσες λεκτικής ευρετηρίασης που υποδιαιρούνται σε ελεγχόμενα συστήματα και σε ελεύθερο κείμενο. Τα Ελεγχόμενα συστήματα υποδιαιρούνται περαιτέρω σε προ-συνδυασμένα και υστερο-συνδυασμένα συστήματα (Hjørland, 2012, Εικόνα 1). Η σημαντικότερη διαφορά των γλωσσών ευρετηρίασης είναι μεταξύ των «συστημάτων ελεύθερου κειμένου» και των όλων των άλλων, όπως καταλήγει ο Hjørland (2012, σελ. 304).



**Σχήμα 5:** οι κατηγορίες των ΣΟΓ ως γλώσσες ευρετηρίασης σύμφωνα με τον Hjørland

Παρόλο που είναι μάλλον αυτονόητο ότι κανένα από τα υπάρχοντα ΣΟΓ δεν φτάνει την πλήρη εκφραστικότητα της φυσικής γλώσσας, εντούτοις πολλοί ερευνητές θεωρούν τα ΣΟΓ ως γλώσσες, αν και «τεχνητές». Προς αυτή την κατεύθυνση, οι Smiraglia και Heuvel αναφέρουν ότι «από τη γλωσσολογική θεωρία έχουμε μάθει ότι τόσο η σύνταξη όσο και η σημασιολογία έχουν αξία για την κατανόηση των εννοιών. Δεδομένου ότι διάφορες ταξινομήσεις έχουν συζητηθεί ως τεχνητές γλώσσες, μας ενδιαφέρει ιδιαίτερα η αλληλεπίδραση μεταξύ σύνταξης και σημασιολογίας στις σχέσεις μεταξύ στοιχειωδών δομών της γνώσης από ιστορική σκοπιά» (R. Smiraglia & Heuvel, 2013, σελ. 361–362). Η Zavalina δηλώνει ότι «η διαδικασία αναπαράστασης του θέματος επηρεάζεται από τους περιορισμούς των διαθέσιμων θεματικών γλωσσών» (Zavalina, 2012, σελ. 143), ενώ ο Svenonius υποστηρίζει ότι «είναι χρήσιμο να θεωρούμε ένα εργαλείο για την παροχή θεματικής πρόσβασης στην πληροφορία ως γλώσσα, καθώς αυτό παρέχει ένα έτοιμο εννοιολογικό πλαίσιο που μπορεί να χρησιμοποιηθεί στην ανάλυση και την περιγραφή της. Αυτό το πλαίσιο χαρακτηρίζει μια γλώσσα από την άποψη του λεξιλογίου, της σημασιολογίας, του συντακτικού και της πραγματολογίας της» (Svenonius, 2000, σελ. 18). Έχοντας κατά νου το Σύστημα Παγκόσμιας Δεκαδικής Ταξινόμησης UDC, οι Salah κ.ά. προτείνουν ότι «ένα σύνθετο ΣΟΓ είναι μάλλον μια γλώσσα για την περιγραφή αντικειμένων με ελεγχόμενο τρόπο παρά ένα ιεραρχικό δέντρο» (Salah κ.ά., 2012, σελ. 56).

Επί του παρόντος, η μοντελοποίηση των ΣΟΓ φαίνεται να επεκτείνεται σε ένα διδιάστατο χώρο. Η μία διάσταση είναι ένας τεχνικός προσανατολισμός και μια προσέγγιση που είναι πρακτική με τη μετάβαση στη νέα τεχνολογική υποδομή. Η

έμφαση δίνεται στον δομικό μετασχηματισμό που ασχολείται με θέματα όπως η μετατροπή ενός δεδομένου θησαυρού σε SKOS, ή πώς θα μετατραπεί το SKOS σε RDF/XML, και ούτω καθεξής. Με αυτόν τον τρόπο η σημασιολογία παραμένει η ίδια και η εκφραστικότητα της νέας “γλώσσας” υποβαθμίζεται στην εκφραστικότητα της παλιάς. Αυτή η πρακτική είναι σαν να μετατρέπεται ένα αρχείο txt σε μια πλουσιότερη μορφή, όπως docx ή html, χωρίς να εξετάζεται η εφαρμογή των νέων διαθέσιμων χαρακτηριστικών, όπως η δημιουργία κεφαλίδων και η μορφοποίηση. Αυτό που συνιστά παράδοξο είναι ότι η προσέγγιση αυτή φαίνεται να χαιρετίζεται ως αποτελεσματική εκμετάλλευση αυτού του σύγχρονου και ισχυρού μορφότυπου! Προφανώς, η χρήση ενός πλουσιότερου μορφότυπου είναι απαραίτητη προϋπόθεση- ωστόσο, δεν αρκεί για την παροχή εκφραστικότερης πληροφορίας.

Εκτός του τεχνολογικού προσανατολισμού, η άλλη διάσταση αφορά στη θεωρητική προσέγγιση, η οποία όμως είναι αρκετά γενική. Υπάρχει μια τεράστια και αξιοσημείωτη βιβλιογραφία αφιερωμένη στις επιστημολογικές αντιλήψεις της ταξινόμησης και των ΣΟΓ. Οι συζητήσεις βασίζονται σε μοντέρνες ή μεταμοντέρνες απόψεις που εστιάζουν σε πτυχές όπως το αν οι έννοιες εκφράζουν μια αντικειμενική πτυχή της πραγματικότητας ή όχι (Mai, 1999)· αν είναι εφικτή μια «διεπιστημονική επικοινωνία [...] που δεν εδράζεται σε συγκεκριμένους κλάδους ή πολιτισμούς» (Szostak, 2011, σελ. 2247)· ή αν οι έννοιες μπορούν να κατανοηθούν στο πλαίσιο συγκεκριμένων επιστημολογικών θεωριών (Hjørland, 2009), (Hjørland & Pedersen, 2005). Ως εκ τούτου, υπάρχει ανάγκη να μετακινηθεί η ταξινόμηση προς μια μεταμοντέρνα αντίληψη (Mai, 2011). Η Olson, με κύριο γνώμονα τις μελέτες του φεμινισμού, επιχείρησε μια πολύ ενδιαφέρουσα κριτική, εστιάζοντας στο τι κρύβεται πίσω από τους όρους και τις δομές των ΣΟΓ των βιβλιοθηκών και πώς αυτές αντανακλούν τις κυρίαρχες απόψεις της κοινωνίας (Olson, 2000), (Olson, 2001). Αυτό δείχνει ότι οι έννοιες δεν είναι ουδέτερες αναπαραστάσεις ενός αντικειμενικού κόσμου, αλλά αντανακλούν την περιθωριοποίηση και τους αποκλεισμούς που υιοθετούνται σε κάθε κοινωνία. Για τον λόγο αυτό ο Sperberg-McQueen αναφέρει ότι «οι υποθέσεις που διέπουν ένα σύστημα ταξινόμησης, μπορεί να γίνουν ουσιαστικά άορατες και συνεπώς να μην υπόκεινται πλέον σε αμφισβήτηση ή επανεξέταση- για τους σκοπούς της επιστημονικής εργασίας, μια τέτοια αορατότητα είναι επικίνδυνη και πρέπει να αποφεύγεται» (Sperberg-McQueen, 2004, Τμήμα 1).

Συμπερασματικά, τα Συστήματα Οργάνωσης Γνώσης, ως τρόπος μοντελοποίησης της γνώσης, διαδραματίζουν βασικό ρόλο στη δημιουργία νέας γνώσης, διότι ο τρόπος με τον οποίο μοντελοποιούμε την κατανόησή μας αντανακλά και επηρεάζει την ικανότητά μας να κατανοούμε. «Με άλλα λόγια, ο τρόπος οργάνωσης και αναπαράστασης της γνώσης εξαρτάται σε μεγάλο βαθμό από την κατανόηση του τρόπου με τον οποίο παράγεται και συνειδητοποιείται η γνώση» (B. Alexiev & Marksburly, 2010, σελ. 364). Αυτού του είδους η εξάρτηση αποκαλύφθηκε σε μια μελέτη η οποία κατέληξε στο συμπέρασμα ότι η ανάγκη για τον οντολογικό φορμαλισμό της OWL οδηγεί τους παραγωγούς οντολογιών να «δημιουργούν» λεκτικά τα οποία δεν υπάρχουν στα εξειδικευμένα λεξικά του λεξιλογίου ενός συγκεκριμένου τομέα (L'Homme & Bernier-Colborne, 2012). Επιπλέον, η μοντελοποίηση της γνώσης δεν επηρεάζει μόνο τα μοτίβα της σκέψης μας, αλλά μπορεί επίσης, υπό ορισμένες προϋποθέσεις, να ελέγξει την

ορθότητα των υποθέσεών μας. Ο Goguen ονομάζει αυτή τη διαδικασία ως “πειραματική φιλοσοφία”, δεδομένου ότι ζητά τη διεξαγωγή πειραμάτων για τον έλεγχο της ορθότητας των υποθέσεων σχετικά με τις γλωσσικές έννοιες (Goguen, 1974, σελ. 546).

Η παραδοχή ότι οι έννοιες είναι τα δομικά στοιχεία (building blocks) όλων των μορφών των ΣΟΓ, συνεπάγεται ότι οι σχέσεις πρέπει να είναι η κόλλα μεταξύ τους, δεδομένου ότι χωρίς ισχυρούς δεσμούς μεταξύ των στοιχείων δεν υπάρχει στέρεο οικοδόμημα. Η πλειονότητα των ΣΟΓ διέπεται από σχέσεις ισοδυναμίας (βασικά από την άποψη της διαχείρισης συνωνύμων ή οιονεί συνωνύμων), ιεραρχικές σχέσεις και συσχετιστικές- όλες αυτές συχνά φέρουν το όνομα «παραδειγματικές σχέσεις». Οι συσχετιστικές σχέσεις αποτελούν μια πολύ γενική και διφορούμενη κατηγορία, αφού «κάθε παραδειγματική δια-ορολογική σχέση που δεν είναι ιεραρχική, ούτε ισοδυναμίας, περιγράφεται σήμερα συνήθως ως “συσχετιστική”» (Clarke, 2001). Από την άλλη πλευρά, υπάρχουν οι Συντακτικές σχέσεις οι οποίες περιλαμβάνουν λεξικές σχέσεις (συχνά ενδο-ορολογικές). Ο Green αναφέρει ότι οι Παραδειγματικές σχέσεις είναι μια κλειστή κατηγορία, ενώ οι Συντακτικές σχέσεις είναι μια ανοικτή κατηγορία (Green, 2001, σελ. 6), η οποία επιτρέπει «την έκφραση νέων, σύνθετων σημασιών», όπως αναφέρει ο Engerer (2017, σελ. 1485)

#### 5.4.1 Το SKOS ως μέσο αναπαράστασης των Συστημάτων Οργάνωσης της Γνώσης (ΣΟΓ)

Το SKOS είναι μια σύσταση (recommendation) του W3C και έχει σχεδιαστεί για την αναπαράσταση δομημένων ελεγχόμενων λεξιλογίων, όπως θησαυροί, συστήματα ταξινόμησης και συστήματα θεματικών επικεφαλίδων. Αποσκοπεί στη δόμηση των δεδομένων ώστε να είναι δυνατή η μηχανική επεξεργασία τους στο περιβάλλον του σημασιολογικού ιστού και των διασυνδεδεμένων δεδομένων. Όπως το θέτει η Shiri: Το SKOS στοχεύει να οικοδομήσει μια γέφυρα μεταξύ του κόσμου των συστημάτων οργάνωσης γνώσης [...] και της κοινότητας των διασυνδεδεμένων δεδομένων, με στόχο να αποφέρει οφέλη και στους δύο (Shiri, 2014, σελ. 18).

Σύμφωνα με το W3C το μοντέλο δεδομένων SKOS θεωρεί ένα σύστημα οργάνωσης γνώσης ως ένα σχήμα εννοιών που περιλαμβάνει ένα συγκεκριμένο σύνολο εννοιών. Αυτά τα σχήματα εννοιών SKOS και οι έννοιες SKOS («skos:Concept») προσδιορίζονται από URIs, επιτρέποντας σε οποιονδήποτε να αναφέρεται σε αυτά με σαφήνεια, από οποιοδήποτε πλαίσιο, καθιστώντας τα μέρος του Παγκόσμιου Ιστού. Το SKOS ορίζει μια έννοια ως μια ιδέα ή αντίληψη- μια μονάδα σκέψης. Ωστόσο, το τι συνιστά μονάδα σκέψης είναι υποκειμενικό και αυτός ο ορισμός έχει σκοπό να είναι υποδηλωτικός και όχι περιοριστικός (W3C, 2009b). Επιπλέον, ένα «skos:Concept» προορίζεται να παρέχει έναν ουδέτερο στόχο για τη μετάβαση μιας μεγάλης ποικιλίας εννοιών ΣΟΓ στον Ιστό Δεδομένων» (Baker κ.ά., 2013, σελ. 39) και δεν αναγκάζει εκείνον που κάνει τη μοντελοποίηση να διακρίνει μεταξύ κλάσεων και στιγμιοτύπων (Maltese & Farazi, 2011, σελ. 87). Συνεπώς, το SKOS δεν καθορίζει τη φύση των εννοιών. Από την άλλη πλευρά, ορίζει ρητά και φορμαλιστικά τις σχέσεις, δηλαδή τις ιδιότητες, των εννοιών. Ως εκ τούτου, οι σχέσεις περιορίζονται σε οριζόντιες και ιεραρχικές, όπως αυτές που ορίζονται σε έναν θησαυρό- παρά την μεγαλύτερη ακρίβεια με τη χρήση

επεκτάσεων όπως η «`skos:broaderTransitive`». Επιπλέον, το SKOS παρέχει σχέσεις για την αντιστοίχιση εννοιών διαφορετικών συστημάτων εννοιών χρησιμοποιώντας ιδιότητες όπως η «`skos:closeMatch`».

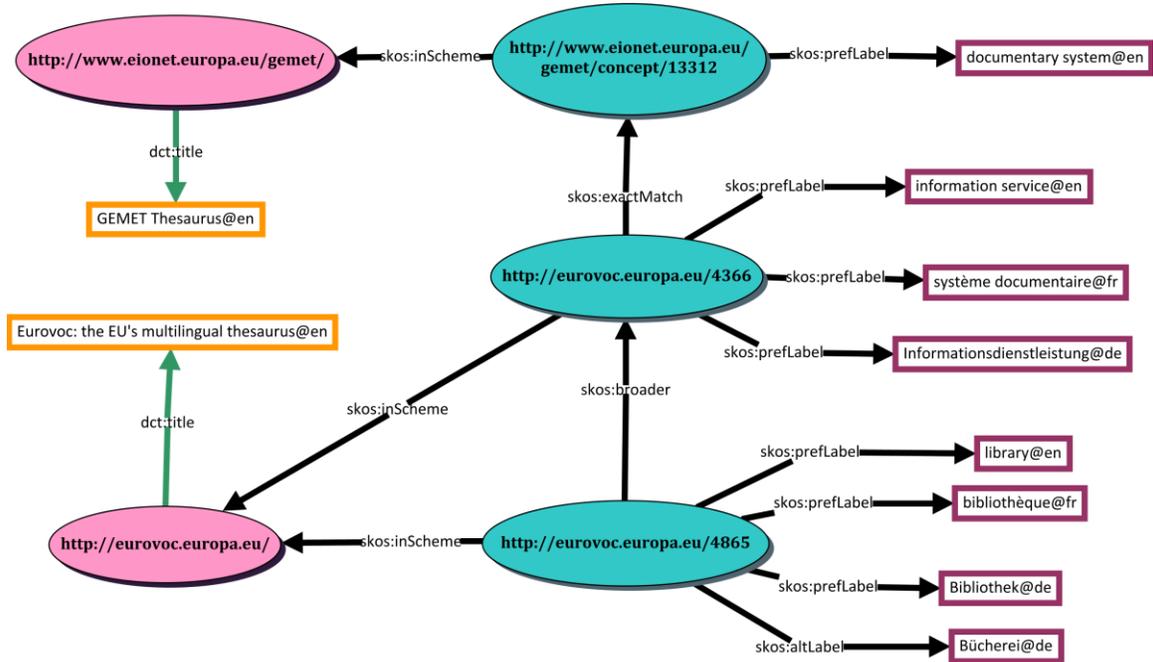
Θα ήταν χρήσιμο να σημειωθεί ότι τα ΣΟΓ είναι σχήματα οργάνωσης της γνώσης, ενώ το SKOS χρησιμοποιείται για την έκφραση των ΣΟΓ στο πλαίσιο του σημασιολογικού ιστού. Πράγματι, λόγω του γεγονότος ότι οι SKOS έννοιες βασίζονται σε URIs και λιγότερο σε ετικέτες φυσικής γλώσσας, δεν υπάρχουν τα όρια που απορρέουν από την εκφραστικότητα της φυσικής γλώσσας. Αυτό δίνει τη δυνατότητα να αναφερόμαστε στα πράγματα όχι με το όνομά τους αλλά χρησιμοποιώντας ένα URI. Τέλος, πρέπει να σημειωθεί ότι το SKOS επιτρέπει ευελιξία στη διατύπωση των URIs χρησιμοποιώντας ιδιότητες όπως η «`skos:hiddenLabel`».

Το SKOS έχει έναν σημαντικό περιορισμό ο οποίος είναι ασύμβατος με την πρακτική καταγραφής των φορέων πολιτισμικής κληρονομιάς. Δεν επιτρέπει τη μοντελοποίηση των ονομάτων των οντοτήτων, δηλαδή εφόσον πρόκειται για RDF literals, δεν υπάρχει δυνατότητα δημιουργίας σχέσεων μεταξύ των ονομασιών των οντοτήτων. Η απόφαση αυτή ελήφθη «προκειμένου να διατηρηθούν οι κύριες προδιαγραφές του SKOS όσο το δυνατόν πιο απλές» (Baker κ.ά., 2013, σελ. 42). Παρόλα αυτά, η μοντελοποίηση των ονομάτων μιας οντότητας είναι σημαντική, έτσι προέκυψε το «SKOS eXtension for Labels» (W3C, 2009a) για να καλύψει το κενό, επιτρέποντας σχέσεις και μεταξύ των ονομασιών (nomens).

Κλείνοντας αυτή την ενότητα, σημειώνεται ότι πλέον η μετάβαση και η συμμόρφωση με τις αρχές του SKOS φέρνουν στην επιφάνεια ορισμένα λανθάνοντα ζητήματα των ΣΟΓ, τα οποία δεν είχαν εντοπιστεί προηγουμένως λόγω του τρόπου χρήσης των ΣΟΓ (Mader κ.ά., 2012). Αυτό οφείλεται κυρίως στο γεγονός ότι η μετάβαση των ΣΟΓ σε SKOS υλοποιείται κυρίως σε δομικό επίπεδο. Ειδικά για τη μετάβαση των θησαυρών, οι Kless κ.ά. ασκούν κριτική στις μεθόδους που χρησιμοποιούνται για τον ανασχεδιασμό των θησαυρών, καθώς δεν αξιοποιούν τις δυνατότητες των τυπικών γλωσσών όπως η OWL όσον αφορά τη διαχείριση της σημασιολογίας για τον εντοπισμό λογικών σφαλμάτων για τη βελτίωση των λεξιλογίων (Kless κ.ά., 2016). Η προαναφερθείσα αντίληψη και πρακτική φαίνεται να έχει εδραιωθεί όχι μόνο στην κοινότητα που είναι υπεύθυνη για τη μετάβαση των ΣΟΓ αλλά και για μεγάλο όγκο δεδομένων που μετασχηματίζονται σε διασυνδεδεμένα δεδομένα (Jain κ.ά., 2010).

Στο Σχήμα 6 αναπαρίστανται τρεις όροι, από δύο διαφορετικούς θησαυρούς, και κάποιες από τις μεταξύ τους σχέσεις. Τα ελλειπτικά σχήματα αφορούν σε κόμβους, οι ακμές σε ιδιότητες και τα παραλληλόγραμμα σε τιμές (RDF literals). Επί της ουσίας το σχήμα δείχνει δύο όρους από τον θησαυρό Eurovoc και έναν από τον θησαυρό GEMET. Μια γρήγορη ανάγνωση μεταξύ των σχέσεων δείχνει ότι ο όρος του Eurovoc <http://eurovoc.europa.eu/4865> (που στα αγγλικά ονομάζεται library, στα γαλλικά bibliothèque κοκ) συνδέεται μέσω της σχέσης «ευρύτερος όρος (`skos:broader`)» με τον όρο <http://eurovoc.europa.eu/4366> (ο οποίος στα αγγλικά ονομάζεται information service, στα γαλλικά système documentaire κοκ). Και ο τελευταίος συνδέεται, μέσω της σχέσης «πλήρης ταύτιση (`skos:exactMatch`)», με τον όρο <http://www.eionet.europa.eu/gemet/concept/13312> του GEMET, που ονομάζεται

«documentary system». Αυτό πρακτικά σημαίνει ότι αυτό που ο Eurovoc ονομάζει «information service» ο GEMET το ονομάζει «documentary system». Άρα, ένα σύστημα που ακολουθεί τις λογικές των διασυνδεδεμένων δεδομένων 'γνωρίζει' ότι πρόκειται για την ίδια έννοια ανεξάρτητα από το όνομα το οποίο χρησιμοποιείται για αυτήν. Όπως είναι πασιφανές, η έμφαση δίνεται στις έννοιες και όχι στα ονόματα μέσω των οποίων αυτές είναι γνωστές.



Σχήμα 6: γραφική αναπαράσταση σε SKOS

Στο Απόσπασμα κώδικα 5 παρουσιάζεται ο RDF/XML κώδικας που αφορά στα δεδομένα που παρουσιάζονται για τους κόμβους <http://eurovoc.europa.eu/4366> και <http://eurovoc.europa.eu/4865>.

**Απόσπασμα κώδικα 5: RDF/XML αναπαράσταση δύο εγγραφών σε SKOS**

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#">
<skos:Concept rdf:about="http://eurovoc.europa.eu/4366">
  <skos:prefLabel xml:lang="en">information service</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">ystème documentaire</skos:prefLabel>
  <skos:prefLabel xml:lang="de">Informationsdienstleistung</skos:prefLabel>
  <skos:inScheme rdf:resource="http://eurovoc.europa.eu"/>
  <skos:exactMatch
rdf:resource="http://www.eionet.europa.eu/gemet/concept/13312"/>
</skos:Concept>

<skos:Concept rdf:about="http://eurovoc.europa.eu/4865">
  <skos:prefLabel xml:lang="en">library</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">bibliothèque </skos:prefLabel>
  <skos:prefLabel xml:lang="de">Bibliothek</skos:prefLabel>
  <skos:altLabel xml:lang="de">Bücherei</skos:altLabel>
  <skos:broader rdf:resource="http://eurovoc.europa.eu/4366"/>
  <skos:inScheme rdf:resource="http://eurovoc.europa.eu"/>
</skos:Concept>

</rdf:RDF>
```

## 5.5 Ανάλυση αντιπροσωπευτικών συστημάτων οργάνωσης και αναπαράστασης γνώσης

Η ενότητα αυτή έχει ως στόχο να αναλύσει αντιπροσωπευτικά δείγματα συστημάτων οργάνωσης και αναπαράστασης της γνώσης με στόχο εξετάσει την εκφραστική τους δύναμη αλλά και την ικανότητά τους να είναι μηχανικά επεξεργάσιμα στο πλαίσιο του σημασιολογικού ιστού. Ο απώτερος στόχος είναι να διαπιστωθεί αν οι μεθοδολογίες που εφαρμόζουν τα αντίστοιχα συστήματα μπορούν να είναι κατάλληλες για την οργάνωση και αναπαράσταση εθνογραφικών δεδομένων και ανθρωπολογικής γνώσης.

### 5.5.1 Σύνολα δεδομένων και μεθοδολογία ανάλυσής τους

Για την εκπλήρωση του στόχου της παρούσας φάσης της μελέτης, εξετάζονται συγκεκριμένα Συστήματα Οργάνωσης Γνώσης (ΣΟΓ) με ιδιαίτερη έμφαση στη δομή τους. Για τις ενδο-ορολογικές σχέσεις, καθώς και για τις ιδιότητες-σχέσεις της οντολογίας, εφαρμόζονται τεχνικές επεξεργασίας φυσικής γλώσσας. Τα ΣΟΓ είναι αντιπροσωπευτικά εκείνων που χρησιμοποιούνται περισσότερο από φορείς πολιτισμικής κληρονομιάς και κυρίως τις βιβλιοθήκες. Πρόκειται για ένα σύστημα θεματικών επικεφαλίδων, το Library of Congress Subject Headings (LCSH), έναν θησαυρό, τον πολύγλωσσο Eurovoc, ένα σύστημα ταξινόμησης, το Dewey Decimal Classification (DDC) και μια οντολογία, το CIDOC-CRM.

Οι όροι των ΣΟΓ περιλαμβάνουν λέξεις-κλειδιά ή σύντομες φράσεις, δηλαδή ελλειπτικές προτάσεις. Αυτό σημαίνει ότι σε κάποιες γλώσσες, όπως η αγγλική, δεν υπάρχουν αρκετά συμφραζόμενα για να γίνει μια αποτελεσματική μορφοσυντακτική ανάλυση. Για παράδειγμα, εάν η λέξη «search» παρέχεται έξω από συμφραζόμενά της, δεν μπορεί να προσδιοριστεί μορφοσυντακτικά αν είναι ουσιαστικό ή ρήμα. Αντίθετα, στην πρόταση «Yesterday I performed a successful search», η λέξη “search” αναγνωρίζεται ως ουσιαστικό, ενώ στην πρόταση «Search for the tickets», η ίδια λέξη αναγνωρίζεται ως ρήμα. Σε άλλες γλώσσες, όπως η ελληνική, η φινλανδική, η ουγγρική, η τουρκική και η ρωσική, λόγω της πλούσιας μορφολογίας τους και της υψηλής κλιτικότητας, η μορφοσυντακτική ανάλυση πραγματοποιείται με μεγαλύτερη ακρίβεια ακόμη και με περιορισμένες πληροφορίες από τα συμφραζόμενα. Πληροφορίες για τη μορφολογική τυπολογία των γλωσσών δίνει ο Comrie (1989) ενώ πιο συγκεκριμένα, πληροφορίες για την κλιτική μορφολογία, βρίσκονται στο έργο του Stump (2001). Επιπλέον, το έργο του Pirkoia (2001) υπογραμμίζει τι σημαίνει η μορφολογική τυπολογία των γλωσσών, δηλαδή τα κοινά ή διαφορετικά γλωσσικά χαρακτηριστικά, ειδικά για τον χώρο της Ανάκτησης Πληροφοριών (information retrieval).

Στη μελέτη μας, λαμβάνοντας υπόψη αυτά τα χαρακτηριστικά των γλωσσών, χρησιμοποιήθηκαν οι ελληνικές εκδόσεις των προαναφερθέντων ΣΟΓ, δεδομένου ότι η ελληνική γλώσσα παρέχει πιο στέρεο έδαφος για τον εντοπισμό μερών του λόγου στα πλαίσια ελλειπτικών προτάσεων. Θα πρέπει να αναφερθεί, ωστόσο, ότι ακόμη και στην ελληνική γλώσσα υπάρχουν ορισμένες περιπτώσεις όπου ο εντοπισμός του μέρους του λόγου αποτελεί πρόκληση όσον αφορά τη συντακτική ασάφεια που προκαλείται από τη μορφολογία των λέξεων. Ένα παράδειγμα είναι η περίπτωση του πληθυντικού αριθμού των θηλυκών ουσιαστικών στην ονομαστική ή αιτιατική πτώση που θα μπορούσε επίσης να σχολιαστεί ως το δεύτερο πρόσωπο των ρημάτων στον ενικό. Τέτοια είναι η περίπτωση του «υποδιαίρεσεις», το οποίο μπορεί να είναι ο πληθυντικός της υποδιαίρεσης ή ρήμα. Ωστόσο, στο πλαίσιο των ΣΟΓ, τέτοιες περιπτώσεις είναι εξαιρετικά σπάνιες, καθώς λίγοι ρηματικοί τύποι εντοπίστηκαν κατά τη διάρκεια της προ-επεξεργασίας.

Πιο συγκεκριμένα, χρησιμοποιήθηκε ο πολύγλωσσος θησαυρός Eurovoc της ΕΕ (έκδοση 4.4), η ελληνική έκδοση του οποίου περιλαμβάνει 6.883 όρους. Όσον αφορά το LCSH, χρησιμοποιήθηκε η έκδοση των αρχείων καθιερωμένων αποδόσεων της Εθνικής Βιβλιοθήκης της Ελλάδος (ΕΒΕ), η οποία περιέχει την ελληνική μετάφραση επιλεγμένων

LC Subject Headings. Από 55.204 θεματικές επικεφαλίδες συνολικά απομονώθηκαν αυτές που δηλώθηκαν ειδικά ως μετάφραση της LCSH, οι οποίες αριθμούσαν 19.986 εγγραφές. Όλα τα υποπεδία MARC21 εκτός από τα \$a και \$x απορρίφθηκαν, καθώς είτε περιλάμβαναν γεωγραφικές υποδιαίρεσεις (\$z), δηλαδή ονομαστικές οντότητες, χρονολογικές υποδιαίρεσεις (\$y), είτε μετέφεραν πληροφορίες που δεν ενέπιπταν στο αντικείμενο της παρούσας μελέτης. Οι καταχωρίσεις που τελικά χρησιμοποιήσαμε ήταν σε επίπεδο υποπεδίου, δηλαδή κάθε \$a και \$x αποτελούσαν μια ξεχωριστή καταχώρηση. Αυτό οδήγησε σε 30.073 καταχωρίσεις. Στη συνέχεια εξαλείψαμε ορισμένα προβλήματα κωδικοποίησης και στίξης και τέλος εξασφάλισαμε ότι δεν υπήρχε επανάληψη των υποπεδίων. Ο αριθμός των μοναδικά ταυτοποιημένων υποπεδίων έφθασε τα 10.308. Το επίπεδο υποπεδίου επιλέχθηκε για την επεξεργασία αυτών των δεδομένων επειδή τα υποπεδία θεωρούνται οι μικρότερες αυτοτελείς λειτουργικές μονάδες μιας επικεφαλίδας. Όσον αφορά το DDC, χρησιμοποιήθηκε η 13η συνοπτική έκδοση που μεταφράστηκε στα ελληνικά από το Ελληνικό Εθνικό Κέντρο Τεκμηρίωσης. Πρέπει να σημειωθεί ότι η ελληνική έκδοση ενσωματώνει τη μεταφρασμένη 21η αμερικανική έκδοση για τους τομείς της ελληνικής λογοτεχνίας και ιστορίας καθώς και της ιστορίας της Κύπρου. Το DDC απέδωσε 5.398 καταχωρίσεις, οι οποίες μειώθηκαν σε 3.811 μετά τη διαδικασία αποδιπλοποίησης (deduplication). Σε όλες τις περιπτώσεις χρειάστηκε περαιτέρω επεξεργασία προκειμένου να γίνει ομοιόμορφος χειρισμός των δεδομένων. Τέλος, χρησιμοποιήθηκε η ελληνική μετάφραση της έκδοσης 4.1 του εννοιολογικού μοντέλου αναφοράς CIDOC (CIDOC-CRM). Για τους σκοπούς της παρούσας μελέτης, οι κλάσεις και οι ιδιότητες αντιμετωπίστηκαν χωριστά.

Κάθε εξεταζόμενο ΣΟΓ αποτελούσε ένα αυτόνομο σύνολο, δηλαδή δεν ήταν αποσπασματικό υποσύνολο ενός μεγαλύτερου όγκου δεδομένων. Παρά το γεγονός ότι το corpus της ΕΒΕ προέρχεται από το LCSH, δεν λογίζεται ως αποσπασματικό υποσύνολο αυτών, καθώς δεν αποκλείει σκόπιμα κάποιο τμήμα του αρχικού συνόλου. Ο περιορισμός του είναι ότι αφορά την ελληνική εκδοτική κατανομή και, ως εκ τούτου, ο εμπλουτισμός του προσαρμόζεται ανάλογα. Στην ουσία αποτελεί αντιπροσωπευτικό δείγμα, καθώς φιλοξενεί τίτλους από όλους τους τομείς. Η ελληνική έκδοση του DDC ακολουθεί την αντίστοιχη αγγλική έκδοση, με μια μόνο απαραίτητη επέκταση για τον τομέα της ελληνικής ιστορίας. Η έκδοση του Eurovoc είναι η πλήρης ελληνική έκδοση, δηλαδή περιλαμβάνει όλες τις περιγραφές, και το CIDOC-CRM είναι μετάφραση ένα προς ένα του πρωτοτύπου. Επιπλέον, πρέπει να αναφέρουμε ότι χρησιμοποιήθηκε μόνο η καθιερωμένη μορφή κάθε έννοιας- έτσι, αποκλείστηκαν οι μη προτιμώμενοι όροι, οι σημειώσεις πεδίου εφαρμογής και κάθε άλλο στοιχείο των εγγραφών.

Το λογισμικό που χρησιμοποιήθηκε για την επισήμανση των Μέρων του Λόγου (ΜτΛ) έχει αναπτυχθεί από το Ινστιτούτο Επεξεργασίας του Λόγου (ΙΕΛ) και αποτελεί μέρος μιας σουίτας εργαλείων Επεξεργασίας Φυσικής Γλώσσας που βασίζεται τόσο σε αλγόριθμους μηχανικής μάθησης όσο και σε προσεγγίσεις βασισμένες σε κανόνες (Prokorporidis κ.ά., 2011). Το λογισμικό είναι ελεύθερα διαθέσιμο ως διαδικτυακή υπηρεσία στη διεύθυνση <http://nlp.ilsp.gr/soaplab2-axis/>. Σε αυτό το σημείο θα πρέπει να αναφερθεί ότι το εργαλείο αυτό αποτελεί πόρο τρίτου, ο οποίος χρησιμοποιείται στο πλαίσιο της παρούσας μελέτης ως τελικό προϊόν. Το σκεπτικό αυτής της επιλογής ήταν

ότι το εργαλείο αναπτύχθηκε από επίσημο ερευνητικό ίδρυμα το οποίο ασχολείται με τη μελέτη της ελληνικής γλώσσας, προφορικής και γραπτής, καθώς και με την ανάπτυξη εργαλείων υπολογιστικής επεξεργασίας.

Σε γενικές γραμμές, ακολουθήσαμε έναν κοινό τρόπο για την προεπεξεργασία των δεδομένων. Δημιουργήσαμε ένα απλό αρχείο κειμένου με τα απαιτούμενα στοιχεία από κάθε ΣΟΓ, το οποίο στη συνέχεια μετατράπηκε σε αρχείο xces xml προκειμένου να υποβληθεί στον tagger. Όπως φαίνεται στο Απόσπασμα κώδικα 6, η υπηρεσία (service) παρείχε ένα αρχείο xces xml με ετικέτες που αφορούσαν στα μέρη του λόγου. Το αρχείο υποβλήθηκε σε επεξεργασία προκειμένου να συγχωνευτεί με το αρχικό αρχείο και να εμπλουτιστεί με πρόσθετες πληροφορίες που απαιτούνται για την ανάλυση των δεδομένων. Όλα τα σύνολα δεδομένων οριστικοποιήθηκαν ως αρχεία csv. Επιπλέον, ορισμένα σύμβολα (!, % και \*) αφαιρέθηκαν προκειμένου να γίνει αποτελεσματικότερη επεξεργασία των δεδομένων. Όλες οι μετρήσεις αναφέρονται στον αριθμό των μοναδικών καταχωρίσεων ανά ΣΟΓ, πράγμα που σημαίνει ότι οι διπλές τιμές αφαιρέθηκαν κατά την προεπεξεργασία. Κατά τη διαδικασία αποδιπλοποίησης, τα κεφαλαία γράμματα θεωρήθηκαν διαφορετικές τιμές ως προς τα αντίστοιχα μικρά γράμματα. Όσον αφορά την πολιτική μη εξάλειψης των μικρών και των κεφαλαίων γραμμάτων, η απόφαση βασίστηκε στην εφαρμογή μιας ανάλυσης ανά ΣΟΓ, η οποία μας οδήγησε στην ερμηνεία ότι η διαφοροποίηση στη χρήση μικρών ή κεφαλαίων γραμμάτων ήταν σκόπιμη.

#### Απόσπασμα κώδικα 6: παράδειγμα επισημειώσεων ΜτΛ (POS tagging) από ΙΕΛ

```
<p id="NLG.78">
<s id="s78">
<t id="t137" word="Νομικό" tag="AjBaNeSgAc" lemma="νομικός"/>
<t id="t138" word="καθεστώς" tag="NoCmNeSgAc" lemma="καθεστώς"/>
<t id="t139" word="," tag="PUNCT" lemma=","/>
<t id="t140" word="νόμοι" tag="NoCmMaPlNm" lemma="νόμος"/>
<t id="t141" word="κλπ." tag="ABBR" lemma="κλπ."/>
</s>
</p>
```

Βασιζόμενοι στη σχετική βιβλιογραφία (Abreu & Vieira, 2017; Bean κ.ά., 2002, σελ. x), αποφασίσαμε αν τα μέρη του λόγου θα μπορούσαν να θεωρηθούν οντότητες ή σχέσεις. Ως εκ τούτου, θεωρήσαμε ως οντότητες: τα Ουσιαστικά (κοινά και κύρια), τις Συντομογραφίες που περιλάμβαναν μόνο κεφαλαία γράμματα, τις ειδικές Αντωνυμίες, δηλαδή τις δεικτικές, τις προσωπικές και τις σχετικές αόριστες, και τις Ξένες Λέξεις, οι οποίες αποτελούνταν κυρίως είτε από κοινά ουσιαστικά είτε από Ονοματικές Οντότητες. Όσον αφορά την αποκάλυψη των λανθανουσών σχέσεων εντός των υπό μελέτη ΣΟΓ, παρατηρήσαμε ότι, αν και τα ρήματα δεν είχαν έντονη παρουσία, υπήρχαν άλλοι τρόποι

έκφρασης κάποιου είδους σχέσης μεταξύ των εννοιών. Ένα σημαντικό μέρος ήταν οι Προθέσεις όπως οι λέξεις *από, για, σε, σε, με, με, με βάση, σε, κατά, λόγω και με*. Μερικά παραδείγματα τέτοιων περιπτώσεων είναι: Βιβλία για παιδιά, Μηχανήματα στη βιομηχανία, Εγκλήματα κατά της ειρήνης, Διακρίσεις με βάση το φύλο. Ως εκ τούτου, ορίσαμε ως σχέσεις τα Ρήματα, τους Συνδέσμους, τις Προθέσεις, τα Αρνητικά Μόρια, τα Συγκριτικά και Υπερθετικά Επίθετα, τα Επιρρήματα και τις Κτητικές Αντωνυμίες. Τα υπόλοιπα μέρη του λόγου που αναγνωρίστηκαν δεν συμπεριλήφθηκαν σε κάποια από αυτές τις κατηγορίες- ως εκ τούτου, δεν απεικονίζονται στα αποτελέσματα. Η παράλειψη αυτή εξηγεί επίσης γιατί ορισμένα στοιχεία που περιέχει ο Πίνακας 1 δεν αντιστοιχούν στο 100% των αντίστοιχων συνολικών ποσών. Ένας λεπτομερής κατάλογος των μερών του λόγου που αναγνωρίζει ο ετικετογράφος (tagger) ΜτΛ του ΙΕΛ είναι διαθέσιμος στη διεύθυνση [http://nlp.ilsf.gr/nlp/tagset\\_examples/tagset\\_en/](http://nlp.ilsf.gr/nlp/tagset_examples/tagset_en/).

### 5.5.2 Συγκεντρωτικά αποτελέσματα

Σε αυτή την υποενότητα παρουσιάζονται οι μετρήσεις που προέκυψαν από τη διαδικασία επισημείωσης των ΜτΛ. Για να μπορέσει ο αναγνώστης να κατανοήσει καλύτερα τα αποτελέσματα στο πλαίσιο της παρούσας μελέτης, δίνονται οι ορισμοί των όρων «εγγραφές», “tokens” και «λέξεις». Μια έγγραφη είναι κάθε αλφαριθμητική ακολουθία χαρακτήρων ή ομάδες χαρακτήρων που αντιστοιχεί ως μεμονωμένη καταχώρηση στα περιεχόμενα ενός ΣΟΓ. Δηλαδή, για τα εξεταζόμενα δεδομένα, μια έγγραφη είναι κάθε όρος του θησαυρού, κάθε υποπεδίο των επικεφαλίδων της ΕΒΕ και κάθε λεξιλογικό ισοδύναμο σε κάθε σημειογραφία του DDC. Για παράδειγμα, το «Προπόνηση (Αθλητισμός)» είναι μια έγγραφη από την ΕΒΕ. Ένα Token είναι μια σημαντική μονάδα (χαρακτήρας ή ομάδα χαρακτήρων) στην οποία αποδόθηκε ένας χαρακτηρισμός ΜτΛ, είτε πρόκειται για λέξη, είτε για ψηφίο(α) είτε για σημείο στίξης. Η έγγραφη «Προπόνηση (Αθλητισμός)» περιλαμβάνει τέσσερα tokens, δηλαδή δύο ουσιαστικά και δύο παρενθέσεις. Λέξεις είναι οι υπόλοιπες ακολουθίες χαρακτήρων, αν εξαιρέσουμε τα ψηφία και όλα τα σημεία στίξης καθώς και άλλα σύμβολα από τα αναγνωρισμένα tokens- επομένως, το προαναφερθέν παράδειγμα μετράει ως δύο λέξεις.

Ο Πίνακας 1 κατηγοριοποιείται ανά ΣΟΓ και απεικονίζει τα αριθμητικά στοιχεία, καθώς και τις μέσες τιμές και ποσοστά ανάλογα με τη φύση κάθε μέτρησης. Οι τρεις προαναφερθείσες κατηγορίες δεδομένων (εγγραφές, tokens και λέξεις) απεικονίζονται στο πρώτο μέρος (πέντε πρώτες γραμμές) του πίνακα. Επιπλέον, στο πρώτο μέρος του πίνακα, εμφανίζεται ο μέσος αριθμός λέξεων εντός κάθε έγγραφης, ανά ΣΟΓ. Στο δεύτερο του μέρος ο Πίνακας 1 αποτελείται από την κατηγοριοποίηση ΜτΛ των οντοτήτων και των σχέσεων ανά ΚΟΣ. Η στήλη «Αρ. ΜτΛ» απεικονίζει τον αριθμό των λέξεων που έχουν κατηγοριοποιηθεί στο αντίστοιχο μέρος του λόγου- η στήλη «Αρ. ΜτΛ / εγγραφές (μ.ο.)» παραθέτει τον μέσο όρο λέξεων σε κάθε ΣΟΓ σε σχέση με τον συνολικό αριθμό των καταχωρήσεων σε κάθε ΣΟΓ- τέλος, η στήλη «Αρ. ΜτΛ/ words (%)» απεικονίζει κάθε ΜτΛ ως ποσοστό του συνολικού αριθμού λέξεων σε κάθε ΣΟΓ.

Η πλειονότητα των λέξεων κατηγοριοποιήθηκε είτε ως οντότητες είτε ως σχέσεις. Συγκεκριμένα, όπως δείχνει ο Πίνακας 1 το συνολικό ποσοστό κατηγοριοποιημένων

λέξεων είναι 68,29% στο Eurovoc, 73,14% στο LCSH (EBE corpus), 67,42% στο DDC, 84,5% στις κατηγορίες του CIDOC-CRM και 87,20% στις ιδιότητες του CIDOC-CRM. Αυτή η κατηγοριοποίηση δεν περιλαμβάνει τα επίθετα (βασικά) τα οποία κατανέμονται ως εξής: Eurovoc 3.678 (24,41%), LCSH (EBE corpus) 4.929 (24,05%), DDC 3.179 (21,75%), CIDOC-CRM κλάσεις 20 (15,50%) και CIDOC-CRM ιδιότητες 23 (4,20%). Τα ποσοστά αυτά, προστιθέμενα στα ποσοστά κατανομής ΜτΛ που παρουσιάζει ο Πίνακας 1, καλύπτουν την ανάλυση του 92,70% των λέξεων Eurovoc, του 97,19% των λέξεων LCSH (EBE corpus), του 89,17% των λέξεων DDC, του 100% των λέξεων των κλάσεων του CIDOC-CRM και του 91,41% των λέξεων των ιδιοτήτων του CIDOC-CRM.

Τα δεδομένα που απεικονίζει ο Πίνακας 1 παρέχουν μόνο ένα μέρος των στοιχείων που απαιτούνται για να εξεταστούν οι λανθάνουσες σχέσεις που διέπονται από γλωσσικούς κανόνες, δηλ. μορφοσυντακτικά πρότυπα, στα ΣΟΓ. Τα παρατηρήσιμα συντακτικά και σημασιολογικά μοτίβα (σε σύγκριση με μοτίβα που σπάνια ή ποτέ δεν εμφανίζονται στη φυσική γλώσσα) αποτελούν εκδήλωση της ανθρώπινης γλωσσικής επεξεργασίας και μπορούν να χρησιμεύσουν ως βάση για επαγωγές σχετικά με τη νοητική επεξεργασία (Wacholder & Liu, 2006, σελ. 1568). Αυτό το είδος έρευνας έχει διάφορες εφαρμογές. Οι Lioma και Ounis (2008) ασχολήθηκαν με την αναδιατύπωση ερωτημάτων και πρότειναν μια τεχνική που ονομάζεται μορφοσυντακτική αναδιατύπωση ερωτημάτων και βασίζεται στη συσχέτιση ενός Λεξικού με τα αντίστοιχα ρηγά συντακτικά στοιχεία<sup>35</sup> της λεξικής αναπαράστασης των ερωτημάτων. Αν και οι τάσεις της εποχής πρότειναν ότι, στις φυσικές γλώσσες, οι λέξεις είναι οι φορείς του νοήματος, ενώ οι προτάσεις είναι απλώς ρηχοί συντακτικοί σχηματισμοί των μερών του λόγου, υποστήριξαν ότι οι συντακτικές πληροφορίες (δηλαδή η κατηγοριοποίηση σύμφωνα με τα μέρη του λόγου) θα μπορούσαν να αξιοποιηθούν για την αναδιατύπωση ερωτημάτων, ιδίως στην περίπτωση της ανάκτησης πληροφοριών κειμένου. Πιο συγκεκριμένα, υποστηρίζουν ότι η συνεκτίμηση του συντακτικού μπορεί να αυξήσει τη δυνατότητα ανάκτησης πιο σχετικών στοιχείων, καθώς παίζει σημαντικό ρόλο (αν και έμμεσο) στη μεταφορά νοήματος, καθορίζοντας τις σχέσεις που συνδέουν τις λέξεις σε συνεκτικές προτάσεις. Εν ολίγοις, δέχονται ότι οι ρηχές συντακτικές πληροφορίες μπορούν να μοντελοποιήσουν τη δομή της γλώσσας, αποκαλύπτοντας τα μοτίβα εκείνα που είναι πιο πιθανό να εμφανιστούν ή/και να συνυπάρχουν. Πιο συγκεκριμένα, οι ρηχές συντακτικές δομές είναι τμήματα προτάσεων οι λεξικές οντότητες των οποίων έχουν αντικατασταθεί από την αντίστοιχη ρηχή συντακτική κατηγορία τους, προκειμένου να σχηματίσουν τα λεγόμενα τμήματα λόγου (part-of-speech blocks, POS blocks). Συνοπτικά, επιβεβαιώνουν ότι η υψηλή συχνότητα των ρηχών συντακτικών θραυσμάτων αντιστοιχεί σε λεξιλογικές μονάδες που φέρουν νόημα. Και, αν και ορισμένες γλώσσες μπορεί να μοιράζονται περιορισμένα χαρακτηριστικά, εντούτοις, σύμφωνα με τον

---

<sup>35</sup> Η ρηχή συντακτική πληροφορία (shallow syntactic information) αναφέρεται στις συντακτικές πληροφορίες που περιγράφουν τη μορφολογική και βασική συντακτική σχέση των λέξεων σε μια πρόταση, ξεκινώντας από τα μέρη του λόγου (ΜτΛ) που την αποτελούν.

Murphy, οι γλώσσες τείνουν να έχουν παρόμοιες μορφοσυντακτικές ιδιότητες (Murphy, 2002, σελ. 435).

Ο Πίνακας 2 παραθέτει, με φθίνουσα σειρά, τις είκοσι πιο συχνές εμφανίσεις συνδυασμών ΜτΛ, δηλαδή τα συντακτικά μοτίβα, για κάθε αναλυόμενο ΣΟΓ και την οντολογία. Το ποσοστό εμφάνισης κάθε συνδυασμού εμφανίζεται στην παρένθεση. Επιπλέον, η ανάλυση των μοτίβων αποκαλύπτει την κατανομή των μοναδικών συντακτικών μοτίβων ανά ΣΟΓ. Αναλυτικότερα, ο Euronoc έχει 286 μοναδικούς συνδυασμούς POS (που δίνει κατά μέσο όρο 24,06 καταχωρήσεις ανά μοτίβο, κατά μέσο όρο), το LCSH 474 (που είναι 21,75 καταχωρήσεις ανά μοτίβο) και το DDC 1.128 (μέσος όρος 3,38 καταχωρήσεις ανά μοτίβο), ενώ το CIDOC-CRM έχει 8 κλάσεις (μέσος όρος 10,13 καταχωρήσεις ανά μοτίβο) και 41 ιδιότητες (μέσος όρος 5,44 καταχωρήσεις ανά μοτίβο). Όπως προκύπτει, τα μοτίβα ΜτΛ στο Euronoc φαίνεται να είναι πιο συνεπή ως προς τη μορφή τους- στο LCSH τα συντακτικά πρότυπα είναι κάπως πιο χαλαρά, ενώ στο DDC τα επαναλαμβανόμενα συντακτικά πρότυπα κατέχουν ένα μάλλον μικρό ποσοστό. Ως ένα βαθμό, η παρατήρηση αυτή επιβεβαιώνει ότι ο φορμαλισμός επί των λεκτικών αναπαραστάσεων, στις λεκτικές γλώσσες ευρετηρίασης, οδηγεί σε μείωση της ποικιλίας των μοτίβων.

Όσον αφορά τα μέρη του λόγου που κατηγοριοποιήθηκαν ως σχέσεις, τα οποία εμφανίζει ο Πίνακας 1, η περαιτέρω ανάλυση δείχνει ότι, με εξαίρεση τις ιδιότητες CIDOC-CRM, πολύ λίγα από τα προαναφερθέντα μέρη του λόγου κατανέμονται στα είκοσι πιο συχνά συντακτικά μοτίβα. Πιο συγκεκριμένα, τα ρήματα είναι παρόντα μόνο στις ιδιότητες CIDOC-CRM, ενώ εμφανίζονται και στα είκοσι μοτίβα. Οι σύνδεσμοι εμφανίζονται σε δύο μοτίβα στο LCSH, σε πέντε μοτίβα στο DDC και σε δύο μοτίβα στις ιδιότητες CIDOC-CRM- οι κατηγορίες CIDOC-CRM και ο Euronoc δεν φιλοξενούν συνδέσμους στα πιο συχνά μοτίβα τους. Οι προθέσεις εμφανίζονται σε τρία μοτίβα του Euronoc, σε δύο μοτίβα του LCSH καθώς και του DDC, και σε επτά μοτίβα στις ιδιότητες του CIDOC-CRM. Τα επιρρήματα εμφανίζονται μόνο μία φορά σε κάθε ΣΟΓ και στις κατηγορίες CIDOC-CRM, ενώ στις ιδιότητες CIDOC-CRM καταγράφονται δύο μοτίβα με επιρρήματα. Τα δύο υπόλοιπα μέρη του λόγου, δηλαδή τα αρνητικά μόρια και οι κτητικές αντωνυμίες, δεν εμφανίζονται στα είκοσι πιο συχνά συντακτικά μοτίβα.

Ακολουθούν οι σχετικοί πίνακες και στη συνέχεια μια πιο λεπτομερής συγκριτική ανάλυση αναφορικά με την εκφραστικότητα των ΣΟΓ.

	Eurovoc			LC (μέσω ΕΒΕ)			DDC			CIDOC (κλάσεις)			CIDOC (ιδιότητες)		
Αριθμός εγγραφών	6882			10308			3811			81			223		
- Αριθμός tokens	15234			23670			16414			129			552		
- Αριθμός λέξεων	15067			20497			14613			129			547		
- Λέξεις ανά εγγραφή (μ.ο.)	2,19			1,99			3,83			1,59			2,45		
	Αρ. ΜτΛ	Αρ. ΜτΛ / εγγραφές (μ.ο.)	Αρ. ΜτΛ / words (%)	Αρ. ΜτΛ	Αρ. ΜτΛ / εγγραφές (μ.ο.)	Αρ. ΜτΛ / words (%)	Αρ. ΜτΛ	Αρ. ΜτΛ / εγγραφές (μ.ο.)	Αρ. ΜτΛ / words (%)	Αρ. ΜτΛ	Αρ. ΜτΛ / εγγραφές (μ.ο.)	Αρ. ΜτΛ / words (%)	Αρ. ΜτΛ	Αρ. ΜτΛ / εγγραφές (μ.ο.)	Αρ. ΜτΛ / words (%)
<b>ΜτΛ που ορίστηκαν ως Έννοιες/Οντότητες (σύνολα)</b>	<b>9652</b>	<b>1,40</b>	<b>64,06%</b>	<b>13321</b>	<b>1,29</b>	<b>64,99%</b>	<b>7552</b>	<b>1,98</b>	<b>51,68%</b>	<b>108</b>	<b>1,33</b>	<b>83,72%</b>	<b>112</b>	<b>0,50</b>	<b>20,48%</b>
- Ουσιαστικά, κοινά	8443	1,23	56,04%	11825	1,15	57,69%	6443	1,69	44,09%	105	1,30	81,40%	111	0,50	20,29%
- Ουσιαστικά, κύρια	653	0,09	4,33%	1209	0,12	5,90%	997	0,26	6,82%	1	0,01	0,78%	0	0,00	0,00%
- Συντμήσεις	240	0,03	1,59%	11	0,00	0,05%	5	0,00	0,03%	1	0,01	0,78%	0	0,00	0,00%
- Αντωνυμίες <sup>36</sup>	2	0,00	0,01%	0	0,00	0,00%	12	0,00	0,08%	0	0,00	0,00%	0	0,00	0,00%
- Λοιπά, (ξένες λέξεις)	314	0,05	2,08%	276	0,03	1,35%	95	0,02	0,65%	1	0,01	0,78%	1	0,00	0,18%
<b>ΜτΛ που ορίστηκαν ως Σχέσεις (σύνολα)</b>	<b>637</b>	<b>0,09</b>	<b>4,23%</b>	<b>1671</b>	<b>0,16</b>	<b>8,15%</b>	<b>2300</b>	<b>0,60</b>	<b>15,74%</b>	<b>1</b>	<b>0,01</b>	<b>0,78%</b>	<b>365</b>	<b>1,64</b>	<b>66,73%</b>
- Ρήματα	18	0,00	0,12%	34	0,00	0,17%	166	0,04	1,14%	0	0,00	0,00%	234	1,05	42,78%
- Σύνδεσμοι	98	0,01	0,75%	614	0,06	3,00%	970	0,25	6,64%	0	0,00	0,00%	14	0,06	2,56%
- Προθέσεις	368	0,05	2,44%	726	0,07	3,54%	796	0,21	2,44%	0	0,00	0,00%	93	0,42	17,00%
- Μόρια, αρνητικά	35	0,01	0,23%	17	0,00	0,08%	30	0,01	0,21%	0	0,00	0,00%	0	0,00	0,00%
- Επίθετα <sup>37</sup>	16	0,00	0,11%	22	0,00	0,11%	83	0,02	0,57%	0	0,00	0,00%	6	0,03	1,10%
- Επιρρήματα	82	0,01	0,54%	213	0,02	1,40%	220	0,06	1,51%	1	0,01	0,78%	17	0,08	3,11%
- Αντωνυμίες, κτητικές	5	0,00	0,03%	27	0,00	0,13%	12	0,00	0,08%	0	0,00	0,00%	1	0,00	0,18%

Πίνακας 1: κατανομή των μερών του λόγου (ΜτΛ) για κάθε ΣΟΓ, βάσει της αυτόματης κατάταξης του NLP tagger του ΙΕΛ.

<sup>36</sup> Αναφορικές αόριστες, Προσωπικές και Δεικτικές

<sup>37</sup> Συγκριτικού και Υπερθετικού βαθμού

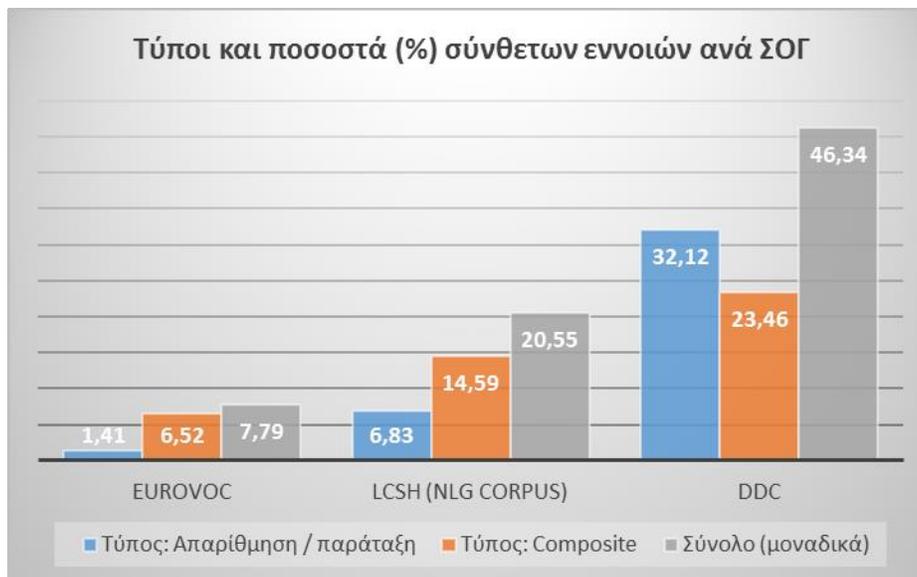
Eurovoc		LCSH (ΕΒΕ)		DDC		CIDOC (Κλάσεις)		CIDOC (Ιδιότητες)	
Μτλ μοτίβο	Απόλυτος αριθμός (ποσοστό)	Μτλ μοτίβο	Απόλυτος αριθμός (ποσοστό)	Μτλ μοτίβο	Απόλυτος αριθμός (ποσοστό)	Μτλ μοτίβο	Απόλυτος αριθμός (ποσοστό)	Μτλ μοτίβο	Απόλυτος αριθμός (ποσοστό)
Adj + N	2182 (31,71%)	N	3353 (32,53%)	N	634 (16,64%)	N	38 (46,91%)	V + Adp	59 (26,46%)
N	1364 (19,82%)	Adj + N	1936 (18,78%)	Adj + N	387 (10,15%)	N + N	23 (28,40%)	V	40 (17,94%)
N + N	819 (11,90%)	N + Punct + Adj	677 (6,57%)	N + N	156 (4,09%)	Adj + N	13 (16,05%)	V + N + Art	21 (9,42%)
N + Art + N	454 (6,60%)	N + N	548 (5,32%)	Dig	112 (2,94%)	N + Adj + N	2 (2,47%)	V + N	17 (7,62%)
Res	191 (2,78%)	N + OPunct + N + CPunct	369 (3,58%)	N + Conj + N	112 (2,94%)	Adj + Adj + N	2 (2,47%)	V + N + Adp	11 (4,93%)
N + Adj + N	171 (2,48%)	N + Adp + N	361 (3,50%)	Adj	63 (1,65%)	N + Res + Abr	1 (1,23%)	V + Adj + N	8 (3,59%)
Adj + Adj + N	143 (2,08%)	N + Conj + N	311 (3,02%)	N + Adj + N	61 (1,60%)	N + Adj	1 (1,23%)	V + N + N	6 (2,69%)
N + Adp + N	118 (1,71%)	Adj	261 (2,53%)	Adj + Adj + N	50 (1,31%)	Adv + N	1 (1,23%)	V + Conj + V + N + Art	5 (2,24%)
Adj	103 (1,50%)	Adj + Adj	134 (1,30%)	N + Art + N	44 (1,15%)	-	-	V + Adj + N + Art	5 (2,24%)
Adj + N + N	93 (1,35%)	Adj + N + Punct + Adj	124 (1,20%)	Adj + N + Conj + N	40 (1,05%)	-	-	V + Conj + V + N	4 (1,79%)
N + Art + Adj + N	68 (0,99%)	N + Punct + N	120 (1,16%)	Adj + Conj + Adj + N	37 (0,97%)	-	-	V + Art + N + Adp	4 (1,79%)
Adj + Adj	59 (0,86%)	N + Art + N	108 (1,05%)	N + Punct + N + Punct + N	33 (0,87%)	-	-	V + Adv + N	4 (1,79%)
N + Adp + Art + N	42 (0,61%)	N + Adj + N	85 (0,82%)	Adv + N	31 (0,81%)	-	-	V + Adj + Adj	4 (1,79%)
N + Art + Abr	42 (0,61%)	N + OPunct + Adj + N + CPunct	77 (0,75%)	N + Conj + Adj + N	31 (0,81%)	-	-	N + V + Adp	3 (1,35%)
N + N + N	39 (0,57%)	N + Res	72 (0,70%)	Adj + N + N	28 (0,73%)	-	-	V + V + Adp	2 (0,90%)
Adj + N + Adj + N	37 (0,54%)	Adv + N	67 (0,65%)	N + Conj + N + N	28 (0,73%)	-	-	V + Art + N + Art	2 (0,90%)
Adv + N	37 (0,54%)	N + Punct + N + Art	60 (0,58%)	Adj + Adj	27 (0,71%)	-	-	V + Adv	2 (0,90%)
Adj + N + OPunct + Abr + CPunct	36 (0,52%)	Adj + Adj + N	48 (0,47%)	N + Adp + Adj + N	27 (0,71%)	-	-	V + Adp + N	2 (0,90%)
N + Adp + Adj + N	36 (0,52%)	Adj + N + Adp + N	48 (0,47%)	N + Adp + N	27 (0,71%)	-	-	V + Adj + Adp	2 (0,90%)
N + Res	34 (0,49%)	Adj + N + Conj + N	45 (0,44%)	Res	26 (0,68%)	-	-	V + V	1 (0,45%)
<b>Sum of 20</b>	<b>6068 (88,17%)</b>	<b>Sum of 20</b>	<b>8804 (85,41%)</b>	<b>Sum of 20</b>	<b>1954 (51,27%)</b>	<b>Sum</b>	<b>81 (100%)</b>	<b>Sum of 20</b>	<b>202 (90,58%)</b>

Πίνακας 2: τα είκοσι πιο κοινά συντακτικά μοτίβα (διατεταγμένα κατά συχνότητα)

Συντομογραφίες: Abr = Σύντμηση · Adj = Επίθετο · Adp = Πρόθεση · Art = Άρθρο · Conj = Σύνδεσμος · CPunct = Τερματισμός στίξης · Dig = Αριθμοί · N = Ουσιαστικό · OPunct = Έναρξη στίξης · Punct = Στίξη · Res = Λοιπά (ξένη λέξη) · V = Ρήμα.

Βασιζόμενοι σε γλωσσικά κριτήρια, μέσω της μορφοσυντακτικής ανάλυσης, διακρίναμε δύο βασικές κατηγορίες εννοιών. Η πρώτη κατηγορία αφορά έννοιες οι οποίες είναι γλωσσικά αδιαίρετες (atomic) ενώ η δεύτερη κατηγορία αφορά έννοιες οι οποίες είναι γλωσσικά σύνθετες (complex) (Mastora κ.ά., 2017). Εν συντομία, οι αδιαίρετες (atomic) έννοιες είναι οι εννοιολογικά αδιαίρετες μονάδες όπως γίνονται αντιληπτές στο πλαίσιο της φυσικής γλώσσας- σε αυτή την κατηγορία εμπίπτουν επίσης συντακτικά σύνθετοι όροι (compounds) που είναι όροι που μπορούν να διασπαστούν μορφολογικά σε ξεχωριστά συστατικά (λέξεις), όπως ο όρος της ΕΒΕ «Ωράριο εργασίας», ωστόσο ο όρος στο σύνολό του αντιπροσωπεύει μια ενιαία έννοια. Συνοπτικά, αδιαίρετες (atomic) έννοιες αποτελούν όλοι οι μονολεκτικοί όροι και οι πολυλεκτικοί όροι που διαθέτουν τα εξής χαρακτηριστικά: συνδυασμοί δύο ουσιαστικών όταν το δεύτερο βρίσκεται σε γενική (π.χ. ορυχεία χαλκού), επίθετο μαζί με ουσιαστικό, επίθετο μαζί με δύο ουσιαστικά (π.χ. διακυβερνητική νομική πράξη) είτε το δεύτερο ουσιαστικό να είναι στη γενική είτε όλες οι λέξεις να είναι στην ίδια πτώση (Μάστορα, 2016).

Από την άλλη πλευρά, οι σύνθετες έννοιες είναι μονάδες που είναι εννοιολογικά διαιρετές και περιλαμβάνουν δύο υποκατηγορίες. Η μία είναι η «απαρίθμηση/παράταξη», η οποία απαριθμεί παραλλαγές εννοιών χωρίς να δηλώνει ρητά κανενός είδους σχέσεις μεταξύ τους. Τέτοιο παράδειγμα είναι ο όρος του DDC «Ηφαιστεια, σεισμοί, θερμά ύδατα και αέρια». Εδώ κρίσιμα στοιχεία είναι τα κόμματα και οι σύνδεσμοι. Η δεύτερη υποκατηγορία των σύνθετων εννοιών είναι οι περίπλοκες «composites» έννοιες που περιλαμβάνουν δύο ή περισσότερες έννοιες της φυσικής γλώσσας με κάποιου είδους σχέση μεταξύ τους- ο όρος της ΕΒΕ «Εγκλήματα κατά της ανθρωπότητας είναι ένα τυπικό παράδειγμα περίπλοκης έννοιας. Στην περίπτωση αυτή τα κρίσιμα στοιχεία είναι η ύπαρξη προθέσεων. Με βάση την προαναφερθείσα κατηγοριοποίηση των εννοιών, τα δεδομένα που προκύπτουν από τη μορφοσυντακτική ανάλυση καθενός από τους συμβατικούς ΣΟΓ αποκαλύπτουν το ποσοστό (βλέπε Γράφημα 5) των σύνθετων εννοιών (complex concepts) εντός κάθε ΣΟΓ.



**Γράφημα 5:** σύνθετες έννοιες ανά ΣΟΓ

Στο σημείο αυτό, κρίνεται σκόπιμο να διασαφηνιστεί ότι στο Γράφημα 5 δεν απεικονίζονται δεδομένα για το CIDOC-CRM, επειδή ούτε οι κλάσεις CIDOC ούτε οι ιδιότητες CIDOC είναι έννοιες, με το νόημα που έχουν οι έννοιες στα ΣΟΓ. Επομένως, μια σύγκριση αυτού του τύπου θα οδηγούσε σε παραπλανητικά συμπεράσματα. Μια διεξοδική ανάλυση αυτής της θέσης παρουσιάζεται στην υποενότητα «Οντολογίες: CIDOC-CRM» που ακολουθεί.

### 5.5.3 Συγκριτική ανάλυση επιλεγμένων ΣΟΓ και αντιπαραβολή με οντολογική προσέγγιση

Πριν προχωρήσουμε στην κριτική κάθε εξεταζόμενου ΣΟΓ, εξετάζονται ορισμένα γενικά ζητήματα που ισχύουν για όλα τα ΣΟΓ. Τα ζητήματα αυτά αφορούν βασικά τις σχέσεις και όχι τις έννοιες, κυρίως επειδή οι έννοιες έχουν διαφορετική φύση, ενώ οι σχέσεις μοιράζονται πολλά κοινά χαρακτηριστικά. Όπως έχει ήδη αναφερθεί στην υποενότητα 5.4, τα ΣΟΓ ορίζουν τρεις τύπους σχέσεων: ισοδυναμίας, ιεραρχίας και συσχέτισης.

Οι σχέσεις ισοδυναμίας διαχειρίζονται διαφορετικές ονομασίες της ίδιας έννοιας ή οιονεί συνώνυμα που αποτελούν μια ενιαία έννοια στο πλαίσιο του ΣΟΓ. Παρά το γεγονός ότι οι σχέσεις αυτές φαίνονται αρκετά σαφείς, φέρουν έναν ορισμένο βαθμό ασάφειας, καθώς δεν διευκρινίζουν αν η διασταύρωση αφορά ένα συνώνυμο, δηλαδή τη σχέση μιας λέξης με μια εναλλακτική - για παράδειγμα «μουρλός» βλ. «τρελός»- ή έναν σχετικό όρο (έννοια)- για παράδειγμα «αρένα» βλ. «στάδιο». Σύμφωνα με αυτό το σκεπτικό, ένα ΣΟΓ μπορεί να μην καταγράφει ρητά αν μια αναφορά αποδίδεται από έναν στενότερο όρο σε έναν ευρύτερο όρο. Για παράδειγμα στον Ευγονος υπάρχει η δήλωση για «πεπόνι» βλέπε «οπωροκηπευτικό». Άρα αντιμετωπίζεται ως σχέση ισοδυναμίας μια σχέση που είναι ιεραρχική.

Όμως οι ιεραρχικές σχέσεις είναι από τη φύση τους πιο σύνθετες. Οι ιεραρχίες είναι αρκετά γενικές και η ερμηνεία της ιεραρχίας κάπως διφορούμενη. Τρεις βασικοί τύποι ιεραρχιών περιλαμβάνονται στα εξεταζόμενα ΣΟΓ, χωρίς όμως να αναφέρεται ρητά ποια από αυτές τις ιεραρχίες ισχύει σε κάθε περίπτωση. Η πρώτη σχέση είναι η σχέση «is a», η οποία είναι μια τυπική σχέση είδους προς γένος. Η σχέση αυτή βασίζεται στην κληρονομικότητα των ιδιοτήτων και η χρήση της δηλώνει ότι η έννοια κληρονομεί τις ιδιότητες της γενικότερης έννοιας. Για παράδειγμα το πουλόβερ είναι ένα είδος ρούχου. Η δεύτερη σχέση είναι η σχέση «μέρος του», η οποία εφαρμόζεται όταν ένα μέρος ανήκει σε ένα συγκεκριμένο σύνολο - για παράδειγμα, το πόδι είναι μέρος του σώματος. Η τρίτη σχέση είναι εκείνη του «στιγμιότυπου» (instance of), η οποία εφαρμόζεται όταν ένα συγκεκριμένο άτομο σχετίζεται με την κλάση στην οποία ανήκει - για παράδειγμα, το Λονδίνο είναι ένα στιγμιότυπο της κλάσης πόλη.

Αυτές οι τρεις υποκατηγορίες ιεραρχικών σχέσεων θα μπορούσαν να εξειδικευτούν περαιτέρω. Ο Johansson αναγνωρίζει τέσσερις διαφορετικούς τύπους της σχέσης “is a”. Ο πρώτος είναι η υπαγωγή κάτω από ένα γένος, που είναι ο παραδοσιακός τρόπος δημιουργίας δέντρων ταξινόμησης των φυσικών ειδών- ειδικότερα, για τη δημιουργία των περίφημων ιεραρχιών των φυτών και των ζώων στη βιολογία. Η δεύτερη, την οποία ονομάζει «προσδιορίσιμη-υπαγωγή» (determinable-subsumption), δεν αφορά τα

φυσικά είδη, αλλά τις ποιότητες (ιδιότητες) που οδηγούν σε διαφορετικού είδους γενικεύσεις. Για παράδειγμα, ως προσδιοριστικό, το κατακόκκινο ανήκει στο κόκκινο. Οι άλλες δύο σχέσεις είναι η «εξειδίκευση του» (“specification of”) και η «εξειδικεύω το» (specialization of)- για παράδειγμα, το «ζωγραφίζω προσεκτικά» είναι εξειδίκευση του “ζωγραφίζω” και το «ζωγραφίζω ένα τραπέζι» εξειδικεύει το τι ζωγραφίζω (Johansson, 2008).

Η τρίτη κατηγορία σχέσεων είναι εκείνη της γενικής συσχέτισης, η οποία είναι η πιο ασαφής και διφορούμενη. Η σχέση μεταξύ δύο εννοιών δηλώνει μια σημασιολογική σύνδεση η οποία όμως δεν προσδιορίζεται περαιτέρω, άρα δεν δηλώνονται λεπτομέρειες ως προς την ακριβή σχέση μεταξύ των Α και Β.

Από τα παραπάνω συνάγεται ότι ακόμη και οι πιο σαφώς καθορισμένες ιεραρχικές σχέσεις εξακολουθούν να διατηρούν ένα ποσοστό ασάφειας. Οι ταξινομίες που δημιουργούνται στον τομέα του LIS, για την κατηγοριοποίηση αυτού του είδους των σχέσεων, αποτελούν απόδειξη της ασαφούς λογικής που κυριαρχεί στις ιεραρχικές σχέσεις (Peters & Weller, 2008, Εικόνα 8). Επίσης, είναι γεγονός ότι τα ΣΟΓ χρησιμοποιούν μόνο ένα μικρό αριθμό παραδειγματικών σημασιολογικών σχέσεων μέχρι στιγμής (Peters & Weller, 2008, σελ. 100).

Κάθε μία από τις επόμενες υποενότητες είναι αφιερωμένη σε ένα συγκεκριμένο ΣΟΓ, όπου παρουσιάζεται λεπτομερής ανάλυση των αποτελεσμάτων. Ακολουθεί σύγκριση και κριτική των συνόλων δεδομένων που μελετήθηκαν.

### 5.5.3.1 Θησαυροί: Eurovoc

Σύμφωνα με το ISO 25964, ο θησαυρός είναι ένα ελεγχόμενο και δομημένο λεξιλόγιο στο οποίο οι έννοιες αντιπροσωπεύονται από όρους, οργανωμένους έτσι ώστε οι σχέσεις μεταξύ των εννοιών να είναι σαφείς και οι προτιμώμενοι όροι να συνοδεύονται από εισαγωγικά λήμματα για συνώνυμα ή οιονεί συνώνυμα (quasi-synonyms) (ISO, 2011, σελ. 12). Οι θησαυροί χρησιμοποιούνται στις βιβλιοθήκες για περισσότερο από μισό αιώνα (Aitchison & Clarke, 2004) και μπορεί να διαφέρουν, τόσο ως προς τη δομή όσο και ως προς τη σημασιολογία τους (Doerr, 2001).

Σε ένα ορισμένο επίπεδο αφαίρεσης, κάθε όρος σε έναν θησαυρό αντιπροσωπεύει μια έννοια. Ο όρος μπορεί να είναι μία λέξη ή ένας σύνθετος όρος «που μπορεί να χωριστεί μορφολογικά σε ξεχωριστά συστατικά» (ISO, 2011, σελ. 3). Για παράδειγμα «ορυχεία σιδήρου». Αν και θα περίμενε κανείς ότι οι σύνθετοι όροι θα αποτελούσαν την εξαίρεση του κανόνα «μία έννοια - μία λέξη», τα στοιχεία που εμφανίζει ο Πίνακας 1 δείχνουν ότι υπάρχει ένας μέσος όρος 2,19 λέξεων ανά όρο στο Eurovoc, ενώ ο Πίνακας 2 δείχνει ότι μόνο 1.364 από τους 6882 όρους του Eurovoc (τουτέστιν το 19,81%) αποτελείται από ένα και μόνο ουσιαστικό. Έτσι, αν και ο αριθμός των λέξεων στο Eurovoc είναι μεγαλύτερος από το LCSH, αποδεικνύεται ότι το LCSH έχει πιο σύνθετες έννοιες από το Eurovoc, όπως φαίνεται στο Γράφημα 5 (20,55% και 7,79% αντίστοιχα). Προφανώς, οι σύνθετες εκφράσεις της φυσικής γλώσσας είναι απαραίτητες για να εκφράσουν αδιαίρετες έννοιες (atomic concepts).

Οι θησαυροί δεν διαθέτουν ενσωματωμένο μηχανισμό για τη δημιουργία ενδο-ορολογικών σχέσεων, ωστόσο η γλωσσική ανάλυση που πραγματοποιήθηκε στο πλαίσιο

της παρούσας μελέτης αποκάλυψε ορισμένες περιπτώσεις όπου οι θησαυροί χρησιμοποιούν γλωσσικές εκφράσεις για τη δημιουργία σύνθετων εννοιών. Για παράδειγμα, η χρήση των επιθέτων, όπως απεικονίζεται στους πίνακες 1 και 2, συνηγορεί ως προς αυτή την πρακτική. Τα προαναφερθέντα στοιχεία, μαζί με την παρατήρηση ότι το 4,33% των λέξεων του Eurovoc είναι κύρια ονόματα επιβεβαιώνουν την παρατήρηση του Stellato (2012) ότι, γενικά, οι θησαυροί δεν κάνουν διάκριση μεταξύ κλάσεων και περιπτώσεων αντικειμένων ή μεταξύ αντικειμένων και εννοιών.

Κλείνοντας την παρούσα ενότητα, πρέπει να σημειωθεί ότι το ISO 25964-1 επιτρέπει τη χρήση ευρύτερων όρων με ή χωρίς προσδιορισμό, δηλαδή «ευρύτερος όρος (γενικά)». Όμως επιτρέπει και τις εξειδικεύσεις «μέρος του» ή «στιγμιότυπο του» (ISO, 2011, σελ. 13). Παρόλα αυτά πολύ λίγοι θησαυροί χρησιμοποιούν αυτές τις εξειδικεύσεις των ιεραρχικών σχέσεων του ISO However (V. Alexiev κ.ά., 2016, σελ. 41) και ο Eurovoc δεν αποτελεί εξαίρεση.

### 5.5.3.2 Θεματικές επικεφαλίδες: LCSH (σώμα EBE)

Το σύστημα θεματικών επικεφαλίδων είναι η πρώτη συστηματική προσπάθεια για την παροχή θεματικής πρόσβασης μέσω λεκτικής απόδοσης των θεμάτων. Αποτελεί ένα ελεγχόμενο λεξιλόγιο του οποίου ο κύριος μηχανισμός είναι η χρήση υποδιαίρεσεων προκειμένου να εξειδικεύσει περαιτέρω τον κύριο όρο που μπορεί να είναι ένα όνομα ή μια έννοια. Η Βιβλιοθήκη του Κογκρέσου ανέπτυξε το LCSH για χρήση στους δικούς της καταλόγους. Όμως θεωρήθηκε επίσης κατάλληλο για τις πολύ μεγάλες δημόσιες βιβλιοθήκες, ορισμένα κολέγια και πολλές πανεπιστημιακές βιβλιοθήκες (Stone, 2000, σελ. 3). Έτσι, η θεματική πρόσβαση, η οποία βασιζόταν στη λεκτική ευρετηρίαση κατά τη διάρκεια του 20ού αιώνα τόσο εντός όσο και εκτός των ΗΠΑ, κυριαρχήθηκε από το LCSH (Heiner-Freiling, 2000). Ακολουθώντας το πνεύμα της εποχής ήταν ένα από τα πρώτα λεξιλόγια βιβλιοθηκών που μετατράπηκε σε SKOS (Harper, 2006), κυρίως επειδή τόσο το SKOS όσο και το LCSH/MARC έχουν ένα μοντέλο προσανατολισμένο στις έννοιες (Summers κ.ά., 2008, σελ. 24).

Τα υποπεδία των LCSH χρησιμοποιούν, κατά μέσο όρο, λιγότερες λέξεις από τον Eurovoc για την αναπαράσταση εννοιών (1,99 και 2,19 αντίστοιχα). Ενώ την ίδια στιγμή έχουν λιγότερες αδιαίρετες (atomic) έννοιες από ότι ο Eurovoc. Τα μέρη του λόγου που δηλώνουν σχέσεις βρέθηκε ότι είναι διπλάσια στην LCSH από εκείνα του Eurovoc. Ειδικά οι σύνδεσμοι είναι 4 φορές περισσότεροι (3% του συνολικού αριθμού των λέξεων). Μια ακόμη πιο προσεκτική ματιά στη χρήση των συνδέσμων αποκαλύπτει μια μάλλον ασυνεπή χρήση. Όσον αφορά ειδικότερα τη χρήση του «και», το οποίο χρησιμοποιείται σε μεγάλο βαθμό ως συνδετικός κρίκος εννοιών, δεν διευκρινίζεται η ακριβής φύση της σύνδεσης που αντιπροσωπεύει. Σε ορισμένες περιπτώσεις χρησιμοποιείται ως τομή (Boolean «AND»), όπως συμβαίνει με τα «Φύλο και νόμος», «Γεωργία και κράτος», και «Μεταφορές και κράτος». Σε άλλες περιπτώσεις χρησιμοποιείται ως ένωση (Boolean «OR»), όπως για παράδειγμα «Γλώσσα και γλώσσες», «Ερμηνεία και κριτική» ή «Δίκαιο καινομοθεσία».

Σε αντίθεση με τους θησαυρούς, οι θεματικές επικεφαλίδες φέρουν ενδο-ορολογικές σχέσεις και διαθέτουν τους αντίστοιχους μηχανισμούς για τη δημιουργία

τους. Ένας τέτοιος μηχανισμός είναι η χρήση των σημείων στίξης. Μια πιο λεπτομερής εξέταση των δεδομένων της παρούσας μελέτης δείχνει ότι η χρήση της στίξης είναι εκτεταμένη εντός των υποπεδίων, καθώς έξι (6) από τα είκοσι (20) πρότυπα περιέχουν σημεία στίξης, όπως παρενθέσεις. Με βάση τα δεδομένα που εικονίζει ο Πίνακας 2, το πέμπτο μοτίβο, που καλύπτει ποσοστό 3,58%, είναι το [N + OPunct + N + CPunct], όπου OPunct είναι το άνοιγμα μιας παρένθεσης και CPunct είναι το κλείσιμο. Όσον αφορά στη σημασία, οι παρενθέσεις χρησιμοποιούνται είτε για να καθορίσουν το πλαίσιο εντός του οποίου χρησιμοποιείται ο όρος είτε για να προσδιορίσουν περαιτέρω την προηγούμενη οντότητα. Τέτοια παραδείγματα είναι οι επικεφαλίδες «Στρες (Ψυχολογία)» και «Πόλεμος (Διεθνές δίκαιο)», όπου υπάρχει μια ενδο-ορολογική παραδειγματική σχέση. Επιπλέον, τα LCSH διαθέτουν μεθόδους, εκτός από τη χρήση παρενθετικού προσδιορισμού, για να καθορίσουν την αναφορική σημασιολογία τέτοιων όρων: για παράδειγμα, θα μπορούσε να ενσωματωθεί σε μια επικεφαλίδα τύπου "στα", π.χ. «Λοιμώξεις στα παιδιά», ή να καθιερωθεί ως μέρος μιας επικεφαλίδας φράσης, π.χ. «Λοιμώξεις παιδιών» - ή να προσδιοριστεί από μια υποκατηγορία, π.χ. «Παιδιά -- Λοιμώξεις» (Svenonius, 2000, σελ. 21). Παρατηρείται ότι όλα τα παραπάνω έχουν επίδραση στην αλφαβητική ταξινόμηση, η οποία φαίνεται αρκετά σημαντική, ακόμη και για την εποχή μας. Η σημασία αυτή καταγράφεται ιδιαίτερα, μεταξύ άλλων, στη χρήση του τρίτου συχνότερου συντακτικού προτύπου, δηλαδή του «N + Punct + Adj», όπου το η στίξη υπαγορεύει τη χρήση κόμματος. Η αντιστροφή ουσιαστικού-επιθέτου εφαρμόζεται έτσι ώστε ο επικεφαλίδα να ταξινομείται αλφαβητικά κάτω από το ουσιαστικό.

Ο Πίνακας 1 και ο Πίνακας 2 δείχνουν ότι υπάρχει αξιοσημείωτη χρήση των Προθέσεων στα LCSH, καταλαμβάνοντας την πρώτη θέση μεταξύ των συμβατικών ΣΟΓ. Αυτός ο συγκεκριμένος τύπος ΜτΛ χρησιμοποιείται για να συνδέσει ουσιαστικά που περιλαμβάνουν μια σύνθετη έννοια. Για παράδειγμα «Εγκλήματα κατά της ιδιοκτησίας» ή «Αθλητισμός για ηλικιωμένους».

Μια άλλη παρατήρηση είναι ότι οι επικεφαλίδες που μοιάζουν πολύ από γλωσσική άποψη έχουν σημαντικές σημασιολογικές διαφορές. Ένα αντιπροσωπευτικό παράδειγμα αυτής της περίπτωσης είναι οι επικεφαλίδες «Φιλοσοφία και θρησκεία» και «Θρησκεία -- Φιλοσοφία». Όπως αναφέρεται στη σημείωση πεδίου εφαρμογής της δεύτερης εγγραφής καθιερωμένου τύπου «*Εδώ έχουν εισάγονται έργα σχετικά με τη φιλοσοφία της θρησκείας. Έργα σχετικά με την αμοιβαία σχέση και επιρροή μεταξύ φιλοσοφίας και θρησκείας καταχωρούνται στην ενότητα Φιλοσοφία και θρησκεία*».

Οι προαναφερθείσες παρατηρήσεις υπογραμμίζουν τις προσπάθειες τυποποίησης που επικεντρώνονται στην τιμή της επικεφαλίδας - ως συμβολοσειρά - προκειμένου να εκφραστεί η διαφορετική σημασιολογία. Είναι προφανές ότι η πρακτική αυτή απαιτεί άρρητη γνώση προκειμένου να ερμηνευτεί, ενώ δεν υπάρχει μέθοδος ή μηχανισμός ρητής δήλωσης των σχέσεων μεταξύ των στοιχείων. Αντ' αυτού, αυτό μπορεί να επιτευχθεί με το συνδυασμό τιμών με οποιονδήποτε τρόπο θεωρείται κατάλληλος. Ως εκ τούτου, ο αφορισμός του Kwaśnik ότι «η LCSH είναι ένα συνονθύλευμα όρων και σχέσεων που δημιουργήθηκε σταδιακά κατά τη διάρκεια ενός αιώνα» (Kwaśnik, 2011, σελ. 21) δεν απέχει και πολύ από την πραγματικότητα.

### 5.5.3.3 Ταξινομικά συστήματα: DDC

Τα ταξινομικά συστήματα αρχικά αναπτύχθηκαν και χρησιμοποιήθηκαν από τις βιβλιοθήκες για τη διαχείριση της ταξινόμησης των φυσικών αντικειμένων, κυρίως των βιβλίων μιας συλλογής. Αργότερα εξελίχθηκαν σε πολυλειτουργικά εργαλεία και, σήμερα, χρησιμοποιούνται ακόμη και στο πλαίσιο των ψηφιακών βιβλιοθηκών, όπου φυσικά αντικείμενα και ράφια δεν έχουν εφαρμογή. Μεταξύ των ταξινομικών συστημάτων, το DDC είναι πολύ δημοφιλές, χρησιμοποιείται σε περισσότερες από 135 χώρες και έχει μεταφραστεί σε περισσότερες από τριάντα γλώσσες σύμφωνα με τα στοιχεία της OCLC. Προκειμένου να είναι αξιοποιήσιμο στο πλαίσιο του σημασιολογικού ιστού, το DDC παρέχεται επίσης ως διασυνδεδεμένα δεδομένα, και η προσπάθεια αντιστοίχισης κάθε αριθμού σε διαφορετικό URI χρονολογείται εδώ και αρκετά χρόνια (Panzer, 2008).

Ο βασικός μηχανισμός του DDC είναι η δημιουργία μιας δενδροειδούς ιεραρχικής δομής με τη χρήση σημειογραφίας. Η σημειογραφική ιεραρχία, όπως αναφέρεται στην εισαγωγή του DDC, εκφράζεται με βάση το μήκος της σημειογραφίας. Οι αριθμοί σε οποιοδήποτε επίπεδο είναι συνήθως υποδεέστεροι μιας κλάσης της οποίας η σημειογραφία είναι κατά ένα ψηφίο μικρότερη και η δομική ιεραρχία σημαίνει ότι όλα τα θέματα (εκτός από τις δέκα κύριες κλάσεις) αποτελούν μέρος των ευρύτερων θεμάτων που βρίσκονται πάνω από αυτά. Άρα ότι ισχύει για το σύνολο ισχύει και για τα μέρη. Αυτή η αλυσίδα δημιουργεί μια μονο-ιεραρχική σχέση όπου κάθε κλάση μπορεί να έχει μόνο μία ευρύτερη κλάση<sup>38</sup>.

Σύμφωνα με την εισαγωγή του DDC (αρχή 4.14), δεδομένου ότι τα μέρη του DDC είναι διατεταγμένα κατά επιστημονικό πεδίο και όχι κατά θέμα, ένα θέμα μπορεί να εμφανίζεται σε περισσότερες από μία κατηγορίες. Όπως όλα τα προσυνδυασμένα συστήματα, το DDC ορίζει εκ των προτέρων την επιλογή μιας κυρίαρχης κατηγορίας κάτω από την οποία περιλαμβάνονται όλες οι πιθανές υποδιαίρεσεις. Τα προβλήματα φαίνεται να αυξάνονται όταν ένα τεκμήριο πρέπει να αντιστοιχιστεί σε μια κυρίαρχη κατηγορία η οποία παρέχεται ως υποκατηγορία εντός του συστήματος. Όπως αναφέρει ο Hjørland σχετικά με το θέμα αυτό: «Ισχυρίζομαι ότι αυτός ο τύπος ανάλυσης, ο οποίος καθορίζει τις προτεραιότητες των οπτικών που πρέπει να ληφθούν για ένα έγγραφο, δεν είναι βέλτιστος σε κάθε περίπτωση. Μπορεί κανείς να φανταστεί ερευνητές που ασχολούνται με τεχνικές πτυχές της εκλογικής διαδικασίας και οι οποίοι επιθυμούν να τις συγκρίνουν σε διάφορες χώρες. Για έναν τέτοιο άνθρωπο οι εκλογές θα ήταν το κεντρικό θέμα, και θα ήταν άβολο αν αυτό αποτελούσε υποδιαίρεση του θέματος της Ιστορίας και της Ινδίας» (Hjørland, 1992, σελ. 178).

Επιπλέον, οι Green & Panzer (2011) εξέτασαν τους τύπους των ιεραρχικών σχέσεων εντός των κατηγοριών του DDC και εντόπισαν διάφορους τύπους σχέσεων - για παράδειγμα, σχέσεις μέρους όλου ή σχέσεις εξειδίκευσης - για τις οποίες δεν εφαρμόστηκε κάποιος συγκεκριμένος φορμαλισμός. Επιπλέον, ο Mazzochi αναφέρει ότι

---

<sup>38</sup> Βεβαίως, αυτό δεν είναι απολύτως ακριβές αφού υπάρχουν εξαιρέσεις, όπως οι διεπιστημονικοί αριθμοί.

«οι ιεραρχίες, που χρησιμοποιούνται από ταξινομικά συστήματα όπως η Δεκαδική Ταξινόμηση Dewey, λειτουργούν πιο απρόβλεπτα. Συνήθως, δεν είναι εφοδιασμένες με τις ίδιες λογικές ιδιότητες των γενικών ιεραρχιών» (Mazzocchi, 2017, σελ. 370). Συμπερασματικά, δεν υπάρχει ισχυρός φορμαλισμός στους τύπους με τους οποίους οι κατηγορίες (ή τα νοηματικά μέρη αυτών) συνδέονται μεταξύ τους στο DDC.

Είναι αναμενόμενο ότι ένα σύστημα σημειογραφίας θα παρουσιάζει περιορισμένη εξάρτηση από τις λεκτικές αναπαραστάσεις, πράγμα που ισχύει κυρίως όσον αφορά τον φορμαλισμό των λεξικών αναπαραστάσεων. Το DDC διατηρεί μια απόσταση από τα άλλα ΣΟΓ όσον αφορά την ποικιλία των μοτίβων, καθώς απαριθμεί ένα μοτίβο για κάθε 3,38 εγγραφές, ενώ τα είκοσι πιο συχνά μοτίβα του μόλις και μετά βίας ξεπερνούν το 50% του συνολικού αριθμού των εγγραφών. Το ποσοστό των εγγραφών που καλύπτουν τα είκοσι πιο συχνά πρότυπα φτάνει το 85,41% στο LCSH και το 88,17% στο Eurovoc. Σύμφωνα με τους Mitchel κ.ά. (2014, σελ. 94) η σημειογραφία (δηλ. ο αριθμός DDC), η λεκτική αναπαράσταση της σημειογραφίας και το URI μιας κλάσης DDC μπορούν να χρησιμοποιούνται εναλλάξ, εφόσον αποτελούν νοήματα του ίδιου θέματος. Επομένως, μια πιο λεπτομερής θεώρηση της γλωσσολογικής ανάλυσης των λεξικών αναπαραστάσεων των θεμάτων μπορεί να παρέχει περισσότερες πληροφορίες για το ίδιο το θέμα.

Ο Πίνακας 1 δείχνει ότι το DDC έχει το χαμηλότερο ποσοστό μεταξύ των τριών παραδοσιακών ΣΟΓ των ΜτΛ που αντιστοιχούν σε οντότητες και το υψηλότερο ποσοστό που αντιστοιχούν σε σχέσεις- σχεδόν το 16% των λέξεών του υποδηλώνουν σχέσεις. Αυτή η παρατήρηση, σε συνδυασμό με το γεγονός ότι το DDC έρχεται πρώτο, μεταξύ των υπό μελέτη ΣΟΓ, όσον αφορά τον αριθμό των οντοτήτων ανά εγγραφή (1,98 ανά εγγραφή), δείχνει ότι υπάρχουν πιο σύνθετες έννοιες στο DDC από ό,τι σε άλλα ΣΟΓ. Οι έννοιες αυτές θα μπορούσαν να είναι μέρη μιας γενικής κατηγορίας (για παράδειγμα 026 «Βιβλιοθήκες, αρχεία, κέντρα πληροφόρησης αφιερωμένα σε συγκεκριμένους επιστημονικούς κλάδους και θέματα») ή πιο σύνθετες σχέσεις μεταξύ εννοιών (όπως 302.5 «Σχέση του ατόμου με την κοινωνία»)- ή μια διασταύρωση Boole (για παράδειγμα «401.4 Γλώσσα και επικοινωνία»). Συγκεκριμένα, η χρήση του «και» στο DDC είναι παρόμοια με τη χρήση του «και» στο LCSH, υπενθυμίζοντας τη δυσκολία που εντοπίστηκε προηγουμένως στη διάκριση μεταξύ των περιπτώσεων στις οποίες χρησιμοποιείται ως ένωση και εκείνων στις οποίες χρησιμοποιείται ως τομή.

Ο Πίνακας 2 δείχνει ότι το τέταρτο πιο συχνό μοτίβο περιέχει ψηφία που αντιστοιχούν κυρίως σε χρονολογικές πληροφορίες. Το DDC χρησιμοποιεί συχνά μεμονωμένα έτη ή άλλους χρονικούς προσδιορισμούς, όπως, για παράδειγμα στον αριθμό 943.08 την τιμή «1866-» που υποδηλώνει την Ιστορία της Γερμανίας από το 1866 και μετά.

#### 5.5.3.4 Οντολογίες: CIDOC-CRM

Είναι προφανές ότι «το πλαίσιο ανάπτυξης και εφαρμογής που προσφέρουν τα διασυνδεδεμένα δεδομένα βοήθησε να μετατραπεί η αντίληψη των οντολογιών από πολύπλοκα και διαστρωματωμένα συστήματα γνώσης που βασίζονται στη φιλοσοφική θεωρία σε πιο ρεαλιστικά εργαλεία που χρησιμοποιούνται για να συρράψουν τα

κομμάτια των πληροφοριών ώστε να δημιουργήσουν το ‘ύφασμα’ του ιστού των διασυνδεδεμένων δεδομένων» (Pattuelli κ.ά., 2015, σελ. 266).

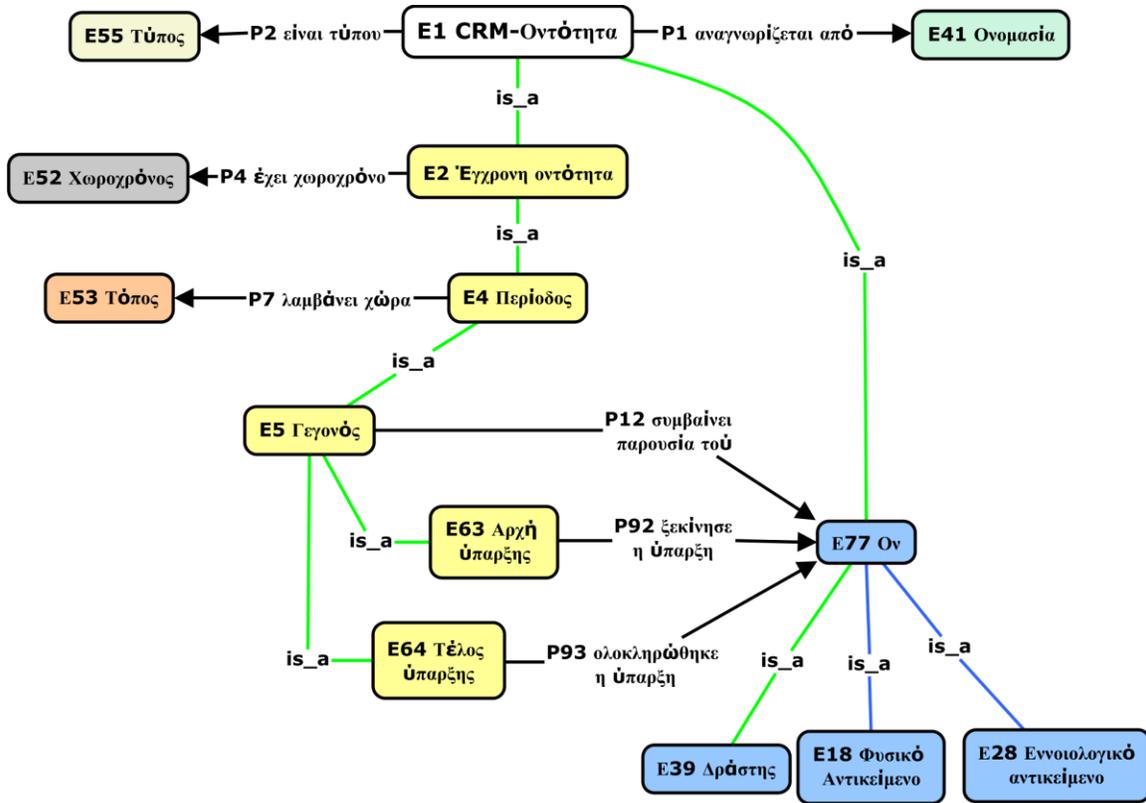
Το CIDOC-CRM είναι μια οντολογία η οποία απευθύνεται στους φορείς πολιτισμικής κληρονομίας (Μουσεία, Βιβλιοθήκες, Αρχεία και Γκαλερί)<sup>39</sup> και έχει αναπτυχθεί από την επιτροπή τεκμηρίωσης του International Council of Museums (ICOM). Σκοπός της είναι να παρέχει έναν τυποποιημένο τρόπο για την ανταλλαγή πολιτισμικής πληροφορίας μεταξύ διαφορετικών συστημάτων τεκμηρίωσης (Bekiarí κ.ά., 2021; Doerr, 2003), ο οποίος θα συμβάλει στην ολοκλήρωση (integration) δεδομένων που σχετίζονται με την πολιτισμική κληρονομιά και αποτελεί διεθνές πρότυπο (ISO, 2023).

Η συμπερίληψή του CIDOC-CRM στην παρούσα φάση της μελέτης δεν βασίζεται στην λειτουργικότητά του ως ΣΟΓ, αλλά για να χρησιμεύσει ως κριτήριο για τη σύγκριση των ΣΟΓ με τις οντολογίες. Το CIDOC-CRM, ως τυπική οντολογία, διαιρείται σε δύο βασικά μέρη: κλάσεις και ιδιότητες, όπου μια κλάση είναι μια «κατηγορία στοιχείων που μοιράζονται μία ή περισσότερες ιδιότητες» (ISO, 2006, σελ. 2) και μια ιδιότητα είναι ένα «χαρακτηριστικό που χρησιμεύει για τον καθορισμό μιας σχέσης συγκεκριμένου είδους μεταξύ δύο κλάσεων» (ISO, 2006, σελ. 4). Ο θεμελιώδης μηχανισμός των οντολογιών γενικά, καθώς και του CIDOC-CRM ειδικότερα, είναι η συσχέτιση δύο κλάσεων (εκείνης του πεδίου ορισμού -Domain και εκείνης τους πεδίου τιμών - Range) μέσω μιας ιδιότητας. Οι κλάσεις και οι ιδιότητες μελετώνται εδώ ως επιμέρους σύνολα δεδομένων προκειμένου να καταδειχθεί η παραλλαγή των ΜτΛ που χρησιμοποιείται σε κάθε μία από αυτές και, μέσω αυτής της σύγκρισης, να καταδειχθεί πόσο ανόμοια είναι η μοντελοποίηση μιας οντολογίας σε σύγκριση με τα παραδοσιακά ΣΟΓ.

Όπως φαίνεται στο Σχήμα 7 η εν λόγω οντολογία δίνει τη δυνατότητα μοντελοποίησης του τρόπου με τον οποίο τα γεγονότα συνδέονται με τους δράστες και τα εννοιολογικά ή φυσικά αντικείμενα τα οποία τα υποκείμενα παράγουν στο χωροχρόνο. Έτσι παρέχεται η δυνατότητα να μοντελοποιηθεί όχι μόνο το αποτέλεσμα μιας πράξης αλλά και η διαδικασία μέσω της οποίας παρήχθη κάτι.

---

<sup>39</sup> Το LAM που έγινε GLAM: στο χώρο της Επιστήμης της Πληροφορίας ο όρος LAM προήλθε από τα αρχικά των λέξεων Libraries, Archives and Museums και χρησιμοποιήθηκε κυρίως για να περιγράψει τη συνεργασία και τις κοινές προκλήσεις μεταξύ βιβλιοθηκών, αρχείων και μουσείων, ειδικά στον ψηφιακό μετασχηματισμό και την τεκμηρίωση των συλλογών τους. Στην συνέχεια, η αναγνώριση της συνεισφοράς τους των Γκαλερί πρόσθεσε το G και το LAM έγινε GLAM.



Σχήμα 7: σχηματική απεικόνιση των βασικών κλάσεων του CIDOC-CRM και των μεταξύ τους σχέσεων<sup>40</sup>

Όσον αφορά τη γλωσσική ανάλυση, πρέπει πρώτα να διευκρινιστεί ότι η οντολογία δεν έχει τα ίδια χαρακτηριστικά με τα εξεταζόμενα ΣΟΓ. Σύμφωνα με το SKOS, κάθε «skos:Concept» είναι ένα στιγμιότυπο (instance) του «owl:Class». Η διάκριση μεταξύ οντολογιών και ΣΟΓ περιγράφεται, μέσω του SKOS, από τους Jupp et al, οι οποίοι σημειώνουν ότι «το SKOS δεν είναι μια γλώσσα για τη μοντελοποίηση οντολογιών, το ίδιο το μοντέλο δεδομένων περιγράφεται στην πραγματικότητα ως οντολογία OWL. Στη συνέχεια, ένα συγκεκριμένο ΣΟΓ αναπαρίσταται ως μια ενσάρκωση αυτής της οντολογίας» (Jupp κ.ά., 2009, σελ. 2). Κατά συνέπεια, τα ΣΟΓ είναι λεξιλόγια τιμών και οι λέξεις που αναλύονται είναι οι τιμές. Οι ετικέτες των κλάσεων και των ιδιοτήτων της οντολογίας δεν είναι αυτού του είδους. Όπως αναφέρθηκε προηγουμένως, ο βασικός λόγος για τη γλωσσική ανάλυση της οντολογίας είχε να κάνει με τον προσδιορισμό της διαφοράς των γλωσσικών χαρακτηριστικών που υπάρχουν στις ιδιότητες, δεδομένου ότι

<sup>40</sup> Τόσο οι πράσινες, όσο και οι μπλε γραμμές απεικονίζουν «is-a» σχέσεις. Η διαφορά είναι ότι στις περιπτώσεις των μπλε γραμμών πρόκειται για έμμεσες υποκλάσεις, ενώ στις περιπτώσεις με τις πράσινες γραμμές οι υποκλάσεις είναι άμεσες.

η καινοτομία στις οντολογίες είναι οι ιδιότητες και όχι η ιεραρχία των κλάσεων η οποία άλλωστε δεν είναι και υποχρεωτική<sup>41</sup>.

Ο Πίνακας 1 δείχνει ότι, εντός των κλάσεων CIDOC-CRM, τα ΜτΛ που αντιστοιχούν σε σχέσεις σχεδόν απουσιάζουν. Αντίθετα, η πλειονότητα των λέξεων που κατηγοριοποιούνται ως σχέσεις (66,73%) έχουν προκύψει από το τμήμα ιδιοτήτων του CIDOC. Τα στοιχεία αυτής της μελέτης δείχνουν ότι αυτή είναι η μόνη περίπτωση όπου τα ρήματα παίζουν κεντρικό ρόλο- πρακτικά κάθε ιδιότητα περιέχει ένα ρήμα. Στα παραδοσιακά ΣΟΓ τα ρήματα σχεδόν απουσιάζουν πλήρως, ενώ στις οντολογίες, υπάρχουν περισσότερες ιδιότητες (δηλαδή σχέσεις) από ότι κλάσεις, διότι όπως δείχνει ο Πίνακας 1 υπάρχουν 223 ιδιότητες έναντι 81 κλάσεων.

#### 5.5.4 Κριτική ανάλυση αποτελεσμάτων

Για τη θεωρία της οργάνωσης της γνώσης, η κατηγοριοποίηση των σχέσεων σε παραδειγματικές και συνταγματικές είναι αρκετά δημοφιλής. Όπως έχει ήδη αναφερθεί, η τελευταία αποτελεί μια ανοικτή κατηγορία που δεν επιτρέπει τον ρητό ορισμό σχέσεων μεταξύ εννοιών. Όμως όταν πρόκειται για τη μοντελοποίηση οντολογιών, η προαναφερθείσα κατηγοριοποίηση επανεξετάζεται, καθώς όλες οι σχέσεις ορίζονται ρητά. Στην υπολογιστική επεξεργασία, όπως γίνεται αντιληπτή και υλοποιείται στο πλαίσιο του σημασιολογικού ιστού, οι έννοιες προσδιορίζονται από τις ιδιότητές τους και όχι από τον ορισμό των λημμάτων τους. Αντιθέτως, οι σημειώσεις σκοπού (score notes) είναι σημαντικές για τη βελτίωση της κατανόησης των ανθρώπων που δημιουργούν ή χρησιμοποιούν το εκάστοτε σύστημα ευρετηρίασης (Nowroosi κ.ά., 2018). Στην ουσία, οι ιδιότητες της RDF τις περισσότερες φορές είναι σχέσεις. Επομένως, στο πλαίσιο του σημασιολογικού ιστού, η απόφαση είναι ειλημμένη: η ονομασία δεν έχει καμία σχέση με τη συλλογιστική. Εκτός από τις περιπτώσεις συγκεκριμένων ειδών τιμών, όπως ημερομηνίες ή άλλη μονάδα μέτρησης που αφορούν είδη δεδομένων τα οποία αντιστοιχούν στις ιδιότητες τύπου Datatype της OWL. Το SKOS URI είναι ένα όνομα (nomen) το οποίο, σε αντίθεση με τα τυποποιημένα ονόματα που χρησιμοποιεί η ευρετηρίαση, δεν προορίζεται για διαχείριση από τον άνθρωπο και ίσως είναι το μόνο nomen που ταυτοποιεί τις έννοιες με σαφήνεια.

Ο βασικός μηχανισμός των οντολογιών «Κλάση Πεδίου Ορισμού-> ιδιότητα -> Κλάση Πεδίου Τιμών» είναι πολύ κοντά στο βασικό συντακτικό της φυσική γλώσσας «Υποκείμενο-> Ρήμα ->Αντικείμενο». Όπως το θέτει ο Hoerpe, το να ξεκινήσουμε να μελετάμε την αναπαράσταση ως ρήμα (δραστηριότητα) και όχι μόνο ως ουσιαστικό (το τελικό προϊόν της δραστηριότητας), υπήρξε μια από τις πιο γόνιμες κινήσεις στις κοινωνικές μελέτες της επιστήμης και της τεχνολογίας (Hoerpe, 2015, σελ. 1078). Διότι η προσέγγιση αυτή δημιουργεί το χώρο για μια συνεπή εννοιολογική δομή που αποκαλύπτει την αλληλεπίδραση μεταξύ των στοιχείων (Campos & Gomes, 2017, σελ.

---

<sup>41</sup> Προφανώς για μια οντολογία το σύνθημα είναι να έχει στον πυρήνα της μια ταξινόμια, άρα να υπάρχει μια ιεραρχία των κλάσεων. Όμως με εξαίρεση το ότι όλα είναι υποκλάσεις του «owl:Thing» δεν είναι υποχρεωτικό μια οντολογία να έχει ιεραρχία κλάσεων.

179). Η εκφραστικότητα των οντολογιών ενισχύεται λόγω του γεγονότος ότι επιτρέπουν τη ρητή δήλωση περιορισμών στις σχέσεις. Ως εκ τούτου, η εκφραστικότητά τους θα μπορούσε να παραλληλιστεί με τη σύνταξη της φυσικής γλώσσας στην βασική της μορφή. Από την άλλη πλευρά, οι οντολογίες είναι πιο ακριβείς από τη φυσική γλώσσα (που σημαίνει ότι είναι απαλλαγμένες από την εγγενή ασάφεια της φυσικής γλώσσας) και, από τη φύση τους, είναι επεξεργάσιμες από υπολογιστές. Επομένως, σε μια υποθετική πυραμίδα εκφραστικότητας, οι οντολογίες θα τοποθετούνταν στην κορυφή, δεδομένου ότι μπορούν να είναι πιο εκφραστικές από τα συμβατικά ΣΟΓ.

Μεταξύ των παραδοσιακών ΣΟΓ, τα πιο κοντινά στις οντολογίες είναι οι θησαυροί, επειδή έχουν έναν αρκετά συνεκτικό ορισμό του τι είναι μια έννοια και αποδίδουν συγκεκριμένες σχέσεις μεταξύ των εννοιών. Αλλά οι σχέσεις των θησαυρών συχνά περιορίζονται σε γενικές συσχετιστικές (η πληροφοριακή αξία των οποίων είναι περιορισμένη) και σε ιεραρχικές. Ωστόσο, παρόλο που μια ιεραρχία παρέχει πληροφορίες για ορισμένα πράγματα και είναι εξαιρετικά σημαντική για τη νόηση (Murphy, 2002, Κεφάλαιο 7), από μόνη της δεν μπορεί να οδηγήσει σε επαρκή συμπερασμό. Όπως αναφέρθηκε ήδη παραπάνω, οι αδιαίρετες (atomic) έννοιες σε συνδυασμό με τον περιορισμένο αριθμό σχέσεων έχουν ως αποτέλεσμα οι θησαυροί να έχουν σχετικά περιορισμένη εκφραστικότητα. Από την άλλη πλευρά, το υψηλό επίπεδο φορμαλισμού τους επιτρέπει να είναι επεξεργάσιμοι από υπολογιστές, αν και αυτή η δυνατότητα επεξεργασίας μπορεί να είναι γόνιμη σε ένα απλοϊκό επίπεδο εξαγωγής συμπερασμάτων, αλλά όχι για ουσιαστικό συλλογισμό. Έτσι, οι θησαυροί δεν πρέπει να θεωρούνται οντολογίες, διότι, όπως σημειώνουν οι Kless κ.ά., οι θησαυροί και οι οντολογίες πρέπει να αντιμετωπίζονται ως δύο ορθογώνια είδη μοντέλων με επιφανειακά παρόμοιες δομές (Kless κ.ά., 2015, σελ. 1348).

Από την άλλη πλευρά, οι LCSH είναι ιδιαίτερες εκφραστικές, καθώς υπάρχουν πολλά είδη σχέσεων μεταξύ και εντός των πεδίων και των υποπεδίων. Όπως έδειξαν τα δεδομένα, όταν οι παραδειγματικές σχέσεις μεταξύ των λειτουργικών τμημάτων δεν είναι επαρκείς, η φυσική γλώσσα αναδύεται για να δώσει λύση. Με αυτόν τον τρόπο δημιουργούνται πολλές σύνθετες έννοιες, όπως απεικονίζεται στο Γράφημα 5. Ένα αντιπροσωπευτικό παράδειγμα αυτής της προσέγγισης είναι η επικεφαλίδα «Μετατροπή του νερού σε κρασί στο γάμο της Κανά (θαύμα) στην τέχνη» (<http://id.loc.gov/authorities/subjects/sh97001686>). Όμως οι LCSH επικεντρώνονται στον φορμαλισμό των τιμών (συμβολοσειρών) και όχι στη σημασιολογία ή στον δομικό φορμαλισμό. Η ευελιξία της δημιουργίας σχέσεων με βάση τη φυσική γλώσσα δεν είναι επεξεργάσιμη από μηχανήματα- τουλάχιστον δεν φέρει το είδος της επεξεργάσιμης δυνατότητας που απαιτείται στο πλαίσιο του σημασιολογικού ιστού. Δεδομένου ότι οι επικεφαλίδες ενσωματώνουν μεγάλο αριθμό εννοιών τόσο μέσω της συσχέτισης των υποπεδίων όσο και μέσω των ενδο-ορολογικών σχέσεων, η εξαγωγή συμπερασμάτων θα μπορούσε να οδηγήσει σε παραπλανητικές συσχετίσεις. Ένα αντιπροσωπευτικό παράδειγμα αυτού του προβληματικού συμπερασμού έχει απεικονιστεί από τον Spero, ο οποίος, ακολουθώντας τις σχέσεις των LCSH, δείχνει τη σύνδεση μεταξύ της επικεφαλίδας «Doorbells» (κουδούνια) και της επικεφαλίδας «Mammals» (θηλαστικά) δίνοντας τον τίτλο στην εργασία του «LCSH is to Thesaurus as Doorbell is to Mammal:

Visualizing Structural Problems in the Library of Congress Subject Headings» (Spero, 2008). Το προβληματικό συμπέρασμα προκύπτει επειδή οι έννοιες δεν ορίζονται με σαφήνεια από τις ιδιότητές τους, δηλαδή τις σχέσεις RDF/SKOS, διότι ο φορμαλισμός των σχέσεων δεν αποδίδεται με απόλυτα συνεπή τρόπο.

Τέλος τα ταξινομικά συστήματα αναπαριστούν τη γνώση με εντελώς διαφορετικό τρόπο, κυρίως λόγω του γεγονότος ότι δεν είναι εννοιοκεντρικά αλλά επιστημονικοκεντρικά. Όσον αφορά τη σημειογραφία του DDC, το σκεπτικό πίσω από αυτήν είναι ότι πρόκειται για ένα συμπυκνωμένο λεκτικό. Για παράδειγμα, το «190» σημαίνει «Σύγχρονη δυτική και λοιπές μη ανατολικές φιλοσοφίες». Όπως φαίνεται στο Γράφημα 5, σχεδόν οι μισοί από τους όρους στο DDC είναι σύνθετοι (δηλαδή εννοιολογικά διαιρετοί) και υλοποιούνται μέσω διαφόρων συντακτικών μοτίβων. Η πιο δημοφιλής κατηγορία είναι η παραταξιακή, η οποία υπερβαίνει το ένα τρίτο των όρων. Μια συνήθης πρακτική για τη διατύπωση αυτής της κατηγορίας είναι η χρήση λέξεων όπως «άλλα» ή «κ.λ.π.» που σηματοδοτεί ανοικτές κατηγορίες. Τέτοιες δομές είναι κατανοητές από τον ανθρώπινο νου, ενισχύοντας έτσι την εκφραστικότητα. Ωστόσο δεν αποτελούν πληροφορίες επεξεργάσιμες από μηχανήματα.

Κατά συνέπεια, η επιλογή ενός σχήματος οργάνωσης ή αναπαράστασης της γνώσης επηρεάζει άμεσα τις δυνατότητες εκφραστικότητας καθώς και τη δυνατότητα επεξεργασίας από τον υπολογιστή. Για να αναπτύξουμε περαιτέρω αυτή τη δήλωση, ας εξετάσουμε τον όρο «Βία των μαθητών προς τους καθηγητές», ο οποίος διαθέτει τρία δομικά στοιχεία, δηλαδή «μαθητές», «καθηγητές» και «βία». Σε ένα υστεροσυνδυασμένο σύστημα, όπου δεν υπάρχει ένδειξη της κατεύθυνσης των σχέσεων, δεν θα μπορούσε να υποδείξει ποιος ήταν το υποκείμενο της βίας, δηλαδή αν αυτή ασκήθηκε από τους καθηγητές προς τους μαθητές ή αντίστροφα. Από την άλλη πλευρά, τα προσυνδυασμένα συστήματα, όπως οι θεματικές επικεφαλίδες ή το DDC, θα αναπαριστούσαν επαρκώς μια τέτοια δήλωση, πιθανότατα με τη χρήση μιας ad hoc λύσης που θα στερούσαν δομικού φορμαλισμού. Μια οντολογία θα μπορούσε να το αναπαραστήσει με ακρίβεια με δύο κλάσεις και μια σχέση (ιδιότητα), προσδιορίζοντας την κατεύθυνση της σχέσης. Και αυτό είναι ιδιαίτερα κρίσιμο, διότι «οι αυστηροί ορισμοί για τον χαρακτηρισμό των τυπικών σχέσεων θα αποτελέσουν ένα σημαντικό βήμα προς την κατεύθυνση της δυνατότητας των επιστημόνων της πληροφορίας να επιτύχουν διαλειτουργικότητα μεταξύ οντολογιών για την υποστήριξη της αυτοματοποιημένης συλλογιστικής σε δεδομένα που προέρχονται από διαφορετικούς τομείς» (Schwarz & Smith, 2008, σελ. 222).

## 5.6 Συμπέρασμα κεφαλαίου

Όπως έδειξε η ανάλυση που προηγήθηκε, η φυσική γλώσσα αποτελεί τον κυρίαρχο μηχανισμό για την αποτύπωση της ανθρωπολογικής γνώσης. Αφού αναλύθηκε, ακροθιγώς, η φυσική γλώσσα ως φορέας του νοήματος, στη συνέχεια παρουσιάστηκαν φορμαλιστικές γλώσσες οι οποίες τοποθετούν στο επίκεντρο τις έννοιες και όχι τις λέξεις, εστιάζοντας σε εκείνες τις φορμαλιστικές 'γλώσσες' που βασίζονται στις τεχνολογίες του σημασιολογικού ιστού.

Στο πλαίσιο της σύγκρισης διαφόρων Συστημάτων Οργάνωσης της Γνώσης (ΣΟΓ), ώστε να κριθεί η καταλληλότητά τους για πιθανή εφαρμογή στην ανθρωπολογία, εξετάστηκε αντιπροσωπευτικό δείγμα από ΣΟΓ. Τα προς εξέταση ΣΟΓ προέρχονταν από τέσσερις βασικές κατηγορίες: θησαυροί, θεματικές επικεφαλίδες, ταξινομικά συστήματα και οντολογίες. Πέραν της θεωρητικής ανάλυσης για την κάθε κατηγορία, επιλέχθηκε ένα αντιπροσωπευτικό δείγμα από κάθε κατηγορία και αναλύθηκαν γλωσσικά, με χρήση τεχνικών επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP) οι όροι που αυτό περιείχε.

Η γλωσσική ανάλυση, μέσω NLP, αποκάλυψε ότι στο πλαίσιο των παραδοσιακών ΣΟΓ δεν υπάρχει σαφής ή συνεπής ορισμός του τι συνιστά μια έννοια. Η μορφοσυντακτική ανάλυση έδειξε ότι δεν υπάρχει σαφής διάκριση μεταξύ ονομαστικών οντοτήτων και εννοιών, δηλαδή τα κύρια ουσιαστικά εμφανίζονται μαζί με τα κοινά ουσιαστικά. Αυτό σημαίνει ότι δεν υπάρχει διαφοροποίηση όπως αυτή που ορίζεται μεταξύ των κλάσεων και των στιγμιотύπων (instances) των οντολογιών. Αυτό το είδος ανάλυσης έδειξε επίσης ότι τόσο στο LCSH όσο και στο DDC εντοπίζονται διάφοροι τύποι ενδοορολογικών σχέσεων, αλλά δεν υπάρχει ενσωματωμένος μηχανισμός ή μεθοδολογία για τον ρητό ορισμό τους. Μόνο η θέση στην ιεραρχία, ή μερικές φορές ένας προσδιοριστής, ορίζει το πλαίσιο στο οποίο η έννοια αποκτά το νόημά της. Ακόμη και έτσι, ο τρόπος με τον οποίο είναι δομημένη η ιεραρχία μπορεί να οδηγήσει σε προβληματικό συμπέρασμα, καθώς δεν βασίζεται πάντα σε σχέσεις τύπου «is-a».

Ειδικά στην περίπτωση των LCSH, δίνεται έμφαση στη δομή των όρων ως συμβολοσειρών που τυποποιούν την τιμή της επικεφαλίδας, προκειμένου να δημιουργηθεί ένα κατάλληλο αναγνωριστικό που να επιτρέπει την αποτελεσματικότερη ταύτιση των συμβολοσειρών (string matching). Οι ενδοορολογικές σχέσεις έρχονται δεύτερες κατά σειρά σπουδαιότητας. Ο μηχανισμός κατασκευής των LCSH διαχειρίζεται τη σημασιολογία πολύ χαλαρά όσον αφορά τις φορμαλιστικές αναπαραστάσεις, ενώ χρησιμοποιεί εκτενώς τη φυσική γλώσσα για την ενίσχυση της ευελιξίας και της εκφραστικότητάς του. Ωστόσο, αυτή η εκφραστικότητα δεν μπορεί να μεταφερθεί στον Σημασιολογικό Ιστό. Επιπλέον, οι LCSH φαίνεται να θεωρούν δεδομένη μια σημαντική ποσότητα λανθάνουσας πληροφορίας, αλλά είναι γνωστό στους ανθρώπους που ασχολούνται με τη μοντελοποίηση δεδομένων ότι «αυτό που φαίνεται το πιο απλό [...] είναι διαβολικά δύσκολο να μοντελοποιηθεί» (Fauconnier & Turner, 2003, σελ. 7). Η συχνή χρήση της απαρίθμησης στο DDC (όπως απεικονίζεται μέσω της προσέγγισης NLP) καθιστά το σχήμα όχι τόσο εννοιοκεντρικό, ενώ οι ιεραρχίες που παράγονται είναι επίσης αδύναμες όσον αφορά τη σημασιολογία. Αντίθετα, η σημασιολογία των σχέσεων στους θησαυρούς είναι πιο συνεπής. Επιπλέον, λόγω της ρητής σημασιολογίας των εννοιών, θα μπορούσαν να υλοποιηθούν περισσότερες σχέσεις -πέραν των ιεραρχικών- παρόμοιες με αυτές που χρησιμοποιούνται στις οντολογίες. Για παράδειγμα, μια δήλωση όπως «Το ελάφι τρώει γρασίδι» μπορεί να εκφραστεί σε έναν θησαυρό χρησιμοποιώντας δύο έννοιες και ένα κατηγορημα με προκαθορισμένη κατεύθυνση. Αντίθετα, στο LCSH και στο DDC τέτοιες δηλώσεις μπορούν να εκφραστούν, και εκφράζονται, ως σύνθετες (δηλαδή εννοιολογικά διαιρετές) έννοιες.

Οι σύγχρονες τάσεις εστιάζουν στη μοντελοποίηση των σχέσεων, δεδομένου ότι αυτό αποτελεί προϋπόθεση για τη διαδικασία συλλογισμού που οδηγεί στην απόδοση νοήματος στις έννοιες. Στα συμβατικά ΣΟΓ η μοντελοποίηση των σχέσεων υπολείπεται, ενώ η φυσική γλώσσα χρησιμοποιείται για τον ορισμό των ενδοορολογικών σχέσεων που επιτρέπουν τη δημιουργία σύνθετων εννοιών. Όμως οι σχέσεις είναι πολύ δύσκολο να μοντελοποιηθούν όσο υπάρχει έλλειψη αδιαίρετων εννοιών (atomic concepts). Όσο πιο σύνθετη (εννοιολογικά διαιρετή) είναι μια έννοια τόσο πιο δύσκολο γίνεται να παραχθούν ακριβείς σχέσεις μεταξύ δύο εννοιών, και μερικές φορές είναι παντελώς αδύνατο. Αυτή η έλλειψη ακρίβειας στην LCSH μπορεί να οδηγήσει στο να συσχετίζονται τα κουδούνια των θυρών με τα θηλαστικά, όπως αναφέρθηκε στην υποενότητα 5.5.4. Όπως τονίζουν οι Wilmont et al, «Μια σχέση περιγράφει τη συμπεριφορά μεταξύ δύο ή περισσότερων εννοιών. [...] Όπως συμβαίνει με τις έννοιες, έτσι και μια σχέση δεν αποκτά νόημα χωρίς ενδελεχή, συγκεκριμένη κατανόηση των εννοιών που συνδέει. Ως εκ τούτου, οι έννοιες και οι σχέσεις εξαρτώνται η μία από την άλλη για να αποκτήσουν νόημα» (Wilmont κ.ά., 2013, σελ. 77). Με αυτόν τον τρόπο δημιουργείται ένα νοηματικό δίκτυο όπου το επίκεντρο της ερμηνείας των εννοιών τοποθετείται στις σχέσεις τους. Μέσα σε ένα σημασιολογικό δίκτυο, με κόμβους και ακμές, είναι απολύτως σαφές ότι κάθε κόμβος συνδέεται με έναν άλλο κόμβο μέσω μιας ακμής, η οποία αντιπροσωπεύει μια σχέση, δηλαδή μια ιδιότητα, η οποία (όπως προέκυψε από τη μελέτη μας στο CIDOC-CRM) εκφράζεται πάντα με ένα ρήμα.

Πρέπει να διανυθεί μεγάλη απόσταση ακόμη για να αποτελέσουν τα ΣΟΓ βάσεις γνώσης και αποτελεσματικά σημασιολογικά εργαλεία στο πλαίσιο του σημασιολογικού ιστού (Zeng & Mayr, 2019, σελ. 20). Αυτό που τονίζεται στην παρούσα μελέτη είναι ότι, προκειμένου να επιτευχθεί αυτό, το ζήτημα δεν περιορίζεται στη μετατροπή των δεδομένων από μια μορφή σε μια νεότερη. Διότι η διαδικασία αυτή, από μόνη της, μπορεί να αποτελέσει τροχοπέδη όσον αφορά την αξιοποίηση των εκφραστικών δυνατοτήτων της νέας μορφής. Αυτή η επιλογή είναι σαν να χρησιμοποιούμε ένα έξυπνο τηλέφωνο μόνο για να κάνουμε τηλεφωνήματα. Το κύριο ζήτημα είναι να υιοθετήσουμε συνολικά μια διαφορετική λογική. «Ο επανασχεδιασμός των ΣΟΓ με οντολογικές αρχές απαιτεί γενικά έναν νέο τρόπο σκέψης» (Kless κ.ά., 2011, σελ. 67), άρα νέα αντίληψη. Συνεπώς, το ζητούμενο δεν είναι απλώς να εκφραστεί ένα συμβατικό ΣΟΓ σε νέες γλώσσες (όπως η OWL), αλλά να αξιοποιηθούν οι νέες δυνατότητες που προκύπτουν μέσω της νέας θεωρητικής και τεχνολογικής υποδομής, προκειμένου να επιτευχθεί ακριβέστερη και πιο εξελιγμένη έκφραση του νοήματός του. Τελικά, παραφράζοντας την περίφημη πρόταση του Wittgenstein «[τ]α όρια της γλώσσας μου σημαίνουν τα όρια του κόσμου μου» (Wittgenstein, 1978), θα μπορούσαμε να πούμε ότι τα όρια των τεχνητών γλωσσών μας σημαίνουν τα όρια των Συστημάτων Οργάνωσης Γνώσης.

Εν κατακλείδι, η έρευνα που προηγήθηκε κατέληξε στο ότι οι οντολογικές αναπαραστάσεις δείχνουν κατάλληλες για την αναπαράσταση της ανθρωπολογικής γνώσης και οι θησαυροί κατάλληλοι για ελεγχόμενα λεξιλόγια τιμών. Διότι οι οντολογίες, παρότι προορίζονται για χρήση κυρίως στο περιβάλλον της τεχνητής νοημοσύνης, χειρίζονται τις έννοιες και τις σχέσεις μεταξύ των εννοιών ακολουθώντας παρόμοια λογική με εκείνη που χρησιμοποιεί ο ανθρώπινος με τη διαφορά ότι, για τη μηχανική

συλλογιστική, τα στιγμιότυπα είναι (ορίζονται από) αυτό που δηλώνουν οι ιδιότητές τους. Άρα το “νόημα” μιας έννοιας είναι αυτό που προκύπτει από τη θέση της εντός ενός δικτύου συσχετίσεων. Με άλλα λόγια το νόημα μιας έννοιας είναι αυτό που υποδηλώνουν οι ιδιότητές της και όχι αυτό που υποδεικνύουν οι ορισμοί ή το όνομά της.



## 6 Οντολογική αναπαράσταση Εθνογραφίας και Κοινωνικής Ανθρωπολογίας

Στο προηγούμενο κεφάλαιο, και ειδικά στην ενότητα 5.3, παρουσιάστηκαν οι βασικές αρχές πάνω στις οποίες στηρίζεται ο σημασιολογικός ιστός. Στο παρόν κεφάλαιο, δανειζόμενος τα λόγια του Leach, θα προσπαθήσω «να αρχίσω να σας πείθω ότι τα τόσο αφηρημένα και φορμαλιστικά θέματα που σας ανέπτυξα ως τώρα έχουν κάποια πρακτική εφαρμογή στην ανάλυση του εθνογραφικού υλικού το οποίο πραγματεύονται τακτικά οι κοινωνικοί ανθρωπολόγοι» (Leach, 1993, σελ. 63). Πριν από όλα όμως θα πρέπει τονιστεί ότι η αναπαράσταση της ανθρωπολογικής γνώσης για το σημασιολογικό ιστό προφανώς και δεν είναι εργασία ενός ανθρώπου ή έστω μια μικρής ομάδας ανθρώπων. Όμως σε αυτό κεφάλαιο, θα προταθεί ένα μοντέλο πάνω στο οποίο θα μπορούσε να βασιστεί η αναπαράσταση της ανθρωπολογικής γνώσης με όρους σημασιολογικού ιστού και θα δοθούν μερικά ενδεικτικά παραδείγματα κωδικοποίησης.

Η αναπαράσταση της γνώσης ανήκει στον πυρήνα των ενδιαφερόντων της ανθρωπολογίας (M. D. Fischer, 1994, σελ. 3). Και ακριβώς επειδή το υλικό της ανθρωπολογίας είναι εγγενώς συγκεχυμένο (Fuentes & Wiessner, 2016, σελ. 53) οι φορμαλιστικές αναπαραστάσεις επιτρέπουν να υποσκελιστούν αρκετές από τις δυσκολίες που απαντώνται στην αποτύπωση σαφών περιγραφών της γνώσης, όπως: η επίλυση, ή έστω η αναγνώριση, της αμφισημίας, και ο έλεγχος της περιγραφής ως προς την πληρότητα ή την συνέπεια και την επαλήθευση της περιγραφής (M. D. Fischer & Finkelstein, 1991, σελ. 120).

Όπως έχει ήδη περιγραφεί στην ενότητα «5.2 Φυσική γλώσσα, γραφή και ανθρωπολογική γνώση», η φυσική γλώσσα και κυρίως η κειμενοποίησή (textualization) της, έχει αποτελέσει το βασικό φορέα τόσο για την παραγωγή όσο και για την αποτύπωση της ανθρωπολογικής έρευνας. Όμως το βασικό μέσο για την αναπαράσταση της ανθρωπολογικής γνώσης στο σημασιολογικό ιστό αποτελούν οι οντολογίες. Οι οντολογίες όμως δεν είναι μια προ-αποφασισμένη, “αντικειμενική” αποτύπωση αλλά αποτελούν ένα σύνθετο κοινωνιοτεχνικό δημιούργημα.

## 6.1 Η οντολογία ως πολιτισμική οπτική

Για το πεδίο της αναπαράστασης γνώσης οι οντολογικές κατηγορίες είναι αναλυτικές κατηγορίες. Δεν χρειάζεται αυτό που περιγράφεται να “υπάρχει” με την διάσταση που αποδίδει η φιλοσοφία στην έννοια των οντολογικών κατηγοριών. Δηλαδή είναι εφικτή η ύπαρξη μιας κατηγορίας «Θεοί» και η ένταξη σε αυτήν του Δία και του Βισνού, χωρίς αυτό να σημαίνει ότι αποδείχτηκε η ύπαρξή τους.

Οι Anticoli και Torrano σχολιάζουν πως μια οντολογία δεν είναι μόνο ένα κοινωνιοτεχνικό δημιούργημα αλλά και ένα πολιτισμικό δημιούργημα (Anticoli & Torrano, 2011b, σελ. 1). Ενώ οι ίδιοι, σε άλλη τους δημοσίευση, σημειώνουν ότι μια οντολογία είναι σαν ένα ζευγάρι γυαλιά: διαμεσολαβεί για αυτό που μπορεί να γίνει αντιληπτό παρέχοντας ένα σύνολο από τύπους εννοιών και σχέσεων μέσω των οποίων ερμηνεύεται η πραγματικότητα. Την ίδια στιγμή, η οντολογία, καλεί για την επίλυση συγκεκριμένων ερωτημάτων με σκοπό τη μοντελοποίηση της πραγματικότητας. Άρα η οντολογία επηρεάζει τόσο την ερμηνεία όσο και την πράξη. Οι οντολογίες και η πράξη είναι στενά αλληλεξαρτώμενες: μια αλλαγή στην πράξη παράγει τροποποίηση της οντολογίας και αντιστρόφως (Anticoli & Torrano, 2011a, σελ. 10).

Ο Seaver, ορίζοντας τους αλγόριθμους γενικά ως όλων των ειδών τους κανόνες που διέπουν τη λειτουργία των υπολογιστών, θέτει το ζήτημα ακόμη πιο αφαιρετικά και σημειώνει «προτείνω να σκεφτόμαστε τους αλγόριθμους όχι ως κάτι ‘εντός’ του πολιτισμού [...] αλλά ‘ως’ πολιτισμό» (Seaver, 2017, σελ. 1) και η Christin συμπληρώνει ότι «οι αλγόριθμοι δεν υπάρχουν ποτέ σε κοινωνικό κενό»<sup>42</sup> (Christin, 2020b). Άλλωστε υπάρχουν μελέτες που αποδεικνύουν ότι οι αλγόριθμοι αναπαράγουν πατριαρχικά στερεότυπα, ενώ παράλληλα πραγματοποιούν φυλετικές διακρίσεις (Parakyriakopoulos & Mboya, 2022) και μάλιστα αυτό μπορεί να επηρεάζεται από τη σύνθεση της ομάδας των προγραμματιστών οι οποίοι τους ανέπτυξαν (Μήτρου, 2023, σελ. 117).

Οι οντολογίες ορίζουν με σαφήνεια κατηγορίες, σχέσεις κτλ αλλά αυτό γίνεται για να αποτυπωθεί με σαφήνεια τι εννοεί ο δημιουργός τους ενώ δεν αποκλείουν την αποτύπωση της ατομικής θεώρησης για τα δεδομένα που αναπαρίστανται, εφόσον οι επιλογές των οριζόμενων (κατηγοριών, σχέσεων κτλ) αποδίδονται με βάση την ερμηνεία του δημιουργού. Συνεπώς, αυτό που καθίσταται δυνατό με τη χρήση των οντολογιών είναι η αποσαφήνιση της πρόθεσης του δημιουργού, και όχι, κατ’ ανάγκη, η αποτύπωση “αντικειμενικών δηλώσεων” για την πραγματικότητα. Άρα και η οντολογία είναι και αυτή μία πολιτισμική κατασκευή, η οποία αντανάκλα μια συγκεκριμένη επιστημολογική προσέγγιση και ως εκ τούτου ούτε είναι, ούτε αναπαριστά, μια αντικειμενική πραγματικότητα.

---

<sup>42</sup> «algorithms never exist in a social vacuum»

## 6.2 Μοντέλο αναπαράστασης με χρήση επιπέδων ολοκλήρωσης (layers of integration)

Αφού αποσαφηνίστηκε ότι κάθε οντολογία εκφράζει μια πολιτισμική οπτική θα παρουσιαστεί ένα μοντέλο το οποίο επιτρέπει τόσο την περιγραφή ενός εκάστου πολιτισμού όσο και την αναγωγή των επιμέρους πολιτισμών σε ένα γενικότερο επίπεδο. Σε κάθε περίπτωση σημείο αφετηρίας είναι η αναπαράσταση γνώσης ενώ για την υλοποίηση χρησιμοποιούνται εργαλεία του σημασιολογικού ιστού και συγκεκριμένα οι γλώσσες RDF, RDFS και OWL.

### 6.2.1 Διαφορετικές οντολογικές κατηγορίες ανά πολιτισμό

Δεν είναι ανάγκη να είναι κάποια/ος ανθρωπολόγος για να γνωρίζει ότι οι ανθρώπινες κοινωνίες δεν μοιράζονται έναν κοινό πολιτισμό και δεν ασκούν κοινές πρακτικές. Συνεπώς, οι διαφορετικές ανάγκες της περιγραφής δεν έχουν να κάνουν μόνο με την προσέγγιση του παρατηρητή (όπως περιγράφεται στην προηγούμενη υποενότητα) αλλά και με τον υπό περιγραφή και ανάλυση πολιτισμό. Πέρα όμως από τις σαφώς διακριτές οντολογικές κατηγορίες (π.χ. περιτομή, νηπιοβάπτισμα) υπάρχουν μια σειρά από κατηγορίες οι οποίες παρότι φαινομενικά φαίνονται να ταυτίζονται -ή να είναι πολύ παρόμοιες-, αν εξεταστούν προσεκτικά προκύπτει ότι οι ιδιότητές τους όχι απλά μπορεί να διαφέρουν από πολιτισμό σε πολιτισμό αλλά μπορεί να είναι και αντιθετικές μεταξύ τους. Στην ενότητα αυτή θα δοθούν μερικά ενδεικτικά παραδείγματα σχετικά με τις διαφορετικές ιδιότητες οι οποίες παρουσιάζουν φαινομενικά παρόμοιες κλάσεις.

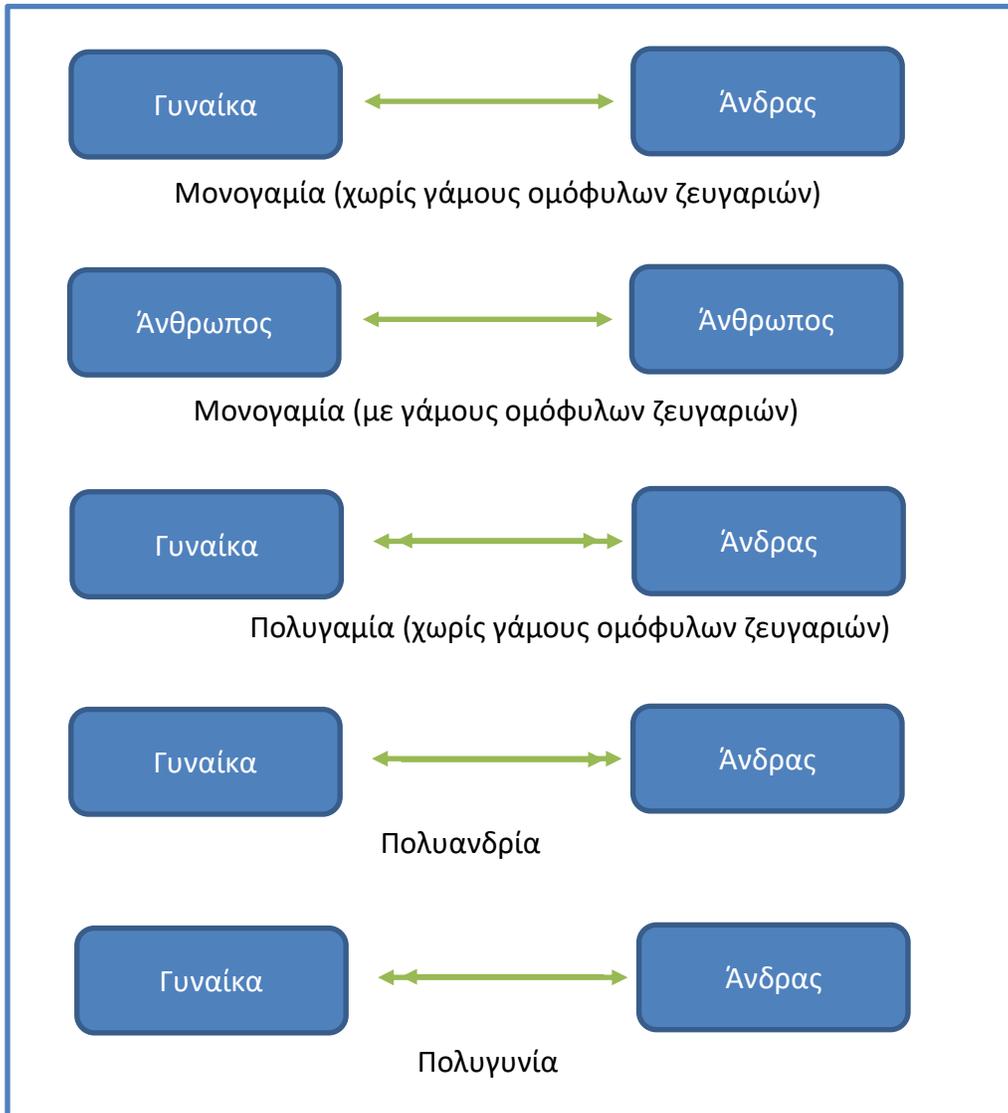
Κύρια πηγή πληροφοριών για την άντληση παραδειγμάτων αποτέλεσε το HRAF<sup>43</sup>. Προφανώς ο στόχος του κεφαλαίου -και της διατριβής εν γένει- δεν είναι η αναλυτική παρουσίαση κανενός πολιτισμού, ούτε κάποιας συγκεκριμένης πολιτισμικής πρακτικής. Άρα τα όποια κοινωνικά φαινόμενα αναφέρονται στη συνέχεια είναι αποσπασματικά και δεν στοχεύουν στην ανάλυση ή την παρουσίαση του πολιτισμού στον οποίο λαμβάνουν χώρα αλλά παρουσιάζονται εδώ ως αντιπροσωπευτικά παραδείγματα για την αναπαράσταση ανθρωπολογικής γνώσης.

Η συγγένεια θεωρείται ένας θεσμός με παγκόσμια ισχύ ο οποίος απαντάται στο σύνολο των πολιτισμών. Την ίδια στιγμή όμως ίσως αποτελεί ένα από τα τυπικότερα παραδείγματα διαφοροποίησης από μια κοινωνία σε μία άλλη, αφού οι διαφορές στους τρόπους με τους οποίους οι άνθρωποι συνδέονται μεταξύ τους μέσω της συγγένειας οριοθετούν ένα πολύ ευρύ φάσμα πρακτικών. Κατά κανόνα η συγγένεια ξεκινάει από τον γάμο. Ας πάρουμε, για παράδειγμα, τους όρους Μονογαμία (Monogamy), Πολυανδρία (Polyandry) και Πολυγυνία (Polygyny)<sup>44</sup>. Στην περίπτωση της μονογαμίας

<sup>43</sup> Η βάση δεδομένων του HRAF έχει παρουσιαστεί στην ενότητα «4.5.1 Βάσεις δεδομένων για Εθνογραφία και Ανθρωπολογία».

<sup>44</sup> Πολυγαμία (Polygamy) εφαρμόζονταν στον πολιτισμό των Μασάι. Polyandry ασκούν οι Toda (Emeneau, 1941). Πολυγυνία (Polygyny) ο Korytoff σημειώνει για τους Suku ότι περίπου το 20% των ανδρών είχαν πάνω από μία σύζυγο (Korytoff, 1964). Εντοπίστηκαν στο HRAF.

ένα άτομο δεν μπορεί να είναι παντρεμένο -ταυτόχρονα- με πάνω από ένα άτομο. Στην περίπτωση της πολυανδρίας μια γυναίκα μπορεί να έχει πολλούς συζύγους ταυτόχρονα, ενώ στην περίπτωση της πολυγυνίας ένας άντρας μπορεί να έχει πολλές συζύγους ταυτόχρονα. Στο Σχήμα 8 παρουσιάζονται οι παραπάνω περιορισμοί κάνοντας χρήση του περιορισμού πληθικότητας (cardinality constrain). Όπου υπάρχει μονό βελάκι σημαίνει ότι η σχέση είναι είτε «από ένα» (αριστερό βελάκι) είτε «προς ένα» (δεξί βελάκι). Ενώ όπου υπάρχουν διπλά βελάκια σημαίνει ότι η σχέση είναι είτε «από πολλά» (βελάκια αριστερά) είτε «προς πολλά» (βελάκια δεξιά).



Σχήμα 8: παραδείγματα ανάγκης περιορισμού πληθικότητας στις σχέσεις γάμου

Προς αποφυγή παρεξηγήσεων: οι περιορισμοί πληθικότητας δεν έχουν ως στόχο τη μέτρηση των συζύγων! Ο στόχος είναι η αναπαράσταση της δομής των κοινωνικών φαινομένων τα οποία συναρτώνται άμεσα με τα νούμερα. Διότι ο αριθμός των ατόμων

με τα οποία ένα άτομο μπορεί να έχει συζυγικές σχέσεις επηρεάζει σε μεγάλο βαθμό τόσο τις κοινωνικές δομές όσο και τη ζωή των ανθρώπων εντός αυτών των νορμών. Για παράδειγμα, ένας σύγχρονος Ιορδανός μπορεί να έχει δύο συζύγους. Αυτό όμως προϋποθέτει ότι θα έχει δύο διαφορετικές κατοικίες και η κάθε σύζυγος θα ζει στη δική της μαζί με τα δικά της παιδιά. Ενώ ο χρόνος του συζύγου θα πρέπει να επιμερίζεται αναλογικά ανάμεσα στις δύο κατοικίες. Το ζητούμενο λοιπόν της πληθικότητας δεν είναι να ποσοτικοποιηθεί ο αριθμός των ατόμων αλλά να μοντελοποιηθούν οι κοινωνικές προεκτάσεις της κάθε πρακτικής.

Πέραν όμως της ανάγκης περιορισμού της πληθικότητας, οι συγγενικές σχέσεις διέπονται και από άλλα χαρακτηριστικά. Ένα πολύ σημαντικό τους χαρακτηριστικό είναι ότι είναι αμφίδρομες και διαθέτουν αντίστροφες σχέσεις. Οι αντίστροφες σχέσεις μπορεί να είναι ταυτόσημες με τις πρώτες (π.χ. αν για την Α είναι αδερφή η Β τότε και για την Β είναι αδερφή η Α) ή και όχι (π.χ. όταν η Χ είναι μαμά για την Ψ τότε η Ψ είναι κόρη της Χ). Ακόμη ένα πολύ σημαντικό χαρακτηριστικό των συγγενικών σχέσεων είναι ότι ορίζουν ομάδες ανθρώπων οι οποίες μπορούν να συνδεθούν με συγκεκριμένες σχέσεις με άλλες ομάδες.

Αναπαριστώντας τους παραπάνω περιορισμούς στο περιβάλλον του σημασιολογικού ιστού, μέσω της OWL, υπάρχει η δυνατότητα α) να ελεγχθεί η πληθικότητα και η φορά της, β) να οριστεί πεδίο ορισμού και πεδίο τιμών (domain-range) για το ποιες κατηγορίες μπορούν να σχετιστούν με ποιες και με ποιον τρόπο γ) να οριστεί η φορά των σχέσεων καθώς και οι αντίστροφες σχέσεις. Στο Απόσπασμα κώδικα 7 δίνεται ένα παράδειγμα της σχέσης «νυμφεύεται» σε ένα περιβάλλον μόνιμης μονογαμίας (που δεν επιτρέπει πάνω από έναν γάμο (permanent monogamy)), στο οποίο δεν επιτρέπονται γάμοι για ομόφυλα ζευγάρια. Για την εν λόγω σχέση έχουν δηλωθεί τα εξής χαρακτηριστικά:

- α) έχει αντίστροφη σχέση το «παντρεύεται»
- β) έχει ευρύτερη σχέση το «έχει σύζυγο»
- γ) διαθέτει περιορισμό πληθικότητας στο ένα προς ένα
- δ) η ομάδα ορισμού της (domain) είναι οι «άνδρες»
- ε) η ομάδα τιμών (range) είναι οι «γυναίκες»

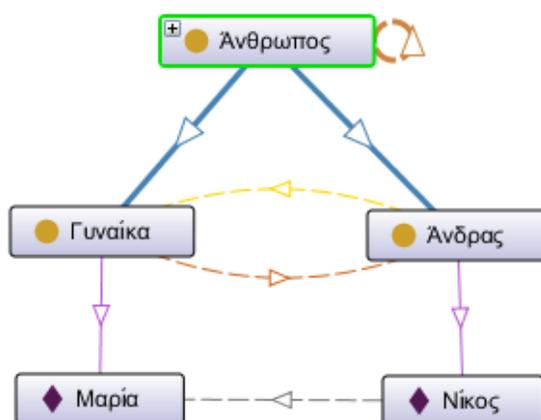
**Απόσπασμα κώδικα 7:** η σχέση «νυμφεύεται» -ως OWL ιδιότητα- σε περιβάλλον μόνιμης μονογαμίας (χωρίς γάμους ομόφυλων ζευγαριών)

```
<owl:ObjectProperty rdf:about="http://MyPhDExamples#νυμφεύεται">
<rdfs:subPropertyOf rdf:resource="http://MyPhDExamples#έχει_σύζυγο"/>
<owl:inverseOf rdf:resource="http://MyPhDExamples#παντρεύεται"/>
<rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
<rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#InverseFunctionalProperty"/>
<rdfs:domain rdf:resource="http://MyPhDExamples#Ανδρας"/>
<rdfs:range rdf:resource="http://MyPhDExamples#Γυναίκα"/>
</owl:ObjectProperty>
```

Όπως αναφέρθηκε παραπάνω η σχέση «νυμφεύεται» έχει ως ευρύτερη ιδιότητα την ιδιότητα «έχει σύζυγο». Αυτό σημαίνει πως όποιος είναι νυμφευμένος συνδέεται με αυτό το άτομο και με τη σχέση «έχει σύζυγο». Τα χαρακτηριστικά όμως της σχέσης «έχει σύζυγο» μπορεί να είναι διαφοροποιημένα αναφορικά με τις υπο-σχέσεις της. Για παράδειγμα η σχέση «έχει σύζυγο» μπορεί να είναι συμμετρική (symmetric), αν έχει πεδίο ορισμού και πεδίο τιμών τον άνθρωπο. Αν η σχέση έχει πεδίο ορισμού τον άνδρα ή την γυναίκα (και είναι συμμετρική) τότε κάθε άντρας θα είναι γυναίκα (και το αντίστροφο) για τον reasoner<sup>45</sup>. Στη σχέση θα μπορούσε να έχει προστεθεί και το χαρακτηριστικό irreflexive, το οποίο θα δήλωνε ότι ένα άτομο δεν μπορεί να σχετισθεί με τον εαυτό του με την ίδια σχέση. Δεν είχε νόημα να μπει ο παραπάνω περιορισμός στις σχέσεις «νυμφεύεται/παντρεύεται» διότι το πεδίο ορισμού είναι διαφορετικό από το πεδίο τιμών, άρα δεν θα ήταν δυνατό ένας άντρας ή μια γυναίκα να παντρευτεί τον εαυτό της αφού το πεδίο τιμών της ιδιότητας είναι μια κλάση στην οποία το άτομο δεν είναι μέλος.

Στο Σχήμα 9 απεικονίζονται (μέσω Protégé) η ιεραρχία των κλάσεων και η σχέση νυμφεύεται ανάμεσα σε δύο συγκεκριμένα άτομα: το Νίκο και τη Μαρία. Πάνω στη κλάση «Άνθρωπος» έχει δηλωθεί η σχέση «έχει σύζυγο».

<sup>45</sup> Όπως έχει ήδη αναφερθεί στο κεφάλαιο της μεθοδολογίας, τα σχόλια για τις αποφάνσεις του reasoner έχουν ελεγχθεί, ως προς την εγκυρότητά τους, μέσω του Protégé.



**Σχήμα 9:** γραφική απεικόνιση (μέσω Protégé) των σχέσεων νυμφεύεται/παντρεύεται και έχει σύζυγο

Στο Απόσπασμα κώδικα 8 παρουσιάζεται η σχέση «νυμφεύεται» -ως OWL ιδιότητα- σε περιβάλλον πολυανδρίας.

**Απόσπασμα κώδικα 8:** η σχέση 'νυμφεύεται' -ως OWL ιδιότητα- σε περιβάλλον πολυανδρίας

```
<owl:ObjectProperty rdf:about="http://MyPhDExamples#νυμφεύεται">
  <rdfs:subPropertyOf rdf:resource="http://MyPhDExamples#έχει_σύζυγο"/>
  <owl:inverseOf rdf:resource="http://MyPhDExamples#παντρεύεται"/>
  <rdf:type
    rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
  <rdfs:domain rdf:resource="http://MyPhDExamples#Άνδρας"/>
  <rdfs:range rdf:resource="http://MyPhDExamples#Γυναίκα"/>
</owl:ObjectProperty>
```

Άρα οι σχέσεις «έχει σύζυγο», «παντρεύεται» και «νυμφεύεται» αφορούν διακριτές έννοιες σε κάθε ένα από τους παραπάνω πολιτισμούς, αφού διαθέτουν διαφορετικά χαρακτηριστικά. Ως εκ τούτου, πρόκειται για διαφορετικές ιδιότητες στο πλαίσιο μιας οντολογίας, διότι είναι άλλη ιδιότητα το «νυμφεύεται» στη δική μας κοινωνία και άλλη στον πολιτισμό των Toda. Ενώ και η γενικότερη ιδιότητα «έχει σύζυγο» είναι άλλη σε αυτές τις δύο κοινωνίες σε σχέση με μια κοινωνία που επιτρέπει γάμους μεταξύ ομόφυλων ζευγαριών.

Από τα παραπάνω καθίσταται προφανές ότι απαιτούνται διαφορετικές κλάσεις και διαφορετικές ιδιότητες για την περιγραφή του κάθε μεμονωμένου πολιτισμού ακόμη και στις περιπτώσεις όπου περιγράφονται αναλυτικές κατηγορίες οι οποίες θα μπορούσαν να θεωρηθούν ομοειδείς. Αυτό με την σειρά του συνεπάγεται ότι κάθε πολιτισμός

απαιτεί μια διαφορετική οντολογία για την αναπαράστασή του. Στην περίπτωση αυτή όμως ανακύπτει ένα πολύ σοβαρό πρόβλημα που έχει να κάνει με τις δυνατότητες γενίκευσης αλλά και σύγκρισης μεταξύ διαφορετικών πολιτισμών. Η λύση του παραπάνω προβλήματος επιτυγχάνεται μέσω ενός μοντέλου το οποίο επιτρέπει τόσο την αποτύπωση κάθε μεμονωμένου πολιτισμού όσο και την γενίκευση σε ένα αφαιρετικότερο επίπεδο μέσω ολοκλήρωσης (integration) των δεδομένων.

Με αυτό ως ζητούμενο, το μοντέλο που παρουσιάζεται στη συνέχεια έχει διττό σκοπό. Αφενός να παρουσιάσει τα ζητήματα που προκύπτουν για την αποτύπωση της εθνογραφικής γνώσης για έναν πολιτισμό και αφετέρου να αναγάγει αυτή τη γνώση σε ένα αφαιρετικότερο επίπεδο το οποίο θα επιτρέπει τη συγκριτική ανθρωπολογία, με τη γενίκευση που αυτή απαιτεί, εντός ενός πληροφοριακού συστήματος.

Τα όρια μεταξύ της εθνογραφίας και της κοινωνικής ανθρωπολογίας είναι κάπως ασαφή και μεταβάλλονται τόσο μέσα στο χρόνο όσο και ανάλογα με την οπτική του ερευνητή. Παρότι οι λεπτομέρειες της οριοθέτησης δεν αποτελούν αντικείμενο του παρόντος κειμένου θα ακολουθηθεί μια στοιχειώδης διαφοροποίηση η οποία θεωρεί πως η διάκριση κυρίως θεμελιώνεται πάνω στο ότι η εθνογραφία περιγράφει έναν πολιτισμό (ή μια πολιτισμική ομάδα) ενώ η ανθρωπολογία ασχολείται με τις συνθήκες και τις δυνατότητες της ανθρώπινης ζωής στο σύνολό της (Ingold, 2017). Ακολουθώντας αυτή την, ομολογουμένως, χονδροειδή διάκριση θα μπορούσαμε να πούμε ότι το πρώτο επίπεδο ανάλυσης αφορά στην αποτύπωση μιας εθνογραφίας ενώ το δεύτερο αφορά στις γενικεύσεις που επιζητά η ανθρωπολογία. Θα επανέλθουμε στη διάκριση αλλά και στην σύζευξη των δύο στα συμπεράσματα του κεφαλαίου.

### **6.2.2 Πρώτο επίπεδο: αναπαριστώντας έναν πολιτισμό (εθνογραφία)**

Σύμφωνα με το καθιερωμένο παράδειγμα (paradigm) της κοινωνικής ανθρωπολογίας, για την παραγωγή μιας εθνογραφίας η/ο εθνογράφος κάνει επιτόπια, συμμετοχική, παρατήρηση ακολουθώντας μια ποιοτική προσέγγιση η οποία θα του επιτρέψει να κατανοήσει “εκ των έσω” τον πολιτισμό που μελετά. Όμως στο κεφάλαιο αυτό δεν ενδιαφέρει καθόλου η μεθοδολογία δημιουργίας της σχετικής γνώσης αλλά το πώς η γνώση αυτή δύναται να αναπαρασταθεί μέσω τεχνολογιών σημασιολογικού ιστού. Τα ζητήματα που άπτονται της μεθοδολογίας έρευνας, καθαρά από την άποψη των κοινωνικών επιστημών (π.χ. συμμετοχική παρατήρηση ή όχι), θα μπορούσαν να μας απασχολήσουν μόνο στο επίπεδο της καλύτερης τεκμηρίωσης της περιγραφής για την καταγραφή της προέλευσης (provenance) της πληροφορίας. Άρα το ζητούμενο του κεφαλαίου δεν είναι το πώς ο ανθρωπολόγος κατέκτησε μια γνώση ή κατέληξε σε ένα συμπέρασμα. Με την χρήση υπολογιστικών τεχνολογιών για την εξαγωγή συμπερασμάτων ασχοληθήκαμε στο κεφάλαιο «Υπολογιστική επεξεργασία εθνογραφικών δεδομένων: ιστορική αναδρομή, παρόν και προοπτικές». Εδώ το ενδιαφέρον στρέφεται στον τρόπο με τον οποίο η γνώση που υπάρχει -ανεξάρτητα από τη μέθοδο μέσω της οποίας αποκτήθηκε- μπορεί να μοντελοποιηθεί.

Το πρώτο επίπεδο αναπαράστασης έχει να κάνει με τον περιγραφικό πυρήνα της εθνογραφίας. Στον ελληνικό συλλογικό τόμο «Αναστοχασμός, ετερότητα και ανθρωπολογία οίκοι: διλήμματα και αντιπαραθέσεις» (Γκέφου-Μαδιανού, 1998) υπάρχει ένα κείμενο του Peter Loizos με τίτλο «Ο θάνατος του θετικισμού είναι υπερβολή». Στο κείμενο αυτό ο Loizos αναφέρεται στην περιγραφική εθνογραφία, σημειώνοντας πως η ανθρωπολογία δεν μπορεί να αρνείται τον περιγραφικό της ρόλο. Ενώ ξεκαθαρίζει πως ούτε τεχνικά απλές είναι αυτές οι περιγραφές, ούτε θα είναι ποτέ τέλειες και οριστικές, διότι θα έχουν πάντα προσωρινό χαρακτήρα και θα είναι συνάρτηση των δυνατοτήτων παρατήρησης που διαθέτουμε (Loizos, 1998, σελ. 284–289).

Οι διαφωνίες για το βαθμό του περιγραφικού ρόλου της ανθρωπολογίας είναι πολύ παλιές. Ίσως έλαβαν την πιο τυπική τους μορφή τους με την διάκριση ανάμεσα σε *emic* και *etic* προσεγγίσεις. Ο Harris, στο άρθρο «History and significance of the emic/etic distinction», θέτει τη διάκριση σε μια πιο απλοποιημένη και χειροπιαστή βάση. Θεωρεί ότι στο κοινωνικό γίνεσθαι υπάρχει ένας δυισμός όπου από τη μία υπάρχει η συμπεριφορά, η οποία είναι άμεσα παρατηρήσιμη και από την άλλη οι ανθρώπινες σκέψεις και συναισθήματα τα οποία δεν είναι άμεσα παρατηρήσιμα. Εντάσσει την «*etic*» προσέγγιση στην πρώτη περίπτωση και την «*emic*» στη δεύτερη (Harris, 1976). Έτσι, με βάση τον Harris, η *emic* πλευρά έχει έναν πιο γνωσιακό/υποκειμενικό χαρακτήρα ενώ η *etic* πλευρά έχει έναν πιο “αντικειμενικό/συμπεριφορικό” χαρακτήρα. Ο Pike, διαφωνεί με τον Harris έχοντας κυρίως τη γλώσσα ως αφετηρία της προσέγγισής του. Συγκεκριμένα, χρησιμοποιεί τη γλώσσα ως σημείο εκκίνησης για την ανάλυση της λεκτικής συμπεριφοράς την οποία και μεταφέρει στο μη λεκτικό υλικό (*nonverbal material*)<sup>46</sup>. Για την παρούσα διατριβή η διάκριση βασίζεται στον ορισμό που αναφέρεται στη Wikipedia «...*Emic* knowledge and interpretations are those existing within a culture, that are ‘determined by local custom, meaning, and belief’ [...] and best described by a ‘native’ of the culture. *Etic* knowledge refers to generalizations about human behavior that are considered universally true, and commonly links cultural practices to factors of interest to the researcher, such as economic or ecological conditions, that cultural insiders may not consider very relevant». [http://en.wikipedia.org/wiki/Emic\\_and\\_etic](http://en.wikipedia.org/wiki/Emic_and_etic).

Έτσι, το πρώτο επίπεδο σχετίζεται κυρίως με την *emic* πλευρά της περιγραφής. Αυτό όμως δεν σημαίνει ότι θα απουσιάζουν τα “*etic*” στοιχεία. Δηλαδή το ότι, για παράδειγμα, οι Πιραχά (*Pirahã*) δεν έχουν την έννοια του δεξιά-αριστερά<sup>47</sup> αλλά αντ’

<sup>46</sup> Οι διαφορετικές οπτικές ανάμεσα σε Harris και Pike ήταν θέμα ενός συμποσίου που έλαβε χώρα κατά την 87<sup>η</sup> ετήσια συνάντηση της American Anthropological Association το 1988. Τα πρακτικά του εν λόγω συμποσίου κατέληξαν στη δημιουργία του συλλογικού τόμου με τίτλο «*Emics and Etics: The Insider/Outsider Debate*» (Headland κ.ά., 1990)

<sup>47</sup> Γενικά για τον προσανατολισμό και την κίνηση στο χώρο υπάρχουν δύο βασικές προσεγγίσεις. Η μία είναι εκείνη που είναι εγωκεντρική (ως εκ τούτου σχετική) και περιλαμβάνει έννοιες όπως δεξιά-αριστερά και πάνω-κάτω. Η δεύτερη έχει να κάνει με σταθερά σημεία αναφοράς, όπως τα σημεία του ορίζοντα. Στη βιβλιογραφία συχνά αναφέρεται ότι πολιτισμοί που ζουν στις ερήμους (όπως οι Αβορίγινες της Αυστραλίας) ή σε ζούγκλες (όπως οι Πιραχά της Αμαζονίας) έχουν αναπτύξει μεγάλη ικανότητα στο να

αυτού «πάνω το ποτάμι-κάτω το ποτάμι», δηλαδή προσανατολίζονται με τη ροή του ποταμού δίπλα στον οποίο διαβιούν (Everett, 2009), δεν σημαίνει ότι στο επίπεδο της περιγραφής του πολιτισμού τους από ανθρωπολόγους δεν θα υπάρχει η γεωγραφική τους θέση με βάση το δυτικό σύστημα συντεταγμένων ή η χώρα μέσα στην οποία ανήκει η κοινότητά τους.

Ένα μέρος αυτών των etic χαρακτηριστικών θα μπορούσε να αφορά σε data type properties και όχι σε σχέσεις μεταξύ κλάσεων. Για παράδειγμα ο πληθυσμός της κοινότητας, η έκταση που καταλαμβάνει η κοινότητα, οι γεωγραφικές συντεταγμένες. Κάποιες όμως από τις etic πληροφορίες είναι πληροφορίες σχέσεων μεταξύ οντοτήτων, όπως για παράδειγμα η χώρα (ή οι χώρες) στην οποία κατοικούν, οι γειτονικοί πληθυσμοί με τους οποίους έχουν φιλικές ή/και εχθρικές σχέσεις.

Συνοπτικά θα μπορούσε να λεχθεί ότι το βασικό ζήτημα που απαιτείται για την αναπαράσταση ενός πολιτισμού είναι η καταγραφή των βασικών εννοιών και το πώς διασυνδέονται μεταξύ τους μέσα στον εκάστοτε πολιτισμό; Δηλαδή, με όρους OWL, ποιες θα είναι οι κλάσεις και ποιες οι σχέσεις αυτών των κλάσεων. Προφανώς για την αποτελεσματικότερη υλοποίηση υπάρχει ανάγκη τόσο από οντολογικές προσεγγίσεις όσο και από λεξιλόγια τιμών (value vocabularies) τα οποία θα τις εξειδικεύουν.

### 6.2.3 Δεύτερο επίπεδο: γενικεύοντας σε ένα αφαιρετικό επίπεδο (κοινωνική ανθρωπολογία)

Η ανθρωπολογία, ως επιστήμη που έχει αντικείμενο την πολιτισμική ετερότητα, βασίζεται, από τη φύση της, στη σύγκριση. Στο πλαίσιο της ανθρώπινης κατανόησης του κόσμου αντιπαραβάλλεται το οικείο του παρατηρητή με την ετερότητα του παρατηρούμενου. Όμως στο πλαίσιο ενός πληροφοριακού συστήματος τίποτα δεν είναι οικείο ή ξένο. Κάθε μία από τις οντολογικές αναπαραστάσεις των επιμέρους πολιτισμών είναι μια αυτόνομη περιγραφή. Όπως ήδη αναλύθηκε στην προηγούμενη ενότητα, οι κλάσεις και οι ιδιότητές τους (ειδικά αυτές που αφορούν στην εσωτερική δομή της πολιτισμικής ομάδας και τον τρόπο που η ίδια αντιλαμβάνεται τον κόσμο, δηλαδή η 'μικρή' πλευρά) διαφέρουν από πολιτισμό σε πολιτισμό. Το πρόβλημα που προκύπτει είναι ότι η επιμέρους περιγραφή των πολιτισμών δεν επιτρέπει τη σύγκριση μεταξύ τους. Για τη λύση αυτού του προβλήματος μπορεί να γίνει χρήση ενός αφαιρετικότερου επιπέδου περιγραφής στο οποίο να αντιστοιχίζονται τα επιμέρους στοιχεία εκάστου πολιτισμού.

Μένοντας στο παράδειγμα των συγγενικών σχέσεων, που αναφέρθηκε προηγουμένως, θα μπορούσε να τεθεί το εξής ερώτημα: υπάρχει κάποια γενική κατηγορία η οποία θα μπορούσε να ονομαστεί συγγένεια και αν ναι τότε ποιες είναι οι σχέσεις που θα έπρεπε να περιλαμβάνονται σε αυτή; Μια πρώτη απάντηση θα μπορούσε να είναι: το σύνολο (δηλαδή το λογικό OR) των σχέσεων που έχουμε

---

διαχειρίζονται τα σταθερά σημεία αναφοράς και να κινούνται σε σχέση με αυτά. Αυτή η ικανότητα αποδίδεται στην ανάγκη να μην χαθούν σε ένα χώρο όπου το τοπίο επαναλαμβάνεται με το ίδιο μοτίβο για μεγάλες αποστάσεις.

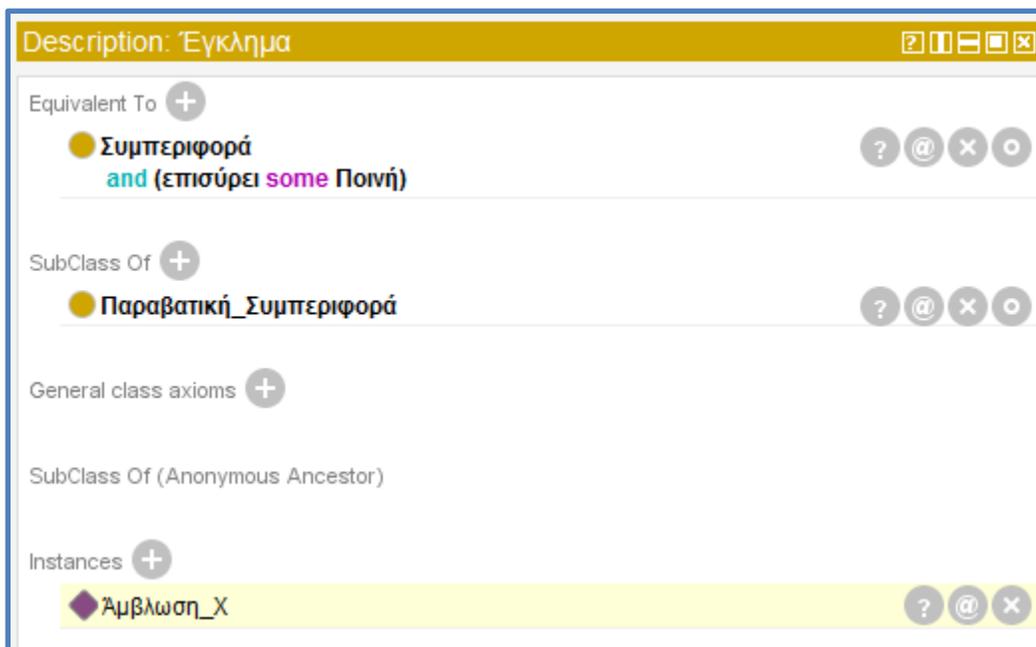
κατηγοριοποιήσει ως συγγενικές στους επιμέρους πολιτισμούς. Όμως ποια είναι τα χαρακτηριστικά αυτών των σχέσεων και κατά πόσο είναι κοινά στους διάφορους πολιτισμούς;

Μένοντας στο παράδειγμα του γάμου -και στηριζόμενοι στα δεδομένα που παρουσιάστηκαν στην ενότητα 6.2.1- προκύπτει ότι στο γενικό επίπεδο η σχέση του γάμου συνδέει ένα άτομο με ένα άλλο άτομο (ανεξάρτητα από φύλο και περιορισμούς πληθικότητας) και είναι μη ανακλαστική (irreflexive). Αυτά με την προϋπόθεση ότι δεν υπάρχουν πολιτισμοί όπου μπορεί να υπάρχει γάμος μεταξύ ανθρώπου και ζώων ή/και πνευμάτων. Ή μήπως επειδή είναι δυνατόν να υπάρχει γάμος μεταξύ ζώων και ανθρώπων πρέπει να αλλάξει ο ορισμός; Ή μήπως αυτού του τύπου οι “γάμοι” επιθυμούμε να αποτελούν άλλη κατηγορία σχέσεων; Αυτός ο βαθμός λεπτομέρειας (granularity), δηλαδή ο βαθμός στον οποίο μια μεγαλύτερη οντότητα εξειδικεύεται, ή ο βαθμός στον οποίο ομάδες μικρότερων δυσδιάκριτων οντοτήτων έχουν ενωθεί για να δημιουργήσουν μεγαλύτερες διακριτές οντότητες, είναι πολύ κρίσιμος παράγοντας για να είναι εφικτή η γενίκευση. Όμως δεν είναι μέρος της διατριβής η απάντηση αυτών των ερωτημάτων. Οι ανθρωπολόγοι, ως μέλη ενός επιστημονικού κλάδου, οφείλουν να απαντήσουν σε αυτά τα ερωτήματα και όχι μια διατριβή ή πολλώ μάλλον οι/η πληροφορικοί/η!

Κατά συνέπεια, οι κατηγορίες και οι σχέσεις του ανώτερου επιπέδου είναι γενικεύσεις καθολικής ισχύος παρότι μπορεί να αφορούν σε κοινωνικές πρακτικές με διαφορετικά, ή και αντικρουόμενα, χαρακτηριστικά. Με άλλα λόγια, το ποια συμπεριφορά είναι εγκληματική και ποια όχι μπορεί να διαφέρει από πολιτισμό σε πολιτισμό. Οπότε δεν είναι δυνατόν να διαμορφωθεί μια γενική λίστα εγκληματικών συμπεριφορών. Επίκαιρο παράδειγμα η αλλαγή του νομικού πλαισίου που ορίζει τις αμβλώσεις στις ΗΠΑ. Παρόλα αυτά, είναι εφικτό να δοθεί μια γενική περιγραφή, σε ένα αφαιρετικό επίπεδο, του τύπου έγκλημα είναι μια παραβατική συμπεριφορά η οποία επισύρει κάποια ποινή. Αυτός ο ορισμός έχει γενική ισχύ ανεξάρτητα από τι είναι παραβατικό και τι όχι σε κάθε πολιτισμό. Άρα το «επισύρει» μπορεί να είναι μια ιδιότητα η οποία έχει πεδίο ορισμού την κλάση «Συμπεριφορά» και πεδίο τιμών την κλάση «Ποινή». Στο Σχήμα 10 παρουσιάζεται μια ενδεικτική φορμαλιστική αποτύπωση του παραπάνω ορισμού στο Protégé και η (αυτόματη) ένταξη σε αυτόν μιας συγκεκριμένης άμβλωσης. Αυτό δεν συνεπάγεται ότι κάθε άμβλωση θεωρείται εγκληματική πράξη αλλά μόνο σε εκείνους τους πολιτισμούς όπου μια τέτοια πράξη δύναται να επιφέρει ποινή<sup>48</sup>.

---

<sup>48</sup> Ο συγγραφέας χρησιμοποιεί το παράδειγμα της άμβλωσης διότι θεωρεί ότι είναι ένα επίκαιρο και εύληπτο παράδειγμα για τις διαφορετικές προσεγγίσεις του τι μπορεί να οριστεί ως παραβατική συμπεριφορά και το πώς αυτό μπορεί να μεταβάλλεται μέσα στο χρόνο ακόμη και στον ίδιο πολιτισμό. Σε κάποιες αμερικάνικες πολιτείες ήταν έγκλημα, έπαψε να είναι και πρόσφατα επανήλθε ως παραβατική συμπεριφορά. Όπως εύστοχα παρατηρεί ο Sahlins για τις αλλαγές και την ρευστότητα των πολιτισμών “One of the current arguments against the coherence of cultures and the possibility of doing any kind of systematic ethnography is that, like a certain famous philosophical river, cultures are always changing. Such is the flux that one can never step in the same culture twice. [...] So to paraphrase John Barth, reality is a nice place to visit (philosophically), but no one ever lived there” (Sahlins, 2002, σελ. 7).



Σχήμα 10: φορμαλιστικός ορισμός εγκληματικής πράξης και αυτόματης ένταξης σε αυτή την κλάση μιας πράξης η οποία αποτελεί έγκλημα μόνο σε κάποιους πολιτισμούς

Στο Απόσπασμα κώδικα 9 παρουσιάζεται ο XML/OWL κώδικας που αντιστοιχεί στον ορισμό μιας συμπεριφοράς ως έγκλημα από τη στιγμή που αυτή επισύρει κάποια ποινή. Ο εν λόγω κώδικας επιτρέπει στο reasoner να εντάξει μια πράξη στην κλάση των εγκλημάτων.

Απόσπασμα κώδικα 9: ορισμός του εγκλήματος - σε OWL- ως Συμπεριφορά που επισύρει Ποινή

```

<owl:Class rdf:about="http://MyPhDExamples#Έγκλημα">
  <owl:equivalentClass>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <rdf:Description rdf:about="http://MyPhDExamples#Συμπεριφορά"/>
        <owl:Restriction>
          <owl:onProperty rdf:resource="http://MyPhDExamples#επισύρει"/>
          <owl:someValuesFrom rdf:resource="http://MyPhDExamples#Ποινή"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </owl:equivalentClass>
</owl:Class>

```

Πρέπει να καταστεί απολύτως σαφές ότι η Χ πράξη έχει ένα μονοδιάστατο χαρακτήρα για το επίπεδο του φυσικού κόσμου. Όμως ο χαρακτηρισμός της σε έγκλημα, έπαινο ή κάτι άλλο, αποτελεί αξιακή κρίση η οποία βασίζεται σε πολιτισμικά συμφραζόμενα. Άρα η αποχή από το φαγητό είναι μια πράξη που λαμβάνει χώρα στον φυσικό κόσμο. Όμως η νηστεία δεν είναι απλά η αποχή από το φαγητό αλλά μια αξιακή διάσταση που αποδίδεται στην αποχή. Αυτό σημαίνει ότι η ιεραρχία των κλάσεων αλλά και των ιδιοτήτων που αφορούν στον φυσικό κόσμο (τρώω, σκοτώνω κτλ) δεν ταυτίζονται με το νόημα που οι πράξεις αυτές αποκτούν μέσα στα εκάστοτε πολιτισμικά συμφραζόμενα. Για παράδειγμα η πράξη της χορήγησης μίας θανατηφόρας δόσης φαρμάκων σε έναν άνθρωπο -με τη χρήση μιας σύριγγας- έχει ως, ντετερμινιστικό, αποτέλεσμα το θάνατο αυτού που δέχεται τη δόση. Αυτό είναι το γεγονός στα πλαίσια του φυσικού κόσμου. Όμως στα πλαίσια του κοινωνικού γίνεσθαι αυτός που έκανε την θανατηφόρα δόση μπορεί να είναι εγκληματίας διότι δολοφόνησε κάποιον (άρα η πράξη θεωρείται φόνος) γιατρός, ο οποίος έκανε ευθανασία σε κοινωνία όπου αυτό επιτρέπεται (άρα η πράξη θεωρείται μια ιατρική πράξη) μουσικός πράκτορας σε εντεταλμένη υπηρεσία (άρα η πράξη θεωρείται ηρωική για τη χώρα του και εχθρική για αυτήν του θύματος) ή ο δήμιος σε μια χώρα η οποία εφαρμόζει τη θανατική ποινή (άρα, για τον δήμιο, η πράξη του ήταν απλώς άλλη μια μέρα στη δουλειά!) κοκ.

Όταν όμως επιθυμούμε να κάνουμε γενίκευση για το πώς κάτι γίνεται αντιληπτό από το σύνολο των ανθρώπων, δεν έχουμε να κάνουμε με ένα συγκεκριμένο δικονομικό σύστημα, ούτε καν με δικονομικά συστήματα γενικώς. Διότι η προς περιγραφή πολιτισμική ομάδα μπορεί να μην έχει ως σημείο αναφοράς ένα σύστημα δικαίου μιας χώρας αλλά τις επιταγές που προτάσσει μια θρησκεία. Έτσι, μένοντας στο προηγούμενο παράδειγμα, από τη μία έχουμε τη χορήγηση της δόσης που οδήγησε στο θάνατο μέσα σε ένα ελβετικό κέντρο ευθανασίας (γεγονός) και από την άλλη την ερμηνεία του γεγονότος (φόνος για την ορθόδοξη εκκλησία, ιατρική πράξη για το Ελβετικό δίκαιο).

Άρα αυτό που περιγράφουν οι κοινωνικοί επιστήμονες (και ειδικά οι ανθρωπολόγοι) δεν είναι το τι συμβαίνει στο φυσικό κόσμο αλλά το πώς μια ομάδα ανθρώπων αντιλαμβάνεται και ερμηνεύει τα όσα συμβαίνουν στο φυσικό και κοινωνικό γίνεσθαι. Έτσι, όπως σημειώνουν οι Chung & Lieberman, τα πιρούνια και τα chopsticks μπορεί να είναι πιο κοντά από ότι είναι τα πιρούνια με τα πιρούνια! Διότι η έννοια «πιρουνιών και μαχαιριών» στην αμερικανική κουλτούρα είναι περισσότερο κοντά στην έννοια των «chopsticks» παρά με εκείνη των «πιρουνιών και μαχαιριών» στην κορεατική κουλτούρα (Chung & Lieberman, 2007, σελ. 76). Και αυτό έρχεται σε πλήρη σύμπνοια με τη θεώρηση του Wittgenstein όπως αυτή συνοψίστηκε στη φράση «μην αναζητάτε το νόημα, αναζητήστε τη χρήση<sup>49</sup>» (Goguen, 1974, σελ. 546), ή όπως γράφει ο ίδιος ο Wittgenstein στις Φιλοσοφικές Έρευνες «η σημασία μιας λέξης είναι η χρήση της μέσα στη γλώσσα» (Wittgenstein, 1977, παρ. 43).

---

<sup>49</sup> Don't ask for the meaning, ask for the use.

Κλείνοντας την ενότητα πρέπει να σημειωθεί ότι το δεύτερο επίπεδο πρέπει να αναπτυχθεί υποχρεωτικά με μια προσέγγιση από κάτω προς τα πάνω (bottom-up). Η αντίστροφη προσέγγιση θα ήταν απόλυτα μεροληπτική διότι θα ενέτασσε (ίσως το «υποδούλωνε» είναι ακριβέστερος χαρακτηρισμός!) στο πολιτισμικό και αξιακό σύστημα του/της ανθρωπολόγου τα επιμέρους συστήματα.

### 6.3 Συμπέρασμα κεφαλαίου

Παρουσιάστηκε ένα μοντέλο το οποίο δίνει τη δυνατότητα αναπαράστασης τόσο εθνογραφικών δεδομένων όσο και γενίκευσης αυτών ώστε να μοντελοποιηθεί η ανθρωπολογική γνώση. Ανεξάρτητα από το αν θα ονομάσουμε το πρώτο επίπεδο Εθνογραφία και το δεύτερο Ανθρωπολογία η ουσία είναι ότι αυτά τα δύο αποτελούν ένα ενιαίο λειτουργικό σύνολο. Άλλωστε είναι ευρέως αποδεκτό στο χώρο της ανθρωπολογίας ότι το πρώτο επίπεδο δεν έχει κάποιο ιδιαίτερο νόημα αν απλά αποτελεί μια παράθεση χαρακτηριστικών μιας κοινωνικής ομάδας.

Ο τρόπος που αυτά τα δύο επίπεδα συνδιαλέγονται μεταξύ τους και πώς αυτό επηρεάζει τόσο την emic και etic προσέγγιση, αλλά και το τι είναι Ανθρωπολογία ή Εθνογραφία, συνοψίζεται εύστοχα από τον Sahlins στην ενότητα Etics and Emics του έργου «Waiting for Foucault, Still».

*“No good ethnography is self-contained. Implicitly or explicitly ethnography is an act of comparison. By virtue of comparison ethnographic description becomes objective. Not in the naive positivist sense of an unmediated perception— just the opposite: it becomes a universal understanding to the extent it brings to bear on the perception of any society the conceptions of all the others” (Sahlins, 2002, σελ. 12).*

Θα μπορούσε κάποιος να πει ότι το επίπεδο ολοκλήρωσης είναι ένα μεροληπτικά κατασκευασμένο etic επίπεδο, όμως δεν είναι έτσι. Διότι δεν επιβάλλει μια από πάνω προς τα κάτω προσέγγιση αλλά ενσωματώνει όλες τις προσεγγίσεις σε ένα ενιαίο λειτουργικό σύνολο.

Ο στόχος του κεφαλαίου, και της διατριβής κατ' επέκταση, δεν ήταν η δημιουργία μιας οντολογίας αλλά η πρόταση ενός γενικού μοντέλου το οποίο επιτρέπει τη μοντελοποίηση βάσει τεχνολογιών σημασιολογικού ιστού, δίνοντας ενδεικτικά παραδείγματα πιθανών υλοποιήσεων. Αυτό σημαίνει ότι δεν αποτελεί θέμα της διατριβής η μοντελοποίηση του Χ πολιτισμικού ή κοινωνικού φαινομένου αλλά η εκφραστική δύναμη και η καταλληλότητα των παραπάνω τεχνολογιών ώστε να περιγράψουν τα εν λόγω φαινόμενα, εφαρμόζοντάς τα σε συγκεκριμένα παραδείγματα όπως η συγγένεια. Όμως οι προσεγγίσεις μοντελοποίησης που θα ακολουθηθούν δεν επιβάλλονται μονοσήμαντα από τις εν λόγω τεχνολογίες<sup>50</sup>. Διότι, παρότι η οντολογική

---

<sup>50</sup> Για το λόγο αυτό η προσέγγιση των διακριτών οντολογιών ανά πολιτισμό βρίσκεται σε απόλυτη σύμπτωση με το ρεύμα της «οντολογικής στροφής» (ontological turn) όπως αυτό έχει εκφραστεί τον 21<sup>ο</sup> αιώνα. Στο πλαίσιο της «οντολογικής στροφής» ο όρος οντολογία αναφέρεται στους τρόπους με τους

αναπαράσταση γνώσης προϋποθέτει ένα συγκεκριμένο τρόπο μέσω του οποίου αναπαρίσταται κάτι, δεν υπάρχει μία και μοναδική οπτική αναπαράστασης του κόσμου. Για αυτό αναφερόμαστε συχνά στις οντολογίες ως κοινωνικοτεχνικά δημιουργήματα.

Για να καταστεί σαφέστερη η προηγούμενη δήλωση, θα εξειδικεύσουμε λίγο περισσότερο στο παράδειγμα των ιδιοτήτων παντρεύεται/νυμφεύεται που αναλύθηκαν πριν. Αν χρησιμοποιηθούν οι ιδιότητες παντρεύεται/νυμφεύεται για να περιγραφεί ένας γάμος τότε τίθεται το θέμα του τι συμβαίνει σε περίπτωση διαζυγίου και νέου γάμου; Αν έχει δηλωθεί η προηγούμενη ιδιότητα «νυμφεύεται» (όπως ορίστηκε στο Απόσπασμα κώδικα 7) και δηλωθεί νέος γάμος για τη Μαρία τότε αυτό το άτομο, για το reasoner, θα είναι το ίδιο άτομο (same individual) με το Νίκο. Αν δηλωθεί ως διαφορετικό άτομο από το Νίκο τότε η οντολογία θα είναι ασυνεπής (inconsistent)<sup>51</sup>. Μία πιθανή λύση είναι η χρήση διαφορετικών ιδιοτήτων. Μια ιδιότητα που δηλώνει το «είναι παντρεμένη/νυμφευμένος» (δηλαδή την τρέχουσα κατάσταση και θα έχει περιορισμό πληθικότητας «1») και μια άλλη ιδιότητα για το «ήταν παντρεμένη/νυμφευμένος» (δηλαδή παρελθούσες καταστάσεις και δεν θα έχει περιορισμό πληθικότητας στην περίπτωση που δεν υπάρχει περιορισμός στον αριθμό των γάμων).

Όμως αυτή δεν είναι η μοναδική λύση μοντελοποίησης. Μια αρκετά διαφορετική προσέγγιση μοντελοποίησης θα μπορούσε να βασίζεται στο CIDOC-CRM για το οποίο υπήρξε μια γενική παρουσίαση στην υπο-ενότητα 5.5.3.4. Στην περίπτωση που ακολουθούνταν το CIDOC-CRM -ή κάποια προσέγγιση συμβατή με αυτό- θα έπρεπε να δημιουργηθεί μια κλάση η οποία θα αφορά στο γάμο ως γεγονός στο οποίο συμμετέχουν τα άτομα. Δηλαδή δεν υπάρχει μια απευθείας σχέση μεταξύ των ατόμων, αλλά η σχέση προκύπτει μέσω του γεγονότος του γάμου. Προφανώς αυτή είναι δομή είναι πολύ πιο εκφραστική αφού ο εκάστοτε γάμος (το εκάστοτε στιγμιότυπο) θα έχει με τη σειρά του τις δικές του ιδιότητες. Έτσι δεν δηλώνεται απλά ότι ένα άτομο συμμετέχει, ως σύζυγος, σε ένα συγκεκριμένο γάμο αλλά ταυτόχρονα μπορούν να προστεθούν και άλλες ιδιότητες όπως το πότε ξεκίνησε ο γάμος, το αν έχει τελειώσει, τα παιδιά που προέκυψαν από αυτόν κλπ.

Γενικά το CIDOC-CRM έχει έναν εξαιρετικό μηχανισμό για να μοντελοποιήσει τα γεγονότα και την ιστορική τους διάσταση αφού συνδυάζει με επιτυχία χρονικές περιόδους, γεγονότα, δράστες και αντικείμενα (νοητικά ή φυσικά) που εκείνοι παράγουν (βλέπε Σχήμα 7). Άρα ο πυρήνας του βασικού του μηχανισμού δείχνει ότι περιέχει αξιόλογα στοιχεία για να ληφθεί υπόψη ως πιθανή υποψήφια οντολογία. Όμως διαθέτει έναν πολύ βασικό περιορισμό που τελικά δεν επιτρέπει τη χρήση του σε ένα σημαντικό μέρος των ανθρωπολογικών δεδομένων διότι δεν προβλέπει την περιγραφή οντοτήτων οι οποίες δεν υπάρχουν στον πραγματικό κόσμο. Έτσι δεν μπορεί να μοντελοποιήσει θεούς, μυθολογία κτλ. Παρόλα αυτά, η εξαιρετική οπτική του για τη μοντελοποίηση των

---

οποίους διαφορετικές κοινωνίες αντιλαμβάνονται την κοινωνική πραγματικότητα και τον φυσικό κόσμο (Holbraad & Pedersen, 2017).

<sup>51</sup> Υπενθυμίζεται πως όλες οι αποφάνσεις αυτού του τύπου έχουν ελεγχθεί με τον Hermit Owl Reasoner μέσω του λογισμικού Protégé.

γεγονότων και της εξέλιξης αυτών μέσα στο χρόνο αξίζει να ληφθεί υπόψη διότι έχει εφαρμογή στην αποτύπωση φαινομένων που μελετά η ανθρωπολογία.

Ο φορμαλισμός που περιγράφηκε σε αυτό το κεφάλαιο δίνει από τη μία πλευρά μεγάλη εκφραστική δυνατότητα και από την άλλη εξαιρετική ακρίβεια, ενώ ταυτόχρονα παρέχει δυνατότητες υπολογιστικού συλλογισμού. Παράλληλα όμως διαθέτει ένα ακόμη πλεονέκτημα. Μια φορμαλιστική περιγραφή μπορεί να μετατραπεί σε φυσική γλώσσα πολύ πιο εύκολα από ότι το αντίθετο. Έχουν αναπτυχθεί διάφορα εργαλεία (συνήθως ονομάζονται OWL verbalizers)<sup>52</sup> τα οποία μπορούν να μετατρέψουν μια οντολογία σε φυσική γλώσσα. Μια συστηματική εργασία πάνω στο θέμα της μετατροπής της OWL σε φυσική γλώσσα έχουν αναπτύξει οι Ανδρουτσόπουλος κ.α. με την ονομασία NaturalOWL (Androutsopoulos κ.ά., 2013).

Όπως συνήθως συμβαίνει με κάθε μοντελοποίηση, υπάρχουν σημεία που χρήζουν ιδιαίτερης προσοχής διότι θα έχουν καταλυτική επίδραση στη συμπερασματολογία. Ένα τέτοιο ζήτημα προκύπτει από τον τρόπο που οι υποκλάσεις σχετίζονται με τις κλάσεις. Η σχέση «is-a» αποτελεί έναν εξαιρετικό μηχανισμό γενίκευσης αφού η μεταβατική ιδιότητα που την χαρακτηρίζει σηματοδοτεί ότι μια κλάση κληροδοτεί όλα τα χαρακτηριστικά της στις άμεσες και έμμεσες υποκλάσεις της. Όμως στο χώρο των κοινωνικών επιστημών αυτού του είδους η στερεοτυπική σκέψη πολύ απέχει από την πραγματικότητα. Η ρευστή -και πολλές φορές ασαφής- φύση της κοινωνικής πραγματικότητας μπορεί να εκληφθεί ως μια στατική και απόλυτη δομή με κίνδυνο η συλλογιστική του αυτόματου συμπερασμού να καθίσταται απολύτως προβληματική. Για παράδειγμα, η επίσημη γλώσσα της Νιγηρίας<sup>53</sup> είναι η Αγγλική. Το παράδειγμα που ακολουθεί έχει να κάνει με την αποφυγή του λογικού άλματος: όταν μια χώρα έχει ως επίσημη γλώσσα τη Χ δεν σημαίνει, κατ' ανάγκη, ότι αυτή είναι μια γλώσσα που γνωρίζουν όλες οι εθνοτικές ομάδες που κατοικούν σε αυτή τη χώρα. Αν λοιπόν αυτή η ιδιότητα αποδοθεί σε επίπεδο χώρας αυτό θα σημαίνει ότι όλες οι εθνοτικές ομάδες μιλούν αυτή τη γλώσσα. Επίσης, με ανάλογη συλλογιστική, αν ένα άτομο ανήκει σε μία κλάση τότε έχει όλα τα χαρακτηριστικά αυτής της κλάσης. Άρα, στο παραπάνω παράδειγμα, κάθε Νιγηριανός μιλάει αγγλικά. Κάτι το οποίο προφανώς και δεν ισχύει και θα οδηγήσει σε λανθασμένα συμπεράσματα στην πορεία.

Αυτό που περιγράφει η προηγούμενη παράγραφος δεν μια είναι εγγενής αδυναμία της οντολογικής αναπαράστασης γνώσης. Διότι είναι εφικτό να χρησιμοποιηθεί η διάκριση ανάμεσα στο «owl:allValuesFrom» (Universal quantification) και το «owl:someValuesFrom» (Existential Quantification) ώστε να αποφευχθούν εσφαλμένα συμπεράσματα. Η διαφορά μεταξύ του πρώτου και του δεύτερου έγκειται στο ότι το «owl:someValuesFrom» αφορά καθολικό περιορισμό (δηλαδή όλα τα

---

<sup>52</sup> Ενδεικτική τέτοια προσπάθεια διατίθεται σε on-line περιβάλλον στη διεύθυνση [http://attempto.ifi.uzh.ch/site/docs/owl\\_to\\_ace.html](http://attempto.ifi.uzh.ch/site/docs/owl_to_ace.html). Όμως αντίστοιχη λειτουργικότητα υποστηρίζεται και από το Ontorion.

<sup>53</sup> Σύμφωνα με το Ethnologue (<https://www.ethnologue.com/country/NG/>), στη Νιγηρία ομιλούνται περίπου 520 γλώσσες.

μέλη πληρούν τη συνθήκη), ενώ το «owl : someValuesFrom» αφορά τουλάχιστον ένα μέλος που πληροί την συνθήκη, χωρίς να ισχύει για όλα τα μέλη. Με τον τρόπο αυτό μπορούν να αποφευχθούν οι λανθασμένες γενικεύσεις.

Από την άλλη πλευρά είναι γεγονός ότι για να επιτευχθεί ο αυτόματος συλλογισμός απαραίτητη προϋπόθεση είναι η συμφωνία, έστω μιας ομάδας ανθρωπολόγων, για τις βασικές ιδιότητες οι οποίες είναι αναγκαίες για να περιγράψουν επαρκώς ένα κοινωνικοπολιτισμικό (sociocultural) φαινόμενο με τη μορφή γράφου γνώσης. Και τα όσα προηγήθηκαν σε αυτό το κεφάλαιο είχαν ως στόχο την ανάπτυξη ενός μοντέλου το οποίο επιτρέπει τη δημιουργία αυτών των γράφων γνώσης. Όμως σε καμία περίπτωση δεν εξασφαλίζει την συμφωνία των ανθρωπολόγων για το τι αυτοί οι γράφοι θα περιλαμβάνουν.

Τέλος, θα πρέπει να διασαφηνιστεί ότι δεν είναι απαραίτητη η γενική συμφωνία για το τι θα περιλαμβάνεται στους γράφους γνώσης, αφού είναι δυνατόν να υπάρχουν διαφορετικές οπτικές για το ίδιο κοινωνικοπολιτισμικό φαινόμενο, οι οποίες μπορεί να συνοδεύονται με τις κατάλληλες δηλώσεις για την προέλευση (provenance) της πληροφορίας κάθε φορά.



## 7 Συζήτηση

«Ο άνθρωπος είναι το μόνο ον που παράγει μια αδιάκοπη πανσπερμία νέων τεχνολογιών [...] που μεταμορφώνουν αδιάλειπτα τις ημέρες και τους αιώνες του» (Χατζηβασιλείου, 2023, σελ. 18). Και παρότι, από την εποχή των πρωτεργατών της κοινωνικής ανθρωπολογίας, ο Malinowski έγραφε πως ο εθνογράφος πρέπει να εμπνέεται από τη γνώση των πιο σύγχρονων αποτελεσμάτων της επιστημονικής μελέτης (Malinowski, 2014, σελ. 8), οι ανθρωπολόγοι σε αρκετές περιπτώσεις κλείστηκαν σε ένα δικό τους κόσμο. Έτσι συχνά πυκνά διακρίνουν τους συναδέλφους τους (ως ένα βαθμό και δικαιολογημένα) ανάμεσα σε εκείνους που πήγαν και είδαν και σε εκείνους που δεν πήγαν και δεν γνώρισαν από πρώτο χέρι. Ενδεικτικό, για τον τρόπο που αντιμετωπίζεται η επιτόπια έρευνα, είναι το κείμενο του Nigel Barley:

*Εκείνοι είχαν τραβήξει τα πάνδεινα, είχαν υποστεί δοκιμασίες και στερήσεις σε θάλτους και ζούγκλες, δεν θα επέτρεπαν σε κανένα σχολιαρόπαιδο να κόψει δρόμο.*

*Αν τύχαινε να βρεθούν σε δύσκολη θέση σχετικά μ' ένα θεωρητικό ή μεταφυσικό ζήτημα, κουνούσαν το κεφάλι τους λυπημένα, ρουφούσαν ράθυμα την πίπα τους ή χάιδευαν τα γένια τους και μουρμούριζαν κάτι για τους «πραγματικούς ανθρώπους» που δεν προσαρμόζονταν στις αφηρημένες έννοιες όσων «δεν είχαν κάνει ποτέ έρευνα πεδίου». Εκδήλωναν αληθινή συμπόνια για κείνους τους στερημένους τύπους, ενώ για τους ίδιους η υπόθεση ήταν σαφέστατη. Την είχαν ζήσει, την είχαν δει με τα μάτια τους. Δεν είχαν τίποτα άλλο να πουν (Barley, 2005, σελ. 8).*

Μόνο που, όπως ήδη αναφέρθηκε προηγουμένως, για την επιστήμη μικρή -αν όχι καθόλου- σημασία έχει το τι είδε ή τι βίωσε ο παρατηρητής. Σημασία έχει κατά πόσο αυτό που είδε ή βίωσε μπορεί να το μεταφέρει στην υπόλοιπη κοινότητα των επιστημόνων, να καταστήσει κοινωνούς τους συναδέλφους του στις παρατηρήσεις και τα βιώματά του. Διότι το ζητούμενο είναι να κατανοήσουν και οι υπόλοιποι. Αυτή η κατανόηση των συναδέλφων είναι ιδιαίτερα κρίσιμη και τη μη επαρκή επίτευξή της στηλιτεύει ο Fischer γράφοντας: «όπως συμβαίνει με όλους τους κλάδους που εστιάζουν στους ανθρώπους και την παραγωγή τους, η ανθρωπολογία τείνει να είναι πλούσια σε πληροφορίες και φτωχή σε κατανόηση» (M. D. Fischer, 2006, σελ. 3).

Για να επιτευχθεί με αποτελεσματικό τρόπο η κατανόηση από τους επιστήμονες, οι πληροφορίες που ανταλλάσσονται πρέπει να είναι λειτουργικά δομημένες ώστε να επιτελούν επιτυχώς το ρόλο τους. Η παρούσα διατριβή εστιάζει σε προσεγγίσεις που χρησιμοποιούν φορμαλιστικές μεθόδους οι οποίες είναι πιο πρόσφορες για αλγοριθμική επεξεργασία. Ο φορμαλισμός αυτού του τύπου δεν έχει στόχο την απλοποίηση των νοημάτων, αλλά την αποτύπωση του νοήματος με τη μεγαλύτερη δυνατή σαφήνεια / ακρίβεια, ως εκ τούτου με τις λιγότερες αμφισημίες. Διότι οι εκφραστικές δυνατότητες ενός συστήματος δεν σχετίζονται με το επίπεδο του φορμαλισμού του αλλά με το επίπεδο ανάλυσης της λεπτομέρειας (granularity) που αυτό μπορεί να υποστηρίξει. Συνεπώς, ο βαθμός ανάλυσης -και κατ' επέκταση η εκφραστικότητα- που επιτρέπει εκάστοτε σύστημα ευθύνεται για τις όποιες απλοποιήσεις ή γενικεύσεις, και όχι ο βαθμός του φορμαλισμού του.

Έχει ήδη γίνει αναφορά στον πρόλογο της παρούσας διατριβής ότι τα μοντέλα μας δεν αποτελούν μια απλουστευμένη και στείρα αναπαράσταση της πραγματικότητας, αλλά έναν δημιουργικό τρόπο για να την αντιληφθούμε. Διότι δεν υπάρχει ένας προεγκαταστημένος βιολογικός μηχανισμός ο οποίος επιτρέπει την ολική και αντικειμενική κατανόηση της πραγματικότητας. Τα μοντέλα αυτά μπορεί να μην γίνονται καν αντιληπτά από τα άτομα που τα χρησιμοποιούν. Όμως, για το χώρο της επιστήμης, η σαφής φορμαλιστική αποτύπωση των μοντέλων είναι καθοριστικής σημασίας.

Ειδικά για τον χώρο των κοινωνικών επιστημών ο Van Der Leeuw τονίζει ότι η αξία της φορμαλιστικής μοντελοποίησης έχει πολλαπλά οφέλη. Ένα πλεονέκτημα των μοντέλων είναι ότι επιτρέπουν στους ερευνητές να περιγράψουν με οικονομία ένα ευρύ φάσμα σχέσεων εφαρμόζοντας τον βαθμό ακρίβειας που συνήθως δεν επιτυγχάνεται από τα μοναδικά εναλλακτικά εργαλεία που έχουμε για να τις περιγράψουμε: δηλαδή τις φυσικές γλώσσες. Ένα ακόμη πλεονέκτημα των φορμαλιστικών μοντέλων είναι ότι οι τομείς εφαρμογής τους είναι απεριόριστοι. Εξίσου σημαντικό στο πλαίσιο των κοινωνικών επιστημών είναι το γεγονός ότι τα φορμαλιστικά μοντέλα δε σχηματίζονται στην ίδια γλώσσα η οποία χρησιμοποιείται για να περιγράψει τα φαινόμενα που μοντελοποιούνται. Αυτό έχει πολλαπλά οφέλη, το πιο σημαντικό από τα οποία είναι πιθανότατα ότι μας επιτρέπει να αφαιρετικοποιήσουμε και να τονίσουμε χαρακτηριστικά τα οποία είναι σχετικά κατά την άποψή μας. Επιπλέον, συγκεκριμένοι τύποι φορμαλιστικών μοντέλων είναι ικανοί να περιγράψουν με μεγάλη ακρίβεια και οικονομία χώρου τις αλλαγές που συμβαίνουν σε περίπλοκα σύνολα σχέσεων. Εξαιτίας αυτών των ιδιοτήτων, η μοντελοποίηση είναι πλέον κατάλληλη για τον φορμαλισμό δυναμικών θεωριών για συγκεκριμένα φαινόμενα, τα οποία μπορούν στη συνέχεια να συγκριθούν με τις δικές μας παρατηρήσεις (Van Der Leeuw, 2004). Και πρέπει να σημειωθεί πως το να σκέφτεται κάποιος ως επιστήμονας της πληροφορικής σημαίνει πολλά περισσότερα από το να προγραμματίζει έναν υπολογιστή. Κυρίως απαιτεί να μπορεί να σκεφτεί σε πολλαπλά επίπεδα αφαίρεσης (Wing, 2006, σελ. 34).

Το ότι η υπολογιστική λογική, ενδεχομένως, είναι εντελώς ξένη από τον πολιτισμό που μελετά ο ανθρωπολόγος, σε τίποτα δεν επηρεάζει την προτεινόμενη μέθοδο. Και η γραφή μπορεί να είναι απολύτως ξένη για κάποιους πολιτισμούς όμως αυτή χρησιμοποιεί η σύγχρονη ανθρωπολογία για να τους περιγράψει. Στην ανθρωπολογία

ούτε η έρευνα, ούτε η αποτύπωσή της γίνεται με τα μέσα που διαθέτουν οι “άλλοι”, αλλά με αυτά που διαθέτει η/ο ανθρωπολόγος. Άλλωστε, η επιστημονική έρευνα στο σύνολό της (η ανθρωπολογική δεν αποτελεί εξαίρεση) είναι, σύμφωνα με τον έγκριτο επιστημολόγο Thomas Kuhn, «a strenuous and devoted attempt to force nature into the conceptual boxes supplied by professional education» (T. S. Kuhn, 1970, σελ. 5). Εδώ λοιπόν παρουσιάζεται η προοπτική της αλλαγής παραδείγματος (paradigm) που εισάγει η αναπαράσταση και διαχείριση της ανθρωπολογικής γνώσης με νέα μέσα. Όμως η όποια αλλαγή παραδείγματος δεν επιβάλλεται αλλά συντελείται εντός του επιστημονικού κλάδου από τους επιστήμονες που τον υπηρετούν. Άρα προϋπόθεση για να επιτευχθεί είναι η ουσιαστική ενασχόληση των ανθρωπολόγων με αυτή την οπτική.

## 7.1 Προαπαιτούμενα / προϋποθέσεις για την αναπαράσταση της ανθρωπολογικής γνώσης στο σημασιολογικό ιστό

Όπως προαναφέρθηκε, η κωδικοποίηση της ανθρωπολογικής γνώσης για το σημασιολογικό ιστό προφανώς δεν είναι εργασία ενός ανθρώπου ή έστω μια μικρής ομάδας. Επίσης δεν είναι μια εργασία που αφορά τους πληροφορικούς. Τουλάχιστον όχι μόνο σε αυτούς. Απαιτεί σύμπραξη των δύο τομέων. Για τη σχέση των δικτύων -και της μαθηματικής θεωρίας των γράφων- με την Ανθρωπολογία η συζήτηση έχει αρχίσει από πολύ παλιά. Ήδη από τη δεκαετία του 1970 ο Wolfe επισημαίνει την άνθιση του network thinking στην ανθρωπολογία συνδέοντάς την με τις τεχνολογίες επεξεργασίας δεδομένων (Wolfe, 1978). Και παρότι η πρότασή του προφανώς δεν βασίζεται -πώς θα μπορούσε άλλωστε;- στην RDF, η βασική ιδέα παραμένει η ίδια. Αλλά από τότε δεν συντελέστηκε πολύ μεγάλη πρόοδος. Όμως πλέον έχουν αλλάξει δύο βασικοί παράγοντες: πρώτον, η πληροφορική είναι ενσωματωμένη στη ζωή μας και οι γνώσεις γύρω από τους υπολογιστές αφορούν σχεδόν στο σύνολο του δυτικού πληθυσμού. Δεύτερον, και ίσως πιο σημαντικό, η ενασχόληση με ζητήματα που άπτονται της πληροφορικής μπορεί να γίνει σε πολύ υψηλό επίπεδο αφαίρεσης. Δεν πρόκειται για μια τεχνική διαδικασία, και, πολλές φορές, η όποια κωδικοποίηση μπορεί να μην απαιτεί καν τη γνώση παραγωγής υπολογιστικού κώδικα αλλά μόνο λογικής. Δηλαδή το κέντρο ενδιαφέροντος δεν είναι τεχνικά ζητήματα αλλά ζητήματα εννοιολόγησης (conceptualization).

Άλλωστε, για να κωδικοποιηθούν ανθρωπολογικά δεδομένα, όπως πολύ σωστά το θέτει ο Fischer, δεν χρειαζόμαστε αμιγώς ανθρωπολογικές υπολογιστικές μεθόδους διότι οι μέθοδοι είναι εγγενώς ανεξάρτητες από επιστημονικούς τομείς (M. D. Fischer, 1994, σελ. 1-2). Άρα, δεν απαιτείται η ανακάλυψη από τους ανθρωπολόγους μιας νέας μεθόδου. Βεβαίως όμως απαιτείται «computational thinking» από τους ανθρωπολόγους αλλά με τη λογική που το θέτει η Wing: δηλαδή όχι ως προγραμματισμό αλλά ως εννοιολόγηση, όχι ως έναν τρόπο που σκέφτονται οι μηχανές, αλλά ως έναν τρόπο που σκέφτονται οι άνθρωποι, όχι ως μέθοδο μετρήσεων αλλά για το πώς οι άνθρωποι χρησιμοποιούν τις ιδέες τους για να προσεγγίσουν και να λύσουν προβλήματα (Wing, 2006).

Έτσι αν, για παράδειγμα, θέλουμε να επεκτείνουμε το γράφο που εμφανίζεται στο Σχήμα 3 τότε μπορούμε να πούμε ότι ο Μαλινόφσκι μελέτησε τους Trobriands. Ποια είναι η σχέση του “μελέτησε”, πού και ποιος την ορίζει; Ποια η σχέση του διεξήγαμε “επιτόπια έρευνα” και πού και ποιος την ορίζει; Είναι οι Trobriands αυτοί που ορίζει ο ακόλουθος σύνδεσμος <https://ehrafworldcultures.yale.edu/cultures/ol06/description>; Αυτά είναι ερωτήματα που μόνον οι ανθρωπολόγοι μπορούν, και πρέπει, να απαντήσουν.

Για να τα απαντήσουν, όμως, θα πρέπει πρώτα να μπορούν να απαντήσουν σε θέματα που αφορούν ένα πολυεπίπεδο πρόβλημα το οποίο ξεκινάει από (η καταλήγει σε - ανάλογα από το σημείο εκκίνησης) μια πιο φιλοσοφική θεώρηση και σε καμία περίπτωση δεν μπορεί να χαρακτηριστεί ως «τεχνικό». Γενικά η οντολογική αναπαράσταση γνώσης μπορεί να ταξινομηθεί σε τρεις βασικές κατηγορίες. Η γενικότερη κατηγορία αφορά στο επίπεδο της θεμελιώδους οντολογίας (foundational ontology ή upper ontology). Το επίπεδό αυτό είναι αρκετά πιο κοντά στη φιλοσοφία αφού ορίζει τι θα μπορούσε να είναι αντικείμενο, τι κατηγορία κοκ. Δεν περιγράφει, κατ' ανάγκη, συγκεκριμένες κλάσεις και σχέσεις. Εδώ υπάρχουν δύο βασικές θεωρήσεις. Η μία θεώρηση εκφράζεται μέσω της λογικής που ακολουθεί η Basic Formal Ontology (BFO), η οποία είναι περισσότερο συμβατή με τη θεώρηση των φυσικών επιστημών, διότι προϋποθέτει την ύπαρξη μιας αντικειμενικής πραγματικότητας (Arg κ.ά., 2015; Smith, 2015). Η άλλη προσέγγιση, όπως η DOLCE, παρέχει μια εννοιολογική βάση για την περιγραφή και κατανόηση της ανθρώπινης εμπειρίας του κόσμου, περιγράφοντας το πώς οι άνθρωποι αντιλαμβάνονται και μιλούν για τον κόσμο (Gangemi κ.ά., 2002). Η δεύτερη κατηγορία οντολογιών, αφορά στις οντολογίες πυρήνα (core ontologies). Το επίπεδο αυτό ορίζει έναν ελάχιστο αριθμό κλάσεων και των μεταξύ τους σχέσεων, οι οποίες είναι απαραίτητες για την αναπαράσταση της γνώσης μέσα στο συγκεκριμένο πλαίσιο. Η τρίτη κατηγορία αφορά στις οντολογίες τομέα (domain ontologies) οι οποίες είναι εξειδικευμένες πάνω σε πολύ συγκεκριμένα θέματα. Άρα στο μοντέλο που προτείνεται στο κεφάλαιο 6, το πρώτο επίπεδο αφορά σε μια core ontology ενώ το δεύτερο επίπεδο (εκείνο της ολοκλήρωσης) αφορά σε μια domain ontology. Ο βαθμός συμμόρφωσης αυτών στο επίπεδο της upper ontology με τη μία ή την άλλη θεώρηση είναι θέμα που καλούνται να απαντήσουν οι ίδιοι οι ανθρωπολόγοι.

## 7.2 Οφέλη για την ανθρωπολογία

Τα οφέλη μιας τέτοιας προσέγγισης δεν θα μπορούσαν να αποτελέσουν μια πεπερασμένη λίστα, γνωστή εκ των προτέρων, και αυτό διότι πρόκειται για μια διαδικασία που από τη φύση της είναι προστιθέμενης αξίας. Παρόλα αυτά στη συνέχεια θα γίνει απαρίθμηση - σαφώς αποσπασματική - κάποιων πλεονεκτημάτων.

Γενικά για την επιστήμη, οι φορμαλιστικές μέθοδοι επιτρέπουν την ακρίβεια και τη σαφήνεια, διαθέτουν μεγαλύτερες δυνατότητες επαλήθευσης και αναπαραγωγιμότητας (reproducibility) των συμπερασμάτων μέσω διαφανούς συλλογιστικής, δίνουν τη δυνατότητα ευκολότερου εντοπισμού λογικών σφαλμάτων, συμβάλουν στον σαφέστερο ορισμό των θεωρητικών προβλημάτων, συστηματοποιούν τη γνώση και αποτελούν προαπαιτούμενο για την διεπιστημονικότητα. Άρα όλα τα παραπάνω έχουν εφαρμογή και στην κοινωνική ανθρωπολογία. Περιοριζόμενοι όμως στις φορμαλιστικές μεθόδους

με τις οποίες ασχολήθηκε η παρούσα διατριβή θα μπορούσαμε να εστιάσουμε σε πιο συγκεκριμένα οφέλη.

Το πρώτο, περισσότερο προφανές και πιο άμεσο, πλεονέκτημα είναι να οργανωθεί αποτελεσματικότερα η έρευνα, για παράδειγμα, μέσω της καταγραφής ηλεκτρονικών σημειώσεων πεδίου (field notes). Δεν είναι το ίδιο απλά να τα γράφεις σε ένα κείμενο διότι σου δίνει τη δυνατότητα επισημειώσεων, αναζήτησης, distant reading κτλ όπως αναλύθηκε διεξοδικά στο κεφάλαιο 4. Επίσης παρέχεται η δυνατότητα ύπαρξης διεύρυνσης των μεθόδων που χρησιμοποιεί η ανθρωπολογία. Παράλληλα οι εν λόγω μέθοδοι επιτρέπουν πειστικότερη τεκμηρίωση επί των αποτελεσμάτων αφού η επεξεργασία των δεδομένων γίνεται και μέσω συστημάτων τα οποία δεν έχουν “λόγους” να είναι μεροληπτικά. Σε καμία περίπτωση δεν ισχυρίζομαι ότι οι δημιουργοί των συστημάτων, τα στοιχεία που εισήχθησαν σε αυτά και οι τρόποι επεξεργασίας τους δεν είναι μεροληπτικοί. Διότι αυτό είναι κάτι που συμβαίνει πολύ συχνά και ως εκ τούτου τα αποτελέσματα μπορεί να είναι απολύτως μεροληπτικά απλά ντυμένα με το μανδύα μιας επιστημονικοφανούς απόδειξης. Όμως ακριβώς η ενασχόληση των ανθρωπολόγων με αυτές τις μεθόδους θα βοηθήσει στην ουσιαστική διαφάνεια (transparency) της εκάστοτε υπολογιστικής επεξεργασίας.

Ένα, επίσης προφανές και άμεσο, πλεονέκτημα σχετίζεται με την ανακάλυψη πόρων (resource discovery). Είναι σχεδόν αυτονόητο ότι, με τη χρήση υπολογιστικών μεθόδων, όχι μόνο αυξάνεται το υλικό εθνογραφικού ενδιαφέροντος που μπορεί να είναι διαθέσιμο στον ιστό αλλά αυξάνονται εκθετικά και οι δυνατότητες αναζήτησης και πλοήγησης σε αυτό το υλικό. Είτε πρόκειται για πόρους που περιέχουν ανθρωπολογικές πληροφορίες (επιστημονικές μελέτες, εικόνες, βίντεο κτλ) είτε για σύνολα δεδομένων (data sets) που θα μπορούσαν να χρησιμοποιηθούν στην ανθρωπολογική έρευνα. Ένας δυνητικά άπειρος αριθμός κριτηρίων μπορεί να φιλτράρει το σχετικό υλικό και να παρέχει σημαντική πληροφόρηση στον χρήστη του συστήματος.

Άμεση συνέπεια της παραπάνω διαδικασίας είναι η ενεργοποίηση επιπλέον δυνατοτήτων συγκριτικής ανθρωπολογίας. Αφού πλέον η πληροφορία δεν είναι διαθέσιμη μόνο ως κείμενο αλλά και ως νόημα, η διαδικασία επιλογής συγκεκριμένων πλευρών της ανθρώπινης δραστηριότητας και η κατ’ αντιπαράσταση εξέτασή τους μπορεί σε μεγάλο βαθμό να γίνει από τον υπολογιστή βάσει των εξεζητημένων κριτηρίων που μπορεί να θέσει κάθε ερευνητής, επειδή η αποτύπωση της γνώσης θα διαθέτει τον αθροιστικό χαρακτήρα ενός ενιαίου αρχείου. Ο Fischer τονίζει πως «η γνώση είναι η παρακαταθήκη της ανθρωπολογίας [...]. Η ανθρωπολογία είναι η περιγραφή και η αναπαράσταση της γνώσης που κατέχουν οι ανθρώπινες ομάδες» (M. D. Fischer, 1994, σελ. 6). Με την περιγραφή της με όρους σημασιολογικού ιστού μπορούμε να αποτυπώσουμε αυτή τη γνώση σε ένα περιβάλλον το οποίο δεν εξαρτάται άμεσα από κάποια φυσική γλώσσα (language agnostic) αλλά είναι εννοιοκεντρικό.

Αυτό το language agnostic περιβάλλον δίνει τη δυνατότητα επικοινωνίας της ανθρωπολογικής γνώσης με άλλους επιστημονικούς τομείς αφού μια τέτοια προσέγγιση λαμβάνει υπόψη τη σημασιολογική διαλειτουργικότητα. Για παράδειγμα, στη σύγχρονη Δύση υπάρχει μια βαθιά ριζωμένη η πεποιθήση, όχι μόνο στο ευρύ κοινό αλλά και σε μέρος της επιστημονικής κοινότητας, περί μιας μονόδρομης εξελικτικής πορείας η οποία

τοποθετεί το σύγχρονο δυτικό άνθρωπο -και τον πολιτισμό του- στην κορωνίδα της εξέλιξης. Η ανθρωπολογία έχει ασκήσει πολύ έντονη κριτική σε αυτή την προσέγγιση και έχει προσφέρει πλήθος εμπειρικών δεδομένων που την αμφισβητούν. Στο κεφάλαιο 4 αναφερθήκαμε στη McKinnon, η οποία δίνει πλήθος εθνογραφικών παραδειγμάτων τα οποία έρχονται σε ευθεία αντίθεση με πολλές από τις διαπιστώσεις σύγχρονης εξελικτικής ψυχολογίας (McKinnon, 2009). Άλλο ένα εμβληματικό έργο που κινείται στο ίδιο μήκος κύματος, αμφισβητώντας αυτές τις θεωρίες, είναι το «The Dawn of Everything: A New History of Humanity» των Graeber και Wengrow. Μέσω αυτού οι συγγραφείς του καλούν στην επανεξέταση βασικών εννοιών (όπως η ελευθερία της μετακίνησης, η ελευθερία της επανοργάνωσης των κοινωνικών σχέσεων, η μη υποταγή σε διαταγές) αλλά και την εξέταση θεμελιωδών αρχών που σχετίζονται με την αμφιβολία για μια μοναδική προέλευση του κράτους ή της εξουσίας, την απομάκρυνση από την αντίληψη ότι η ιστορία έχει ως “στόχο” ή προορισμό το σύγχρονο κράτος και ότι οι άνθρωποι σε διάφορους πολιτισμούς δεν υπήρξαν παθητικοί αποδέκτες ενός μοιραίου εξελικτικού δρόμου, αλλά πειραματίστηκαν με διαφορετικές μορφές πολιτικής οργάνωσης (Graeber & Wengrow, 2022). Η αναπαράσταση γνώσης επιτρέπει βαθύτερη κατανόηση, καλύτερο διαμοιρασμό και ένταξη των παραπάνω τοποθετήσεων σε ένα ολοκληρωμένο (integrated) πληροφοριακό σύστημα.

Το παραπάνω πληροφοριακό σύστημα είναι ιδιαίτερος χρήσιμο στο διεπιστημονικό περιβάλλον. Διότι ο τύπος του φορμαλισμού που διαθέτει το καθιστά επαρκώς αφαιρετικό και παράλληλα επαρκώς λεπτομερές. Ταυτόχρονα μπορεί να είναι αρκετά “ρεαλιστικό” ώστε να υποχρεώνει επιστήμονες που προέρχονται από διαφορετικές επιστήμες να εστιάσουν στα ίδια ζητήματα (Van Der Leeuw, 2004, σελ. 123), αυξάνοντας τα επίπεδα της διυποκειμενικότητας (intersubjectivity) όχι μόνο μεταξύ ατόμων αλλά και μεταξύ επιστημονικών τομέων, ενεργοποιώντας μια διυποκειμενική συμφωνία (intersubjective agreement). Έτσι είναι εφικτό να δοθούν απαντήσεις σε ερωτήματα τα οποία δεν έχουν καν τεθεί, διότι οι προσεγγίσεις που διαθέτει η σημερινή θεώρηση δεν μπορούν καν να τα διατυπώσουν.

Επίσης, οι υπολογιστικές μέθοδοι κωδικοποίησης και επεξεργασίας της πληροφορίας παρέχουν εξαιρετική ευελιξία στην παραγωγή πολλών και διαφορετικών οπτικοποιήσεων. Τέτοιου είδους οπτικοποιήσεις είναι -σε αρκετές περιπτώσεις- ιδιαίτερα βοηθητικές. Ήδη στο κεφάλαιο 4 αναφερθήκαμε διεξοδικά στη σημασία των διαγραμμάτων κατά την αποτύπωση των συγγενικών σχέσεων ώστε να μην επιβάλει ο παρατηρητής το δικό του μοντέλο στους παρατηρούμενους. Αλλά και ενδεικτικά παραδείγματα οπτικοποίησης περιλαμβάνονται στα σχήματα και τα γραφήματα της παρούσας διατριβής. Για παράδειγμα τα όσα εμφανίζονται στο Γράφημα 4, και απεικονίζουν το βιβλιογραφικό δίκτυο που σχηματίζουν συγκεκριμένοι όροι στην επιστημονική βιβλιογραφία, πόσο εύκολα θα μπορούσαν να περιγραφούν με λέξεις;

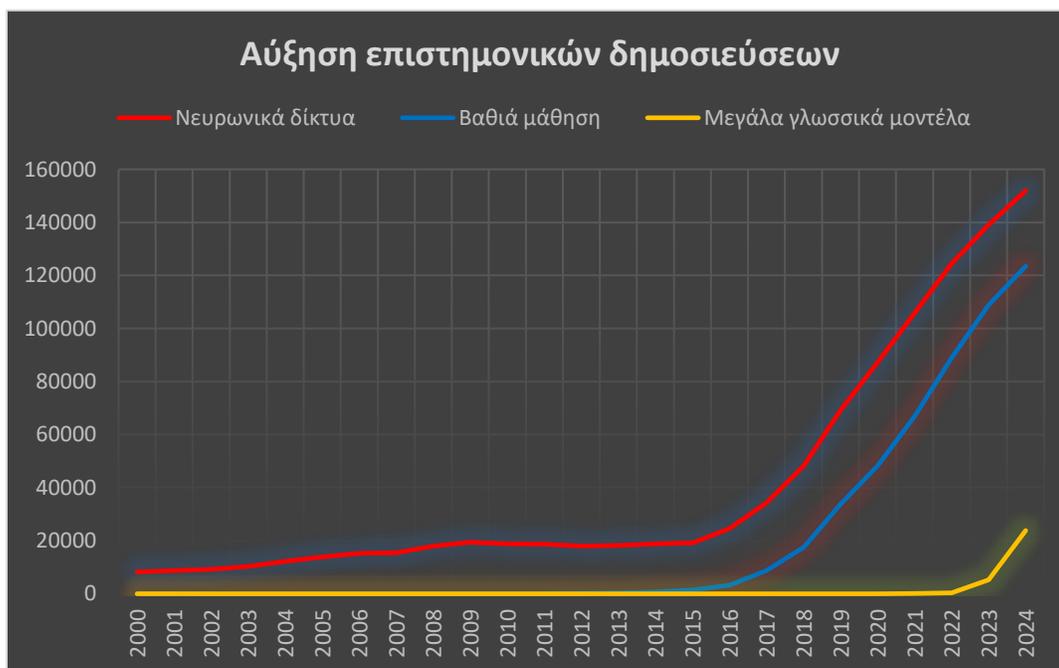
Ένα παράπλευρο, αλλά πολύ σημαντικό όφελος, είναι ότι ο φορμαλισμός που απαιτεί η υπολογιστική αναπαράσταση συμβάλει στην αύξηση της ακρίβειας τόσο των περιγραφών όσο και των συμπερασμάτων μας. Άρα δεν δημιουργείται απλά μια δυαδική αντίστιξη άνθρωπος-υπολογιστής αλλά η μοντελοποίηση για τον υπολογιστή αλλάζει και την οπτική διαχείρισης εκτός του υπολογιστή. Επίσης, μέσω των δυνατοτήτων

συλλογισμού (reasoning) εντός του υπολογιστικού συστήματος, παρέχονται δυνατότητες ακόμη και ελέγχου της ορθότητας των αποφάσεων των επιστημόνων. Και εδώ τίθεται ένα πολύ σημαντικό θέμα για το τι σημαίνει τεχνητή νοημοσύνη και ποιες είναι οι βασικές όψεις της.

### 7.3 Generative AI, Large Language Models, Deep Learning και μορφές Τεχνητής Νοημοσύνης

Με την έναρξη της τρίτης δεκαετίας του 21<sup>ου</sup> αιώνα η εφαρμογή, σε ευρεία κλίμακα, των νευρωνικών δικτύων απογειώνει τις δυνατότητες της υπολογιστικής επεξεργασίας σε δύο επίπεδα. Πρώτον, στο επίπεδο αύξησης της υπολογιστικής ισχύος και, δεύτερον, στον τρόπο της επεξεργασίας της πληροφορίας, διότι τα νευρωνικά δίκτυα επέτρεψαν αυτό που ονομάζεται «βαθιά μάθηση» (deep learning).

Όπως φαίνεται στο Γράφημα 6 ο όρος «νευρωνικά δίκτυα» (neural networks) απασχολεί τους επιστήμονες σε όλη τη διάρκεια του 21<sup>ου</sup> αιώνα αλλά η εκτίναξη της ενασχόλησης της επιστημονικής κοινότητας έρχεται μαζί με την αύξηση του ενδιαφέροντος για την εφαρμογή των νευρωνικών δικτύων στην ανάπτυξη αυτού που ονομάζεται «βαθιά μάθηση» (deep learning).



**Γράφημα 6:** επιστημονικές δημοσιεύσεις για νευρωνικά δίκτυα, βαθιά μάθηση και μεγάλα γλωσσικά μοντέλα (πηγή δεδομένων Scopus)

Όμως η έκρηξη των παραπάνω τεχνολογιών συνοδεύεται με τη χρήση τους στα μεγάλα γλωσσικά μοντέλα (Large Language Models), περισσότερο γνωστά με το ακρωνύμιο LLMs. Ο όρος Large Language Models εμφανίζεται στην επιστημονική βιβλιογραφία το 2020, με μόλις 14 δημοσιεύσεις να τον αναφέρουν την συγκεκριμένη

χρονιά. Ενώ τέσσερα χρόνια αργότερα, εντός του έτους 2024, εμφανίζεται σε περισσότερες 22.500 επιστημονικές δημοσιεύσεις<sup>54</sup>.

Επειδή ένα μέρος της ανά χειράς διατριβής εντάσσεται στο πεδίο της τεχνητής νοημοσύνης, κρίνεται σκόπιμο να αποσαφηνιστούν μερικά ζητήματα που αφορούν τη σχετική ορολογία και συχνά οδηγούν σε παρανοήσεις. Στην καθημερινή γλώσσα του τέλους του πρώτου τετάρτου του εικοστού πρώτου αιώνα, ο όρος τεχνητή νοημοσύνη είναι συνυφασμένος με εργαλεία τύπου ChatGPT, DeepSeek τα οποία βασίζονται στη μηχανική μάθηση (machine learning) και χρησιμοποιούν νευρωνικά δίκτυα. Τα παραπάνω εντάσσονται σε μια γενικότερη κατηγορία που ονομάζεται Generative AI, η οποία αφορά μοντέλα που έχουν την ικανότητα να δημιουργήσουν νέο περιεχόμενο όπως κείμενο, εικόνες, ήχο, κλπ. Για την ακρίβεια τα LLMs που βασίζονται στην παραπάνω τεχνολογία χρησιμοποιούν ένα συγκεκριμένο τύπο που ονομάζεται μετασχηματιστής (transformer) που είναι ειδικά σχεδιασμένος για την επεξεργασία κειμένων. Η λογική αυτή παρουσιάζεται σε μια εργασία της Google η οποία παρουσιάστηκε το 2017 στην μελέτη «Attention is All you Need»<sup>55</sup> (Vaswani κ.ά., 2017).

Τα μοντέλα αυτά είναι επί της ουσίας πιθανοκρατικά (probabilistic) και εκπαιδεύονται με συγκεκριμένα σύνολα δεδομένων. Επειδή είναι πολύ ευέλικτα στη χρήση της φυσικής γλώσσας, δίνουν την αίσθηση ότι καταλήγουν σε λογικά συμπεράσματα, αλλά στην ουσία η «γνώση» τους είναι εμπειρική και όχι φορμαλιστικά ρητή και αποδείξιμη. Αντιθέτως η οντολογική αναπαράσταση γνώσης είναι ένα λογικό μοντέλο το οποίο αποτελείται από φορμαλιστικά ρητή «γνώση» η οποία, χειριζόμενη λογικές δηλώσεις, μπορεί να καταλήξει σε συμπεράσματα τα οποία είναι εξ ορισμού “αληθή”, ελέγχοντας πιθανές ασυνέπειες μεταξύ των δηλώσεων. Η συνέπεια (consistency) της οντολογίας είναι άλλωστε και ο θεμέλιος λίθος για τον ορθό συλλογισμό. Ο τύπος μηχανικής μάθησης που περιγράφηκε παραπάνω και τα μοντέλα που βασίζονται σε αυτόν δεν διαθέτουν τέτοιου είδους δυνατότητες.

Εξειδικεύοντας λίγο περισσότερο, η οντολογική αναπαράσταση βασίζεται σε αυτό που ονομάζεται Assertion Box ενώ η μηχανική μάθηση σε αυτό που ονομάζεται Black Box. Στην περίπτωση του Assertion Box υπάρχουν ρητές δηλώσεις (statements) από τις οποίες μπορεί να προκύψουν, με σαφή και προδιαγεγραμμένο τρόπο, νέες δηλώσεις. Άρα ο τρόπος παραγωγής της γνώσης είναι πλήρως κατανοητός και αποδείξιμος. Αντίθετα, στην περίπτωση του Black Box η εσωτερική του λογική και λειτουργία είναι άορατη ή κρυφή. Κάποιος μπορεί να παρατηρήσει τι απαντάει το σύστημα σε μια ερώτηση, αλλά δεν μπορεί να δει πώς το σύστημα καταλήγει στις αποφάνσεις του.

---

<sup>54</sup> Τα στοιχεία αντλήθηκαν από τη βιβλιογραφική βάση Scopus, με τελευταία ενημέρωση το Μάιο του 2025. Πριν το 2020 ο όρος έχει κάποιες σποραδικές εμφανίσεις, αν και όχι πάντα με το νόημα που έχει αποκτήσει σήμερα.

<sup>55</sup> Ενδεικτικό της επίδρασης που είχε η συγκεκριμένη δημοσίευση στην επιστημονική κοινότητα είναι ο αριθμός των αναφορών που έχει λάβει. Τον Μάιο του 2025 η Scopus της απέδιδε περίπου 84.000 αναφορές ενώ το Google scholar περίπου 180.000.

Άρα, στις περιπτώσεις των black boxes δεν υπάρχει διαφάνεια στον τρόπο με τον οποίο το σύστημα κατέληξε στα συμπεράσματά του, τα οποία προέκυψαν μέσω μιας στατιστικής επεξεργασίας δεδομένων με τα οποία έχει ήδη εκπαιδευτεί. Αντιθέτως στην οντολογική αναπαράσταση -και το σχετικό συμπέρασμο- ο τρόπος παραγωγής γνώσης είναι απολύτως σαφής και δεν βασίζεται σε στατιστικές μεθόδους ή πιθανότητες αλλά στην ανθρώπινη γνώση που έχει μοντελοποιηθεί με κατάλληλο τρόπο ώστε να είναι επεξεργάσιμη από τον υπολογιστή με τρόπο που οδηγεί σε αληθή συμπεράσματα.

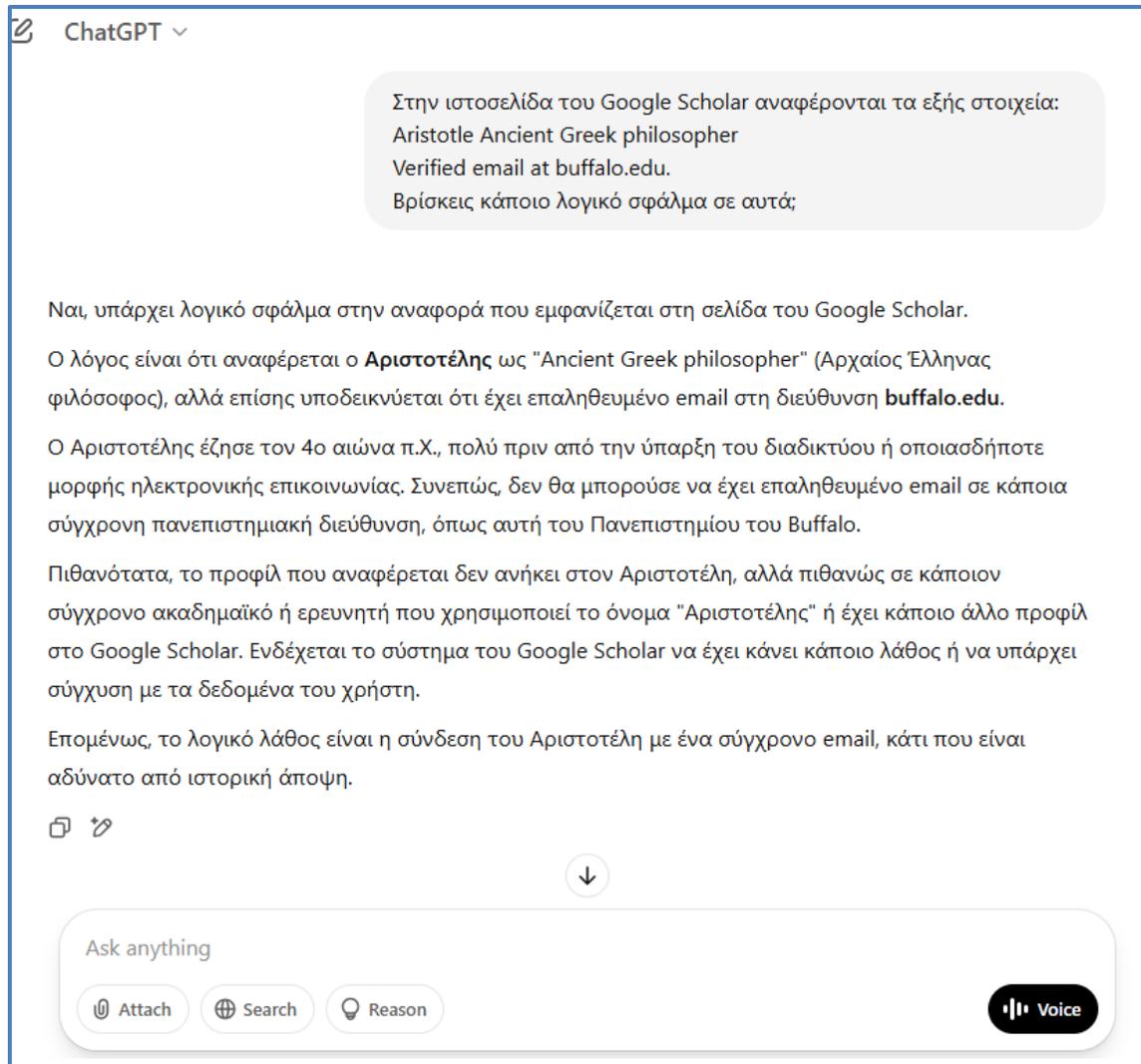
Παρουσιάζονται τρία παραδείγματα από το ChatGPT<sup>56</sup> τα οποία είναι ενδεικτικά του τρόπου με τον οποίο λειτουργεί το μοντέλο. Το πρώτο παράδειγμα (Εικόνα 1) δείχνει την υψηλή ικανότητα ‘συλλογισμού’ που διαθέτει, αφού το μοντέλο μπορεί να ‘διακρίνει’ ότι ο Αριστοτέλης, ως ένας αρχαίος Έλληνας, δεν είναι δυνατόν να διαθέτει email.

Ενώ στη δεύτερη περίπτωση (Εικόνα 2), που του ζητείται να απαριθμήσει τις χώρες που αρχίζουν από το γράμμα Β -γραμμένο στα ελληνικά- αποτυγχάνει παταγωδώς! Δεν διακρίνει το ελληνικό από το λατινικό αλφάβητο ενώ ταυτόχρονα μπλέκει το λατινικό γράμμα V με το λατινικό B. Επίσης, παρότι επιστρέφει ως χώρα το Βιετνάμ (που στα λατινικά αρχίζει από V και στα ελληνικά από Β), δεν επιστρέφει τη Βενεζουέλα (η οποία επίσης αρχίζει στα λατινικά από V και στα ελληνικά από Β), αλλά ούτε το Βατικανό ή τα νησιά Βανουάτου. Αντίστοιχα, ενώ επιστρέφει το Μπαγκλαντές (αρχίζει από λατινικό B αλλά από ελληνικό M) δεν επιστρέφει το Μπουτάν που ανήκει ακριβώς στην ίδια κατηγορία. Άρα η συνέπεια με την οποία αντιμετώπισε τα ονόματα των χωρών είναι ανύπαρκτη. Από τη μία λοιπόν ‘δυσκολεύεται με την αλφαβήτα’ αλλά, από την άλλη, μπορεί να εντοπίσει λογικό σφάλμα στα όσα αναφέρονται στη σελίδα του Google Scholar για τον [Αριστοτέλη](#) κάνοντας αντιπαραβολές της ιστορικής περιόδου που έζησε ένας άνθρωπος με τις τεχνολογικές υποδομές που ήταν διαθέσιμες σε αυτήν!<sup>57</sup>

---

<sup>56</sup> Οι ερωτήσεις έγιναν, τελευταία φορά, στις αρχές Μαΐου του 2025 στην ελεύθερη έκδοση του ChatGPT, χωρίς να είναι συνδεδεμένος κάποιος χρήστης. Είναι πολύ πιθανόν, στο μέλλον, οι ίδιες ερωτήσεις να οδηγήσουν σε άλλες απαντήσεις αφού το μοντέλο διαρκώς ‘εκπαιδεύεται’.

<sup>57</sup> Σε παλαιότερες εκδοχές του μοντέλου ήταν συχνά ακόμη και τα λάθη σε απλές μαθηματικές πράξεις. Ο αναγνώστης, με μια απλή αναζήτηση στο διαδίκτυο, μπορεί να βρει αρκετά παραδείγματα.



**Εικόνα 1:** το ChatGPT για τη σελίδα του Αριστοτέλη στο Google Scholar

ChatGPT Log in

Πόσες είναι οι χώρες που ξεκινάνε από Β?

Υπάρχουν 10 χώρες του κόσμου που ξεκινούν με το γράμμα "B". Αυτές είναι:

1. Βαχρέιν (Bahrain)
2. Μπανγκλαντές (Bangladesh)
3. Μπαρμπάντος (Barbados)
4. Βέλγιο (Belgium)
5. Μπελίζ (Belize)
6. Μποτσουάνα (Botswana)
7. Βουλγαρία (Bulgaria)
8. Μπουρούντι (Burundi)
9. Βολιβία (Bolivia)
10. Βιετνάμ (Vietnam)

Αν έχεις κι άλλες ερωτήσεις, ευχαρίστως να σε βοηθήσω!

Get smarter responses, upload files and images, and more. Log in Sign up for free

Ask anything

Attach Search Reason Voice

ChatGPT can make mistakes. Check important info. See [Cookie Preferences](#).

**Εικόνα 2:** το ChatGPT για χώρες που ξεκινούν από ελληνικό Β

Το επόμενο παράδειγμα, το οποίο εμφανίζεται στην Εικόνα 3, καταδεικνύει με πιο ξεκάθαρο τρόπο ότι δεν υπάρχει διαδικασία συλλογισμού με την έννοια του ελέγχου της συνέπειας των αποτελεσμάτων. Η ερώτηση που υποβάλλεται στο ChatGPT είναι να φτιάξει ένα παράδειγμα κώδικα μιας ιεραρχίας η οποία θα βασίζεται σε ένα στοιχείο από το χώρο ονομάτων (namespace) της RDF. Το θέμα είναι ότι το στοιχείο αυτό δεν υπάρχει στο χώρο ονομάτων της RDF, διότι το «subClassOf» ορίζεται από το RDFS και ουχί από το RDF. Άρα το στοιχείο «rdf:subClassOf» που προτείνει το ChatGPT είναι απλά ανύπαρκτο. Η ειδοποιός διαφορά σε αυτόν τον τύπο λάθους έγκειται στο γεγονός ότι οποιοδήποτε πληροφοριακό σύστημα διαχειρίζεται δεδομένα σημασιολογικού ιστού

επιχειρήσει να επεξεργαστεί τον εν λόγω κώδικα θα 'χτυπήσει' αμέσως διότι θα εντοπίσει ότι αυτό που του δίνεται ως input δεν υπάρχει στο σχήμα αναφοράς.

ChatGPT ▾

Μπορείς να μου φτιάξεις ένα παράδειγμα ιεραρχίας σε XML, χρησιμοποιώντας το `rdf:subClassOf` ?

Βεβαίως! Το `rdf:subClassOf` είναι ένα χαρακτηριστικό που χρησιμοποιείται στο RDF (Resource Description Framework) για να δηλώσει την ιεραρχία ή τις κληρονομίες μεταξύ κλάσεων (classes). Δηλαδή, λέει ότι μια κλάση είναι υποκλάση μιας άλλης.

Ακολουθεί ένα παράδειγμα ιεραρχίας χρησιμοποιώντας το `rdf:subClassOf` σε XML:

```

xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ex="http://example.org/">

  <!-- Ορισμός της βασικής κλάσης "Ζώα" -->
  <rdf:Description rdf:about="http://example.org/Animal">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  </rdf:Description>

  <!-- Ορισμός της υποκλάσης "Θηλαστικά", που είναι υποκλάση των "Ζώων" -->
  <rdf:Description rdf:about="http://example.org/Mammal">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdf:subClassOf rdf:resource="http://example.org/Animal"/>
  </rdf:Description>

  <!-- Ορισμός της υποκλάσης "Σκύλος", που είναι υποκλάση των "Θηλαστικών" -->
  <rdf:Description rdf:about="http://example.org/Dog">

```

**Εικόνα 3:** το ChatGPT προτείνει τη χρήση ανύπαρκτου στοιχείου του χώρου ονομάτων (namespace) της RDF

Προς αποφυγή παρεξηγήσεων, ο συγγραφέας της διατριβής, σε καμία περίπτωση, δεν προσπαθεί να υποτιμήσει ούτε το ChatGPT ούτε -πολύ περισσότερο- αυτού του είδους τις τεχνολογίες. Αντιθέτως, θεωρεί ότι βρισκόμαστε σε μια εποχή όπου οι εν λόγω τεχνολογίες θα λειτουργήσουν καταλυτικά σε πάρα πολλούς τομείς της ανθρώπινης δραστηριότητας και ειδικά της επιστημονικής μεθοδολογίας και πρακτικής. Άλλωστε το ChatGPT, αν δεν ήταν προγραμματισμένο να δηλώνει ότι είναι μηχανή, θα περνούσε άνετα το τεστ του Άλαν Τούρινγκ (Turing, 1950) δίνοντας τη βεβαιότητα ότι επικοινωνούμε με έναν ιδιαίτερα ευφυή -και σχεδόν παντογνώστη- άνθρωπο. Διότι αν

επανερχόμεσταν στην ερώτηση με τις χώρες και διευκρινίζαμε ότι με το B εννοούμε μόνο το ελληνικό B, τότε θα επαναδιατύπωνε την απάντηση. Αυτό βέβαια μάς οδηγεί, για άλλη μια φορά, στην ασάφεια της φυσικής γλώσσας.

Όμως, από τα παραπάνω παραδείγματα, είναι προφανές ότι τα LLMs δεν λειτουργούν ακολουθώντας τη συνέπεια (consistency) και τη συνοχή που διαθέτει ο συλλογισμός στην οντολογική αναπαράσταση γνώσης, παρότι μερικές φορές δίνουν την αίσθηση ότι το πράττουν. Διότι η μηχανική μάθηση επιτρέπει σε ένα υπολογιστικό σύστημα να προσομοιώσει μια πραγματικότητα και να αντιδράσει σε αυτήν όπως αντιδρούν και τα υπόλοιπα υποκείμενα που έρχονται σε επαφή με αυτήν. Αντίθετα η οντολογική αναπαράσταση ορίζει από πριν, με φορμαλιστικό και απόλυτο τρόπο, το τι πρέπει να κάνει ένα υπολογιστικό σύστημα. Αν πάρουμε ως παράδειγμα την αυτόματη οδήγηση τότε μηχανική μάθηση αποτελεί το να βγει στο δρόμο ένα αυτοκίνητο ώστε να 'δει' και να 'μάθει' πώς οδηγούν τα υπόλοιπα αυτοκίνητα και στη συνέχεια να κάνει το ίδιο. Δηλαδή με εμπειρικό τρόπο θα μάθει να διαντιδρά με τον περιβάλλον του όπως διαντιδρούν και τα υπόλοιπα υποκείμενα. Αντιθέτως, μια οντολογική αναπαράσταση θα έθετε πρωτίστως του κανόνες του κώδικα οδικής κυκλοφορίας, και, το αυτοκίνητο, θα έπρεπε να συμμορφώνεται με αυτούς.

Θα μπορούσε να ειπωθεί πως, τηρουμένων των αναλογιών, ο τρόπος που λειτουργούν τα μεγάλα γλωσσικά μοντέλα προσομοιάζει στον τρόπο που μαθαίνει να παράγει λόγο ο ανθρώπινος εγκέφαλος, ενώ ο τρόπος που λειτουργούν οι οντολογικές αναπαραστάσεις είναι πιο κοντά στις λογικές που εδράζεται ο επιστημονικός λογισμός. Διότι τα LLMs δεν είναι, κατ' ανάγκη, ορθολογικά αφού βασίζονται στα δεδομένα με τα οποία εκπαιδεύτηκαν. Έτσι, ο επαγωγικός συλλογισμός που χρησιμοποιούν αυτά τα μοντέλα είναι εξαιρετικός για να εντοπίσει μοτίβα τα οποία βοηθούν στον εντοπισμό ενός μερικού/τοπικού νοήματος αλλά δεν διαθέτουν, κατ' ανάγκη, μια ορθολογική προσέγγιση. Αντιθέτως η οντολογική αναπαράσταση γνώσης είναι απολύτως ορθολογική.

Όπως είναι αναμενόμενο οι δύο προσεγγίσεις δεν είναι αλληλοαποκλειόμενες. Αντιθέτως, είναι σοφό να χρησιμοποιηθούν παράλληλα. Το ενδιαφέρον της επιστημονικής κοινότητας για την παράλληλη χρήση οντολογιών και γράφων γνώσης συνδυαστικά με τα LLMs είναι διαρκώς αυξανόμενο. Ένα πρόσφατο, αλλά πολυδιαβασμένο, κείμενο επί του θέματος είναι το «Unifying Large Language Models and Knowledge Graphs: A Roadmap» (Pan κ.ά., 2024). Διότι οι οντολογικές αναπαραστάσεις και οι γράφοι γνώσης, μέσω του αυστηρού μηχανισμού εξαγωγής συμπερασμάτων που ενσωματώνουν, μπορούν να συμβάλουν στην αξιοπιστία των αποτελεσμάτων των LLMs.

Στο παραπάνω πλαίσιο αναπτύσσονται μέθοδοι συνεργασίας όπως η Neuro-symbolic AI. Η συγκεκριμένη μέθοδος στοχεύει στο συνδυασμό της μηχανικής μάθησης με την αναπαράσταση γνώσης ώστε από τη μία να βελτιωθεί η αξιοπιστία των αποτελεσμάτων και από την άλλη η ευελιξία του συστήματος (Bhuyan κ.ά., 2024; Hitzler κ.ά., 2022). Στο ίδιο πλαίσιο κινείται και η μέθοδος RAG (Retrieval Augmented Generation) η οποία συνδυάζει τα γλωσσικά μοντέλα με εξωτερικές βάσεις γνώσης (όπως γράφους γνώσης) από τις οποίες αντλούνται πληροφορίες που συμβάλλουν στην αξιοπιστία των απαντήσεων (Gao κ.ά., 2024).

Από την άλλη πλευρά, τα μεγάλα γλωσσικά μοντέλα, παρότι είναι μαύρα κουτιά, μπορούν να χρησιμοποιηθούν και για τη δημιουργία οντολογιών. Για αυτή τη σχέση έχουν γραφτεί λιγότερα, από όσα για την παράλληλη χρήση LLMs με γράφους γνώσης, αλλά υπάρχει και εκεί ενδιαφέρον. Ενδεικτικά αναφέρεται η πρόσφατη εργασία των Lippolis, Saeedizade κ.α. (Lippolis, Saeedizade, Keskisarkka, κ.ά., 2025; Lippolis, Saeedizade, Keskisärkkä, κ.ά., 2025).

Για την ώρα, αυτό που έχει αναπτυχθεί περισσότερο είναι η σχέση των LLMs με άλλες μορφές Generative AI. Ένα ενδεικτικό παράδειγμα, από το χώρο των ανθρωπιστικών επιστημών, είναι το πρόγραμμα Aeneas. Πρόκειται για νευρωνικό δίκτυο που έχει σκοπό να βοηθήσει ιστορικούς και επιγραφολόγους στην κατανόηση, χρονολόγηση, και αποκατάσταση αρχαίων επιγραφών. Παίρνει ως είσοδο (input) τόσο το κείμενο της επιγραφής όσο και την εικόνα της επιγραφής, ώστε να συνδυάσει την οπτική με την κειμενική πληροφορία (Assael κ.ά., 2025).

## 7.4 Μελλοντική έρευνα

Οι οντολογικές προσεγγίσεις δεν περιέχουν έναν εγγενή μηχανισμό δημιουργίας μιας σύνθετης δήλωσης του τύπου «μην πίνεις αλκοόλ όταν είσαι έγκυος». Δηλαδή η απόδοση του «όταν» είναι αρκετά σύνθετο να μοντελοποιηθεί και δεν έχει έναν και μοναδικό τρόπο μοντελοποίησης. Για παράδειγμα, η παραπάνω πρόταση θα μπορούσε να μοντελοποιηθεί με τη χρήση του «owl:disjointWith» ανάμεσα στην κλάση «έγκυος» και στην κλάση «κατανάλωση αλκοόλ». Για το σημασιολογικό ιστό, μια τέτοια μοντελοποίηση γίνεται αποτελεσματικότερα μέσω της SWRL (Semantic Web Rule Language). Γενικότερα όμως η μοντελοποίηση του μηχανισμού των απαγορεύσεων πρέπει να λάβει υπόψη της διάφορα θέματα και χρήζει μιας πιο ενδεδειγμένης θεωρητικής προσέγγισης. Διότι υπάρχει μια ειδοποιός διαφορά ανάμεσα σε έναν κοινωνικό κανόνα που λέει τι απαγορεύεται και σε έναν φυσικό νόμο. Ο νόμος της βαρύτητας ισχύει και δεν μπορεί να παρακαμφθεί. Ενώ η απαγόρευση της κλοπής επίσης ισχύει αλλά πολλές φορές τα άτομα παραβαίνουν τον κανόνα. Όμως, στα κοινωνικά φαινόμενα έχει ιδιαίτερη σημασία η μοντελοποίηση της μη συμμόρφωσης και οι συνέπειες που αυτή επιφέρει. Εδώ έχει ενδιαφέρον η δεοντική λογική (deontic logic) (Beller, 2001, 2008; De Giorgis κ.ά., 2022; McNamara, 2006; McNamara & Van De Putte, 2022), η οποία μπορεί να συνδράμει στη μοντελοποίηση προσθέτοντας περαιτέρω θεωρητικό πλαίσιο, φορμαλισμό και εκφραστικότητα.

Επίσης ενδιαφέρον -για το συγγραφέα της διατριβής- παρουσιάζει η σχέση της οντολογικής αναπαράστασης και των γράφων γνώσης ως συν-διαμορφωτές των συμπερασμάτων που παράγονται μέσω των μοντέλων γενετικής τεχνητής νοημοσύνης (generative AI). Ειδικά όσον αφορά στην δυνατότητα των γράφων γνώσης να συνδράμουν, ως πηγές αξιόπιστης συλλογιστικής, για να αυξηθεί η επαληθευσσιμότητα και η αξιοπιστία των συμπερασμάτων των LLMs. Οπότε προσεγγίσεις όπως η RAG και το Neuro-symbolic AI που αναφέρθηκαν παραπάνω δείχνουν μια πολύ ενδιαφέρουσα προοπτική.

## 8 Συμπεράσματα

Το κάθε ένα από τα τρία βασικά κεφάλαια της διατριβής διαθέτει τα δικά του συμπεράσματα. Στο παρόν κεφάλαιο επιχειρείται μια σύνοψη των επιμέρους συμπερασμάτων και η ένταξη αυτών σε ένα ενιαίο πλαίσιο.

Η διατριβή χτίζει πάνω στη θέση του Fortun κ.α. για την ανάγκη πειραματισμού των ανθρωπολόγων με τις ψηφιακές μορφές, με τρόπους οι οποίοι προμηνύουν την επαναδιατύπωση της εθνογραφίας και του εθνογραφικού αρχείου και τον συνεπαγόμενο επανασχεδιασμό της ψηφιακής υποδομής εντός της οποίας θα υπάρξουν (Fortun κ.ά., 2017, σελ. 13). Μέσα σε αυτό το πλαίσιο εξετάστηκε το σύγχρονο τεχνολογικό περιβάλλον ως ένα μέσο εντός του οποίου είναι δυνατό να παραχθεί και να επικοινωνηθεί η ανθρωπολογική γνώση, συμβάλλοντας στην εδραίωση και την ανάπτυξη της ανοικτής επιστήμης. Όχι όμως στο πώς η πληροφορία θα κωδικοποιηθεί σε bits και bytes αλλά σε ένα αφαιρετικό επίπεδο εννοιολόγησης. Άλλωστε η αλληλοσυσχέτιση της ανθρωπολογίας και της πληροφορικής αποτελεί πρόσφορο έδαφος για συζητήσεις περί εννοιολογήσεων αφού, όπως σημειώνει ο Abramson, τα υπολογιστικά εργαλεία επεκτείνουν το δυνητικό σκοπό της ανθρωπολογίας επειδή μπορούν να βοηθήσουν στη γενίκευση (Abramson κ.ά., 2018).

Η διατριβή τεκμηρίωσε πάνω στη θέση ότι οι κωδικοποιήσεις και οι αναπαραστάσεις, που αποτέλεσαν το αντικείμενό της, είναι και αυτές με τη σειρά τους πολιτισμικές κατασκευές, οι οποίες αντανακλούν μια συγκεκριμένη επιστημολογική προσέγγιση και ως εκ τούτου ούτε είναι, αλλά ούτε και αναπαριστούν, μια 'αντικειμενική' πραγματικότητα. Διότι η προσέγγιση που ακολουθήθηκε κατά τη διάρκεια της έρευνας δεν αντιμετωπίζει τους αλγορίθμους και τις οντολογικές αναπαραστάσεις ως αυτόνομα τεχνικά αντικείμενα αλλά ως σύνθετα κοινωνιοτεχνικά συστήματα (Seaver, 2018, σελ. 378). Άρα ο στόχος δεν είναι ένας απλοϊκός και στείρος θετικισμός, αλλά η ανάπτυξη κωδικοποιήσεων οι οποίες είναι συνεκτικές και εσωτερικά συνεπείς ώστε να μπορούν να επεξεργαστούν και να διατηρήσουν αποτελεσματικά στο διηνεκές τα εθνογραφικά δεδομένα, ενώ ταυτόχρονα καλούνται να αποτελέσουν τη βάση ώστε να αναπτυχθούν αναπαραστάσεις που θα επιτρέπουν την αξιόπιστη διάχυση των αποτελεσμάτων της ανθρωπολογικής έρευνας. Διότι, όπως ήδη σημειώθηκε στην ενότητα 4.6, αν ο ερευνητής δεν μπορεί να μοιραστεί αυτό που παρατήρησε, τότε, για την επιστήμη, αυτή η παρατήρηση δεν υφίσταται. Έτσι καθίσταται πολύ κρίσιμη η

διάκριση ανάμεσα στην άρρητη (tacit) και τη ρητή (explicit) γνώση. Και, επειδή η πρώτη δεν μπορεί ποτέ να είναι κριτική (Polanyi, 1997, σελ. 278), αυτό αποτελεί έναν επιπλέον λόγο για το ότι δεν μπορεί να αποτελεί μέρος της επιστήμης. Η επιστήμη είναι πάντα αναπόσπαστα συνδεδεμένη με την κριτική σκέψη και την ερμηνεία.

Η ερμηνεία (interpretation), με τη σειρά της, είναι πολύ κρίσιμη για την κατανόηση και άρρηκτα συνυφασμένη με τα συμφραζόμενα (context). Στο εξώφυλλο της διατριβής υπάρχει η ακολουθία 0011101000101110100101001. Δεν πρόκειται για μια τυχαία απαρίθμηση από μηδέν και ένα, με καλλιτεχνικό και μόνο συμβολισμό. Πρόκειται για τρεις ASCII χαρακτήρες οι οποίοι αντιστοιχούν στην άνω κάτω τελεία, την παύλα και την παρένθεση. Η παράθεση αυτών των τριών χαρακτήρων αποτελούσε τον πρόδρομο του emoji με τη γελαστή φατσούλα. Ενώ ο XML/RDF κώδικας που εμφανίζεται: `<rdf:Description rdf:about="http://smiling.face/emoticon"/>` θα μπορούσε να ήταν ένα URI που αντιστοιχεί στο νόημα που οι άνθρωποι αποδίδουν σε αυτούς τους χαρακτήρες. Δηλαδή στη γελαστή φατσούλα, ανεξάρτητα από τον τρόπο που αυτή εικονίζεται κάθε φορά. Άρα, στο επίπεδο του φυσικού κόσμου, με τη λογική του ηλεκτρικού σήματος, το σήμα που μεταφέρεται (εντός του καλωδίου ή ασύρματα) είναι το 0011101000101110100101001. Στο επίπεδο του ψηφιακού κόσμου, δηλαδή της μετατροπής του ρεύματος σε χαρακτήρες ή σήματα επί της οθόνης, αυτό που εμφανίζεται στην οθόνη είναι το :-). Και τέλος, στο επίπεδο της ανθρώπινης κατανόησης, αυτό σημαίνει ένα γελαστό πρόσωπο, ως εκδήλωση συναισθήματος χαράς.

Όμως τα bits δεν έχουν κανένα νόημα από μόνα τους. Όπως είναι ήδη γνωστό, από τη θεωρία της πληροφορίας του Shannon, ο πομπός και ο δέκτης πρέπει να έχουν συμφωνήσει για το νόημα που έχουν οι πληροφορίες που ανταλλάσσονται (Shannon, 1948). Η ακολουθία 0011101000101110100101001 ερμηνεύεται έτσι αν αφορά τρεις ASCII χαρακτήρες (8 bits / χαρακτήρα). Αυτό όμως συμβαίνει διότι ο συγγραφέας επέλεξε να ερμηνεύσει τα εν λόγω bits ως ASCII χαρακτήρες. Αν τα ίδια bits ήταν τμήμα μιας εικόνας, τότε η εν λόγω ακολουθία χαρακτήρων θα δήλωνε ένα εικονοστοιχείο (pixel) καφεκόκκινου χρώματος (Hex: #3A2D29) στην RGB κλίμακα. Ή αν ήταν μέρος ενός midi μουσικού αρχείου θα δήλωνε τη μουσική νότα F2 με μέτρια ένταση και συγκεκριμένη διάρκεια. Ή, ανάλογα με τον τύπο αρχείου, και κάτι άλλο κάθε φορά. Άρα τα τρία επίπεδα αναπαράστασης συνοψίζονται ως εξής: τι περιέχει το σήμα (επίπεδο 0 και 1), το πώς το σήμα -δηλαδή τα μηδέν και ένα- μετατρέπονται από τον υπολογιστή στο ψηφιακό περιβάλλον και, τέλος, το νόημα που αποδίδεται σε αυτή την πληροφορία. Όπως είναι σαφές από το παράδειγμα, και τα τρία εξαρτώνται από τα συμφραζόμενα.

Άρα, αφού όλα είναι εξαρτώμενα από τα συμφραζόμενα (context depended), είναι πάρα πολύ σημαντικό να δηλώνεται ρητά το νόημα ώστε τοποθετώντας το εκάστοτε σήμα εντός των σωστών συμφραζομένων, να συντελεστεί και η σωστή αποκωδικοποίηση του μηνύματος. Για αυτό είναι σημαντική η σημασιολογία, διότι δίνει το κλειδί για την ορθή ερμηνεία (interpretation) των δεδομένων, ώστε να μπορεί ο δέκτης να αποδώσει στο μήνυμα το νόημα που ο πομπός είχε πρόθεση να μεταδώσει. Όπου πομπός και δέκτης, στο παράδειγμά μας, είναι άνθρωποι και όχι υπολογιστικά συστήματα. Αυτά ήταν απλά το μέσο που μετέφερε το μήνυμα.

Μένοντας στο παράδειγμα με τα emoticons, και συνεχίζοντας επί της ερμηνείας του ατόμου δέκτη, να σημειωθεί ότι οι Δυτικοί χρησιμοποιούν τα πλαγίως προσανατολισμένα emoticons όπως το γελαστό πρόσωπο που μας απασχολεί. Αντίθετα, οι Ασιάτες, ειδικά οι Ιάπωνες, ανέπτυξαν το δικό τους στυλ. Αυτά τα emoticons (ονομάζονται kaomoji) διαβάζονται κανονικά -όχι στο πλάι- και χρησιμοποιούν μεγαλύτερη ποικιλία χαρακτήρων. Έτσι το γελαστό πρόσωπο απεικονίζεται ως (^ ◡ ^). Άρα τελείως διαφορετικοί χαρακτήρες και απεικόνιση, αλλά το νόημα παραμένει το ίδιο.

Μέσα στο παραπάνω ερμηνευτικό πλαίσιο, και για την αποτελεσματική κάλυψη του στόχου της, η μελέτη κινήθηκε σε δύο βασικές διαστάσεις: από τη μία εξέτασε ζητήματα που αφορούν στη χρήση της πληροφορικής ώστε να βοηθηθούν οι ανθρωπολόγοι στην επεξεργασία των δεδομένων τους και την εξαγωγή συμπερασμάτων από αυτά, ενώ από την άλλη εξέτασε τρόπους για την αποτύπωση αυτών των συμπερασμάτων με μηχαναγνώσιμο τρόπο. Παράλληλα, αμφισβήτησε τη φυσική γλώσσα ως αποκλειστικό εργαλείο για την αποτύπωση των αποτελεσμάτων της ανθρωπολογικής έρευνας και διερεύνησε υπολογιστικές κωδικοποιήσεις και αναπαραστάσεις που είναι κατάλληλες για τα εθνογραφικά δεδομένα και την ανθρωπολογική γνώση.

Για το σκοπό αυτό εξέτασε μεθόδους και εργαλεία που μπορούν να λειτουργήσουν επικουρικά για την εξαγωγή συμπερασμάτων: από βάσεις δεδομένων έως λογισμικά ποιοτικής ανάλυσης δεδομένων (CAQDAS), αλλά και συστήματα αυτόματης συνεπαγωγής, εστιάζοντας στην Prolog ως μέσο ανάλυσης της συγγένειας, παρατηρήθηκε η μετάβαση από τη χρήση των υπολογιστών ως απλά εργαλεία μετρήσεων στη χρήση της πληροφορικής ως εργαλείο ανάπτυξης νέων μεθόδων για την ανάλυση πολλών διαφορετικών τύπων δεδομένων -και όχι κατ' ανάγκη αριθμητικών- ώστε να συνδράμει στην επίτευξη τεκμηριωμένων συμπερασμάτων.

Όμως όσο περισσότερο προχωράει η μετάβαση προς την αλγοριθμική επεξεργασία των δεδομένων, τόσο προκύπτει το θέμα της διαφάνειας αυτής της επεξεργασίας. Ο Kemper τονίζει πως όσο περισσότερο οι αλγόριθμοι ενσωματώνονται στη λειτουργία φορέων και θεσμών, τόσο γίνονται πιο επιδραστικοί και αδιαφανείς. Όσοι δημιουργούν τους αλγόριθμους μπορεί να κάνουν αυθαίρετες επιλογές οι οποίες όμως αποκρύπτονται από την κοινή θέα. Έτσι οι αλγόριθμοι αντικατοπτρίζουν τη μεροληψία των δημιουργών τους και μπορεί να επιβάλουν κάποιες προτιμήσεις ή θέσεις έναντι άλλων (Kargouzis, 2025; Kemper & Kolkman, 2018). Αυτός είναι και ο βασικός λόγος για τον οποίον είναι απαραίτητη η διαφάνεια αναφορικά τόσο με τη δημιουργία όσο και τη λειτουργία τους. Και αυτό δεν είναι ένα πρόβλημα που αφορά μόνο τους κοινωνικούς επιστήμονες αλλά και τους πολίτες. Διακρίνοντας το πρόβλημα, κάποιες χώρες προχωρούν σε μέτρα ώστε να διασφαλιστεί η διαφάνεια των ενεργειών που κάνουν τα λογισμικά που χρησιμοποιεί η δημόσια διοίκηση. Έτσι η Ελβετία ανέλαβε νομοθετική πρωτοβουλία και ψήφισε ομοσπονδιακό νόμο -που έχει τεθεί σε ισχύ από το 2024- όπου απαιτείται από τις δημόσιες αρχές να δημοσιεύουν τον κώδικα κάθε λογισμικού που αναπτύσσουν ή αναθέτουν σε τρίτους, εκτός αν ζητήματα ασφαλείας ή δικαιώματα τρίτων το εμποδίζουν. Ο νόμος είναι γνωστός με το ακρωνύμιο EMBAG και προωθεί τη συνεργασία μέσω μιας διακυβερνητικής κοινότητας ανοιχτού κώδικα, η οποία ενθαρρύνει την κοινή ανάπτυξη λύσεων μεταξύ φορέων του δημοσίου.

Πέρα από τη διαφάνεια, το θέμα που αφορά κυρίως τους κοινωνικούς επιστήμονες επεκτείνεται και γύρω από τη συλλογή, τη μοντελοποίηση και την επεξεργασία των ίδιων των δεδομένων. Άλλωστε πλέον μια σειρά από συσκευές καταγράφουν συνεχώς δεδομένα με ποικίλους τρόπους. Άρα στην καθημερινή διάδραση με το περιβάλλον τους οι άνθρωποι παράγουν ένα τεράστιο ψηφιακό αποτύπωμα η ανάλυση του οποίου θέτει νέα διάσταση στην ανθρωπολογική έρευνα, διότι τα ψηφιακά ίχνη (digital traces) μπορούν να αποτελέσουν πολύτιμο υλικό. Όμως αυτά τα δεδομένα δεν μπορούν να αναλυθούν με τις παραδοσιακές μεθόδους. Η επεξεργασία τους προϋποθέτει προηγμένες υπολογιστικές τεχνικές. Μέσα σε αυτό το περιβάλλον σημαντικός κρίνεται ο ρόλος των big data, η ανάλυση των οποίων απαιτεί ακόμη πιο εξελιγμένους τρόπους μετρήσεων (advanced metrics) που πολύ απέχουν από τους απλούς αριθμητικούς υπολογισμούς (calculations). Προφανώς δεν είναι μόνο τα ψηφιακά ίχνη που μπορούν να επεξεργαστούν υπολογιστικά αλλά και ένα μεγάλο μέρος του συμβατικού υλικού θα μπορούσε να αποτελέσει «τροφή» για υπολογιστική επεξεργασία. Όμως τα γεννημένα ψηφιακά δεδομένα είναι σαφώς πιο κοντά στην άμεση υπολογιστική επεξεργασία, ενώ τα αναλογικά θα πρέπει να υποστούν κάποιου είδους ψηφιοποίηση.

Τα όσα διαπραγματεύεται η διατριβή σχετικά με την υπολογιστική ανθρωπολογία και εθνογραφία δεν έχουν σκοπό να αντικαταστήσουν μεθόδους ή πρακτικές που ήδη χρησιμοποιούνται αλλά να επεκτείνουν κάποιες από αυτές ή να δημιουργήσουν κάποιες νέες. Προφανώς τόσο οι επεκτάσεις παλιών προσεγγίσεων όσο και οι νέες που προτείνονται δεν αποτελούν ούτε πανάκεια ούτε λύση για όλα τα είδη προβλημάτων της ανθρωπολογίας. Αντιθέτως φέρουν και αυτές με τη σειρά στο προσκήνιο μια σειρά από προβληματισμούς και ζητήματα. Ο ένας προβληματισμός σχετίζεται με την προστασία ευαίσθητων προσωπικών δεδομένων, διάσταση η οποία ήταν ήδη γνωστή στους ανθρωπολόγους αλλά πλέον τίθεται σε νέα βάση. Ο δεύτερος σχετίζεται με την ενδεχομένως περιορισμένη πρόσβαση στα συμφραζόμενα παραγωγής των δεδομένων, η οποία οδηγεί σε λανθασμένες ερμηνείες. Και ο τρίτος αφορά την μοντελοποίηση που θα μπορούσε να οδηγήσει στην επανάχρηση των δεδομένων. Σε αυτό το πλαίσιο, οι αρχές FAIR (Findable, Accessible, Interoperable, Reusable) προτείνουν μια κατεύθυνση για τη βελτίωση της διαχείρισης των δεδομένων, στοχεύοντας στο να ενισχύσουν ότι τα δεδομένα θα είναι ανιχνεύσιμα, προσβάσιμα, διαλειτουργικά και επαναχρησιμοποιήσιμα.

Ουσιαστικά στο κεφάλαιο 4 διακρίναμε δύο βασικές κατηγοριοποιήσεις: η μία αφορά μεθόδους ανάλυσης υποβοηθούμενες από υπολογιστική ανάλυση και η δεύτερη μεθόδους απευθείας εξαγωγής συμπερασμάτων από τα δεδομένα. Όμως και στη μία και στην άλλη περίπτωση δεν υπάρχει διαφυγή από την ερμηνεία των αποτελεσμάτων, αν και οι τεχνικές αυτές μπορεί να ελαχιστοποιήσουν την ύπαρξη -και μερικές φορές την πρακτική- των αυθαίρετων, μη τεκμηριωμένων, συμπερασμάτων (Abramson κ.ά., 2018, σελ. 259). Και σε αυτή τη φάση τίθεται το ερώτημα: το πώς η γνώση που προέκυψε (είτε από υπολογιστικές μεθόδους, είτε από παραδοσιακές) μπορεί να οργανωθεί και να αναπαρασταθεί με τρόπο πρόσφορο για το ψηφιακό περιβάλλον;

Η συστηματική σύγκριση Συστημάτων Οργάνωσης της Γνώσης (ΣΟΓ), εφαρμόζοντας -μεταξύ άλλων- και εκτεταμένη χρήση NLP έδειξε ότι οι οντολογίες δείχνουν κατάλληλες

για την αναπαράσταση της ανθρωπολογικής γνώσης. Ένας βασικός λόγος για αυτό είναι ο τρόπος που διαχειρίζονται τις σχέσεις διότι, ειδικά στο περιβάλλον του σημασιολογικού ιστού, οι κατηγορίες (κλάσεις) είναι κενές νόημα χωρίς την ύπαρξη σχέσεων, αφού οι σχέσεις είναι εκείνες που τους δίνουν νόημα. Από την άλλη, οι θησαυροί δείχνουν ότι είναι κατάλληλα εργαλεία για εξειδίκευση ως λεξιλόγια ελεγχόμενων τιμών (controlled value vocabularies). Και τα δύο αυτά αποτελούν πρόσφορα μέσα για την αποτύπωση της πληροφορίας στο περιβάλλον του σημασιολογικού ιστού.

Ο σημασιολογικός ιστός σηματοδοτεί μια μετατόπιση του παραδείγματος (paradigm shift) εντός της πληροφορικής δίνοντας έμφαση στη δομή των δεδομένων, σε αντίθεση με την έμφαση που δίνεται στην ανάπτυξη αλγορίθμων για την επεξεργασία αδόμητων δεδομένων. Προφανώς η κωδικοποίηση των δεδομένων κάθε άλλο παρά νέα προσέγγιση είναι. Από το Μοντέλο Οντοτήτων Συσχετίσεων (Entity-Relationship Model) μέχρι την XML, η ιστορία της πληροφορικής είναι γεμάτη από προσπάθειες μοντελοποίησης των δεδομένων. Όμως στο πλαίσιο του σημασιολογικού ιστού η δομή αποτελείται από απεριόριστο αριθμό σχέσεων μεταξύ στοιχείων που αναγνωρίζονται παγκοσμίως (μέσω URI), τα οποία είναι ανεξάρτητα από τα άμεσα συμφραζόμενα. Αυτό σημαίνει ότι δεν έχει σημασία το που εμφανίζεται ένα στοιχείο αφού η ταυτότητά του παραμένει αναλλοίωτη ανεξάρτητα από το πλαίσιο μέσα στο οποίο εμφανίζεται. Αυτή είναι η βασική καινοτομία του σημασιολογικού ιστού.

Άρα ο σημασιολογικός ιστός αποτελεί το απαύγασμα της μοντελοποίησης των δεδομένων που απογειώνει τη διαλειτουργικότητα και επιτρέπει την αποτελεσματική επαναχρησιμοποίηση (re-use) των δεδομένων. Το περιβάλλον αυτό αποτελεί σημαντική συμβολή στην κωδικοποίηση και δημοσιοποίηση πολιτισμικού περιεχομένου με όρους που συμβάλλουν ουσιαστικά τόσο στην επικοινωνία εντός του πεδίου της ανθρωπολογίας όσο και στην ανάπτυξη της διεπιστημονικότητας. Κατ' αυτόν τον τρόπο, τα εθνογραφικά δεδομένα και οι ανθρωπολογικές αναλύσεις μπορούν να έχουν τον αθροιστικό χαρακτήρα ενός αρχείου και την προστιθέμενη αξία που αυτό συνεπάγεται, ενώ ταυτόχρονα να είναι ανοικτά σε νέες αναλύσεις και ερμηνείες.

Η κοινωνική ανθρωπολογία, ως μια επιστήμη άρρηκτα συνδεδεμένη με τη γλώσσα και ειδικά με την γραφή, σπάνια πειραματίζεται με άλλες μεθόδους αναπαράστασης των αποτελεσμάτων της. Αντιθέτως το περιβάλλον του σημασιολογικού ιστού δεν είναι κειμενοκεντρικό αλλά εννοιοκεντρικό και διαθέτει πολλαπλά επίπεδα αφαίρεσης και πολυπλοκότητας. Βέβαια, η παρούσα διατριβή δεν είναι ένα κείμενο που προτείνει την κατάργηση των κειμένων! Απλά επιχειρεί να καταδείξει ότι η γραφή και η παραγωγή κειμένων δεν είναι μονόδρομος για την καταγραφή των αποτελεσμάτων της ανθρωπολογικής έρευνας. Και άλλοι δρόμοι είναι εφικτοί. Ο σημασιολογικός ιστός, οι γράφοι γνώσης και η αναπαράσταση γνώσης μπορούν να συμβάλουν στην αποτύπωση, επεξεργασία και διάχυση της ανθρωπολογικής γνώσης.

Παρότι η προσέγγιση που ακολουθείται είναι εννοιοκεντρική, σε καμία περίπτωση δεν υποτιμάται η σημασία που έχουν οι λέξεις για την απόδοση των εννοιών. Οι λέξεις, ως αναγνωριστικά των εννοιών, είναι ζωτικής σημασίας για τους ανθρώπινους πολιτισμούς. Πριν από περίπου μια δεκαετία, στο πλαίσιο της «πολιτικής ορθότητας»

και του φόβου ότι η γλώσσα μπορεί να συμβάλει στο ρατσισμό, η Σουηδία άλλαξε τα ονόματα ειδών πουλιών που θα μπορούσαν ενδεχομένως να προσβάλουν, όπως «γυφτοπούλι» και «νέγρος» (Noack, 2015). Επίσης, είναι γνωστό ότι υπάρχουν γυναίκες συγγραφείς που υπέγραφαν χρησιμοποιώντας ανδρικά ονόματα -για παράδειγμα η Mary Ann Evans έγραψε ως George Eliot- για ευνόητους, για την εποχή της<sup>58</sup>, λόγους. Άρα τα ονόματα δεν είναι τυχαίες ακολουθίες χαρακτήρων ή ήχων που χρησιμεύουν απλώς ως αναγνωριστικά στοιχεία για τις οντότητες. Έχουν επίσης κοινωνικοπολιτισμικές σημασίες και ερμηνείες, οι οποίες σε πολλές περιπτώσεις είναι λανθάνουσες. Η εννοιοκεντρική προσέγγιση που αναλύθηκε στην παρούσα διατριβή ουδόλως αγνοεί αυτή τη διάσταση. Τα μοντέλα που βασίζονται στις τεχνολογίες του σημασιολογικού ιστού παρέχουν διάφορους τρόπους για τη μοντελοποίηση των λέξεων και ο συγγραφέας της παρούσας διατριβής έχει ασχοληθεί συστηματικά με αυτούς (Peronakis, 2016).

Μέσω της προτεινόμενης μοντελοποίησης, η γνώση των επιστημόνων μπορεί να χρησιμοποιηθεί μέσα σε ένα σαφώς ορισμένο πλαίσιο, η ρητή αποτύπωση του οποίου επιτρέπει την ανάλυση τόσο σε μικρό όσο και σε μεγάλο επίπεδο (και σε πολλά ενδιάμεσα), δημιουργώντας ένα πολύ αποτελεσματικό σύστημα, ειδικά όταν αυτό συνομιλεί με άλλα δεδομένα, αφού το βασικό αντικείμενο των διασυνδεδεμένων δεδομένων και των σημασιολογικών τεχνολογιών είναι η ενθάρρυνση της συνεργασίας (Oldman κ.ά., 2016). Άλλωστε η δύναμη των υπολογιστών στο να γεφυρώσουν το κενό μεταξύ θεωρίας και συλλογής δεδομένων δεν πρόκειται να γίνει πράξη αν δεν τους αναθέσουμε αποστολές πιο σημαντικές από την απλή οργάνωση των σημειώσεών μας και την παρουσίαση επιλεγμένων πληροφοριών σε πίνακες και γραφήματα (Kirpen, 1988, σελ. 319). Το πώς θα επιλέξουν οι ανθρωπολόγοι να συνδιαλλαγούν με την πληροφορική εξαρτάται από εκείνους, αλλά σίγουρα οι όποιες αποφάσεις τους θα επηρεάσουν το περιβάλλον μέσα στο οποίο η γνώση τους θα μπορεί να αυξηθεί, να τεκμηριωθεί, να διαχυθεί και να επιτύχει μια αμφίδρομη επικοινωνία με άλλους επιστημονικούς τομείς.

Όπως ήδη σημειώθηκε στην εισαγωγή αυτού του κειμένου, η οντολογική αναπαράσταση που προτείνεται κάνει αυτό που (θα έπρεπε να) κάνει και η Ανθρωπολογία: διασαφηνίζει ρητά αυτό που υποβόσκει ως γνώση στα άτομα. Με άλλα λόγια η αναπαράσταση της γνώσης, μέσω οντολογιών, προσιδιάζει με την κοινωνική ανθρωπολογία στο σημείο του ότι προσπαθεί να καταγράψει, με σαφή και επιστημονικό τρόπο, τα όσα ήδη είναι γνωστά σε μια ομάδα ατόμων και να χρησιμοποιήσει αυτή την καταγραφή για να βγάλει συμπεράσματα ερμηνεύοντας τη γνώση που κατέγραψε.

Η προτεινόμενη μοντελοποίηση επιτρέπει μια ουσιαστική ολιστική προσέγγιση αφού δίνει τη δυνατότητα να ενσωματωθούν όλες οι επιμέρους οπτικές σε ένα κοινό “αρχείο”. Η ολιστική προσέγγιση ήταν πάντα ζητούμενο για την ανθρωπολογία αλλά στην πράξη μάλλον δεν εφαρμόζεται. Ο Borofsky ανέλυσε τα άρθρα που δημοσιεύτηκαν

---

<sup>58</sup> Η Mary Ann Evans έζησε τον 19<sup>ο</sup> αιώνα και χρησιμοποιούσε ανδρικό λογοτεχνικό ψευδώνυμο. Ο λόγος ήταν, όπως είχε δηλώσει η ίδια, πως επιθυμούσε να παίρνουν τα έργα της στα σοβαρά.

στο περιοδικό *American Anthropologist* για μια περίοδο 100 ετών, και έδειξε ότι η ουσιαστική συνεργασία μεταξύ των ανθρωπολογικών υποκλάδων είναι σε μεγάλο βαθμό ένας μύθος - εφαρμοζόταν μόνο σε 311 από τα 3.264 άρθρα που εξετάστηκαν (ή 9,5% του συνόλου). Η εν λόγω μελέτη καταλήγει κάνοντας έκκληση για νέες μορφές ολισμού (Borofsky, 2002).

Ενώ άλλοι τομείς της τεχνητής νοημοσύνης (όπως η μηχανική μάθηση και τα *black boxes* που αναφέρθηκαν προηγουμένως) αποτυπώνουν και επεξεργάζονται τα δεδομένα με τρόπους οι οποίοι είναι “κατανοητοί” μόνο από τους υπολογιστές, η αναπαράσταση γνώσης βασίζεται σε μια λογική η οποία επιβάλλει ανθρωποκεντρική οπτική αφού η κωδικοποίηση της γνώσης γίνεται με τρόπο που είναι αναγνωρίσιμος τόσο από τον άνθρωπο όσο και από τον υπολογιστή. Όμως δεν είναι δυνατή η μεταφορά της γνώσης του ειδικού χωρίς να προηγηθεί μια ειδική διαδικασία μοντελοποίησης. Όπως σημειώνει ο Guarino «μια βάση γνώσης δεν είναι ένα αποθετήριο γνώσης η οποία εξήχθη από το μυαλό των ειδημόνων (σα να μεταλαμπαδεύτηκε), αλλά το αποτέλεσμα μιας διαδικασίας μοντελοποίησης» (Guarino, 1995, σελ. 625). Η αναπαράσταση σε μηχαναγνώσιμη μορφή των συμπερασμάτων -είτε αυτά προκύπτουν από *data-driven* μεθόδους είτε όχι- χρησιμοποιώντας τεχνολογίες σημασιολογικού ιστού, και συγκεκριμένα οντολογικής αναπαράστασης γνώσης, δίνει τη δυνατότητα επανατροφοδότησης του συστήματος και ικανότητας συμπερασμού (*reasoning*) από την πλευρά των υπολογιστών.

Μπορεί η αναπαράσταση γνώσης να βρίσκεται στον πυρήνα της κοινωνικής ανθρωπολογίας (M. D. Fischer, 1994, σελ. 3) όμως πρέπει να τονιστεί ότι δεν πρόκειται ούτε για αναπαράσταση ενός αντικειμενικού κόσμου ούτε για αντικειμενική αναπαράσταση ενός σχετικού κόσμου. Άλλωστε η γνώση είναι πάντα ελλιπής, αυτό που ζητά η επιστήμη είναι η καλύτερη δυνατή εξήγηση (Doerr, 2025). Η άποψη των ανθρωπολόγων για τον κόσμο (όποια και αν είναι αυτή) επιχειρείται να αναπαρασταθεί. Διότι η οντολογική αναπαράσταση στην οποία αναφερόμαστε «δεν ενδιαφέρεται για την απευθείας ταυτοποίηση και χαρακτηρισμό της πραγματικότητας μέσω κατηγοριών αλλά πώς η πραγματικότητα αναπαρίσταται στο νου και όχι στο πώς αναπαρίσταται από το νου» (Kohne, 2014, σελ. 87). Και βεβαίως έχει εφαρμογή και σε αυτού του τύπου τις αναπαραστάσεις η γνωστή φράση, που αποδίδεται στον Σπινόζα: «Αυτά που ο Παύλος λέει για τον Πέτρο μάς λένε περισσότερα για τον Παύλο παρά για τον Πέτρο», αφού, και στις μη κειμενοκεντρικές αναπαραστάσεις, αυτά που διαλέγουμε να περιγράψουμε εκφράζουν σε μεγάλο βαθμό εκείνους που επέλεξαν αυτά τα χαρακτηριστικά ενώ αποσιώπησαν κάποια άλλα. Απλά ο συγκεκριμένος τρόπος περιγραφής έχει δύο πολύ σημαντικά πλεονεκτήματα σε σχέση με τη φυσική γλώσσα: πρώτον, επιδέχεται λιγότερες παρερμηνείες από ότι η φυσική γλώσσα και δεύτερον, εκτοξεύει τις δυνατότητες υπολογιστικής επεξεργασίας, αυξάνοντας τις δυνατότητες για αύξηση του βαθμού αποτελεσματικότητας στα τρία βασικά χαρακτηριστικά που θέλουμε να έχει η ανθρωπολογία, ήτοι: εγκυρότητα, αξιόπιστη αναπαράσταση και δυνατότητα γενίκευσης (*validity, reliable representation, and generalizability*) (Abramson & Dohan, 2015).

Όπως προκύπτει από την ανάλυση που προηγήθηκε, οι υπολογιστικές τεχνολογίες μπορούν να βοηθήσουν τόσο στην δημιουργία όσο και στην αναπαράσταση της

ανθρωπολογικής γνώσης. Όμως ήδη από τον πρόλογο αυτής της διατριβής αναφερθήκαμε στην αδυναμία δημιουργίας μιας αντικειμενικής και απόλυτης ταξινόμησης του κόσμου. Όπως γράφει ο Peres «Το πρόβλημά μου, με τις ταξινομήσεις, είναι ότι δεν διαρκούν· μόλις έχω τελειώσει να βάζω μια τάξη, η τάξη αυτή έχει ήδη ακυρωθεί» (Peres, 2005, σελ. 229). Κάθε μοντέλο καταγραφής έχει έναν προσωρινό χαρακτήρα, διότι είναι ένα κοινωνιοτεχνικό δημιούργημα που αντανακλά τον τρόπο που οι δημιουργοί του κατατάσσουν τον κόσμο που τους περιβάλλει. Μετά από λίγο θα εκλεπτούν -ή θα τροποποιήσουν ριζικά- τα μοντέλα τους και έτσι νέες ανάγκες επανερμηνείας και μοντελοποίησης θα προκύψουν, συμβάλλοντας στην αέναη διαλεκτική του πώς η αποτύπωση της έρευνας επηρεάζει τον τρόπο διεξαγωγής της και πώς αυτός με τη σειρά του την αποτύπωσή της. Συνεπώς προκύπτει ένας συνεχής μετασχηματισμός των τρόπων με τους οποίους σκέφτεται, αποτυπώνει τις σκέψεις, αλλά και υποβάλλει τα συμπεράσματά του στη βάση της, εκάστοτε έγκυρης, επιστημονικής μεθόδου ο Άνθρωπος.

## Βιβλιογραφικές αναφορές

- Γκέφου-Μαδιανού, Δ. (1998). Αναστοχασμός, ετερότητα και ανθρωπολογία οίκοι: Διλήμματα και αντιπαραθέσεις. Στο *Ανθρωπολογική θεωρία και Εθνογραφία: Σύγχρονες τάσεις* (4η έκδ., σελ. 365–435). Ελληνικά Γράμματα.
- Θεωρία πληροφορίας. (2023). Στο *Βικιπαίδεια*. [https://el.wikipedia.org/w/index.php?title=%CE%98%CE%B5%CF%89%CF%81%CE%AF%CE%B1\\_%CF%80%CE%BB%CE%B7%CF%81%CE%BF%CF%86%CE%BF%CF%81%CE%AF%CE%B1%CF%82&oldid=9964643](https://el.wikipedia.org/w/index.php?title=%CE%98%CE%B5%CF%89%CF%81%CE%AF%CE%B1_%CF%80%CE%BB%CE%B7%CF%81%CE%BF%CF%86%CE%BF%CF%81%CE%AF%CE%B1%CF%82&oldid=9964643)
- Καπιδάκης, Σ., Πεπονάκης, Μ., & Doerr, M. (2024). *Τεκμηρίωση μουσειακών συλλογών: Μεταδεδομένα για αντικείμενα πολιτισμικής κληρονομιάς*. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. <http://dx.doi.org/10.57713/kallipos-955>
- Μάστορα, Ά. (2016). *Συμβολή στη Μελέτη της Εννοιολογικής Συσχέτισης των Όρων που Χρησιμοποιούνται από τους Χρήστες κατά τη Σύνθεση και Ανασύνθεση Ερωτημάτων σε Ψηφιακές Βιβλιοθήκες [Διδακτορική Διατριβή]*. Ιόνιο Πανεπιστήμιο. Τμήμα Αρχειονομίας και Βιβλιοθηκονομίας (TAB).
- Μήτρου, Λ. (Επιμ.). (2023). *Μπορεί ο αλγόριθμος... Να είναι ηθικός, να είναι δίκαιος, να είναι διαφανής, να δικάζει & να διοικεί*. Πανεπιστημιακές Εκδόσεις Κρήτης.
- Παναγιωτίδης, Φ. (2014). *Μίλα μου για γλώσσα: Μικρή εισαγωγή στη γλωσσολογία*. Πανεπιστημιακές Εκδόσεις Κρήτης.
- Παπαηλία, Π., & Πετρίδης, Π. (2015). *Ψηφιακή εθνογραφία*. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. <https://repository.kallipos.gr/handle/11419/6117>
- Τουντασάκη, Ε. (2008). *Ανθρωπολογικές θεωρήσεις της συγγένειας κατά τον 20ό αιώνα*. Ελληνικά Γράμματα.
- Τουντασάκη, Ε. (2015). *‘Το παιδί που μεγαλώνει μέσα σου θα πάρει κι από σένα’: Δωρεά ωαρίων, μητρότητα και συγγένεια*. Πατάκης.
- Τουντασάκη, Ε. (2022). *Διακρατική/διαφυλετική υιοθεσία: ‘ήπια μετανάστευση’ και νέες χωρογραφίες της συγγένειας*. Εκδόσεις Πατάκη.
- Χατζηβασιλείου, Γ. (2023). *Φιλοσοφία της τεχνητής νοημοσύνης: Ένα ταξίδι στο μέλλον*. Διόπτρα.
- Abrams, M. H. (2014). *Τι είναι μια ανθρωπιστική κριτική;* (Α. Μπερλής, Μτφ.). MIET.
- Abramson, C. M., & Dohan, D. (2015). Beyond Text: Using Arrays to Represent and Analyze Ethnographic Data. *Sociological Methodology*, 45(1), 272–319. <https://doi.org/10.1177/0081175015578740>
- Abramson, C. M., Joslyn, J., Rendle, K. A., Garrett, S. B., & Dohan, D. (2018). The promises of computational ethnography: Improving transparency, replicability, and validity for realist approaches to ethnographic analysis. *Ethnography*, 19(2), 254–284. <https://doi.org/10.1177/1466138117725340>
- Abreu, S. C. de, & Vieira, R. (2017). RelP: Portuguese Open Relation Extraction. *Knowledge Organization*, 44(3), 163–177. <https://doi.org/10.5771/0943-7444-2017-3-163>
- Acker, A. (2015). Toward a Hermeneutics of Data. *IEEE Annals of the History of Computing*, 37(3), 70–75. <https://doi.org/10.1109/MAHC.2015.68>

- Aitchison, J., & Clarke, S. D. (2004). The Thesaurus: A Historical Viewpoint, with a Look to the Future. *Cataloging & Classification Quarterly*, 37(3–4), 5–21. [https://doi.org/10.1300/J104v37n03\\_02](https://doi.org/10.1300/J104v37n03_02)
- Albris, K., Otto, E. I., Astrupgaard, S. L., Gregersen, E. M., Jørgensen, L. S., Jørgensen, O., Sandbye, C. R., & Schønning, S. (2021). A view from anthropology: Should anthropologists fear the data machines? *Big Data & Society*, 8(2), 205395172111043655. <https://doi.org/10.1177/205395172111043655>
- Alexiev, B., & Marksby, N. (2010). Terminology as Organized Knowledge. Στο C. Gnoli & F. Mazzocchi (Επιμ.), *Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the Eleventh International ISKO Conference, 23-26 February 2010, Rome, Italy* (σελ. 363–370). Ergon.
- Alexiev, V., Isaac, A., & Lindenthal, J. (2016). On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI). *International Journal on Digital Libraries*, 17(1), 39–48. <https://doi.org/10.1007/s00799-015-0162-2>
- Alvard, M., & Carlson, D. (2020). Identifying Patch Types Using Movement Data from Artisanal Fishers from the Commonwealth of Dominica. *Current Anthropology*, 61(3), 380–387. <https://doi.org/10.1086/708720>
- Anderson, A., Chilczuk, S., Nelson, K., Ruther, R., & Wall-Scheffler, C. (2023). The Myth of Man the Hunter: Women’s contribution to the hunt across ethnographic contexts. *PLOS ONE*, 18(6), e0287101. <https://doi.org/10.1371/journal.pone.0287101>
- Androutsopoulos, I., Lampouras, G., & Galanis, D. (2013). Generating Natural Language Descriptions from OWL Ontologies: The NaturalOWL System. *Journal of Artificial Intelligence Research*, 48, 671–715. <https://doi.org/10.1613/jair.4017>
- Anticoli, L., & Toppano, E. (2011a). How culture may influence ontology co-design: A qualitative study. *International Journal of Information Technology and Web Engineering*, 6(2), 1–17. <https://doi.org/10.4018/jitwe.2011040101>
- Anticoli, L., & Toppano, E. (2011b). The role of culture in collaborative ontology design. *Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applications*, 4:1-4:9. <https://doi.org/10.1145/1980822.1980826>
- Antoničević, S. (2020). Digital Workflow in the Humanities and Social Sciences: A Data Ethnography. Στο J. W. Crowder, M. Fortun, R. Besara, & L. Poirier (Επιμ.), *Anthropological Data in the Digital Age: New Possibilities – New Challenges* (σελ. 59–83). Springer International Publishing. [https://doi.org/10.1007/978-3-030-24925-0\\_4](https://doi.org/10.1007/978-3-030-24925-0_4)
- Antoniou, G., & Van Harmelen, F. (2008). *A Semantic Web Primer* (2ed ed.). MIT Press.
- Applin, S. A., & Fischer, M. D. (2015). New technologies and mixed-use convergence: How humans and algorithms are adapting to each other. *2015 IEEE International Symposium on Technology and Society (ISTAS)*, 1–6. <https://doi.org/10.1109/ISTAS.2015.7439436>
- Arnold, T., & Fuller, H. J. A. (2018). In Search of the User’s Language: Natural Language Processing, Computational Ethnography, and Error-Tolerant Interface Design. *Advances in Usability, User Experience and Assistive Technology*, 36–43. [https://doi.org/10.1007/978-3-319-94947-5\\_4](https://doi.org/10.1007/978-3-319-94947-5_4)
- Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with Basic Formal Ontology*. Massachusetts Institute of Technology.
- Artmann, S. (2010). Computers and Anthropology. Στο J. H. Birx (Επιμ.), *21st century anthropology: A reference handbook* (σελ. 915–924). Sage.

- Assael, Y., Sommerschild, T., Cooley, A., Shillingford, B., Pavlopoulos, J., Suresh, P., Herms, B., Grayston, J., Maynard, B., Dietrich, N., Wulgaert, R., Prag, J., Mullen, A., & Mohamed, S. (2025). Contextualizing ancient texts with generative neural networks. *Nature*, 1–7. <https://doi.org/10.1038/s41586-025-09292-5>
- Asuncion, C. H., & van Sinderen, M. J. (2010). Pragmatic Interoperability: A Systematic Review of Published Definitions. Στο P. Bernus, G. Doumeingts, & M. Fox (Επιμ.), *Enterprise Architecture, Integration and Interoperability* (σελ. 164–175). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-15509-3\\_15](https://doi.org/10.1007/978-3-642-15509-3_15)
- Athanasίου, Σ., Amiridis, Β., Gavriilidou, Μ., Evangelos Gerasopoulos, Alexandros Dimopoulos, Georgina Kaklamani, Fotis Karagiannis, Iraklis Klampanos, Dimitra Kondili, Kostas Koumantaros, Panos Konstantopoulos, Katerina Lenaki, Angelos Likiardopoulos, Natalia Manola, Despoina Mitropoulou, Agiatis Benardou, Nikolaos Boukos, Alexandros Nousias, Maria Daoudaki, ... Fotis Psomopoulos. (2020). *Εθνικό σχέδιο για την ανοικτή επιστήμη = National Plan for Open Science*. Zenodo. <https://doi.org/10.5281/zenodo.3908953>
- Ayimdji, A., Koussoubé, S., Fotso, L. P., & Konfé, B. O. (2011). Towards a “Deep” Ontology for African Traditional Medicine. *Intelligent Information Management*, 03, 244–251. <https://doi.org/10.4236/iim.2011.36030>
- Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., & Summers, E. (2013). Key choices in the design of Simple Knowledge Organization System (SKOS). *Web Semantics: Science, Services and Agents on the World Wide Web*, 20, 35–49. <https://doi.org/10.1016/j.websem.2013.05.001>
- Bakharia, A., & Corrin, L. (2019). Using Recent Advances in Contextual Word Embeddings to Improve the Quantitative Ethnography Workflow. Στο B. Eagan, M. Misfeldt, & A. Siebert-Evenstone (Επιμ.), *Advances in Quantitative Ethnography* (σελ. 299–306). Springer International Publishing.
- Balakireva, L., & Balakirev, F. (2022). Making FAIR Practices Accessible and Attractive. Στο G. Silvello, O. Corcho, P. Manghi, G. M. Di Nunzio, K. Golub, N. Ferro, & A. Poggi (Επιμ.), *Linking Theory and Practice of Digital Libraries* (σελ. 417–424). Springer International Publishing. [https://doi.org/10.1007/978-3-031-16802-4\\_41](https://doi.org/10.1007/978-3-031-16802-4_41)
- Barkhatova, L. A. (2023). The computational analysis of digital traces in ethnographic studies of online communities. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 160(1), 30–56. <https://doi.org/10.1177/07591063231196161>
- Barley, N. (2005). *Ο ανυποψίαστος ανθρωπολόγος: Σημειώσεις σε μία χωμάτινη καλύβα* (Μ. Μπονάτσου & Α. Μπουκάλα, Επιμ.; Γ. Σεφερλής, Μτφ.). Αιώρα.
- Bartlett, A., Lewis, J., Reyes-Galindo, L., & Stephens, N. (2018). The locus of legitimate interpretation in Big Data sciences: Lessons for computational social science from -omic biology and high-energy physics. *Big Data & Society*, 5(1), 2053951718768831. <https://doi.org/10.1177/2053951718768831>
- Bean, C. A., Green, R., & Myaeng, S. H. (2002). Introduction. Στο R. Green, C. A. Bean, & S. H. Myaeng (Επιμ.), *The Semantics of Relationships* (σελ. vii–xvi). Springer Netherlands.
- Beaulieu, A. (2017). Vectors for Fieldwork: Computational Thinking and New Modes of Ethnography. Στο L. Hjorth, H. A. Horst, A. Galloway, & G. Bell (Επιμ.), *The Routledge companion to digital ethnography* (σελ. 29–39). Routledge, Taylor & Francis Group. <https://www.routledge.com/The-Routledge-Companion-to-Digital-Ethnography-1st-Edition/Hjorth-Horst-Galloway-Bell/p/book/9781138940918>

- Behrens, C. A. (1990). Qualitative and quantitative approaches to the analysis of anthropological data: A new synthesis. *Journal of Quantitative Anthropology*, 2(1), 305–328.
- Behrens, C. A., & Read, D. W. (1993). Anthropology: Moving from task-driven to science-driven computing. *Social Science Computer Review*, 11(4), 429–451.
- Bekiari, C., Bruseker, G., & Doerr, M. (2021). *Definition of the CIDOC Conceptual Reference Model version 7.1.1* (σελ. 232). CIDOC CRM Special Interest Group. <https://www.cidoc-crm.org/version/version-7.1.1>
- Beller, S. (2001). A model theory of deontic reasoning about social norms. *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, 63–68.
- Beller, S. (2008). Deontic norms, deontic reasoning, and deontic conditionals. *Thinking & Reasoning*, 14(4), 305–341. <https://doi.org/10.1080/13546780802222258>
- Berners-Lee, T. (2006). *Interview with Tim Berners-Lee* (B. Runciman) [Συνέντευξη]. <https://www.impactlab.com/2006/03/25/interview-with-tim-berners-lee/>
- Bernstein, A., & Dohan, D. (2020). Using Computational Tools to Enhance Comparative Ethnography: Lessons from scaling ethnography for biomedicine. Στο *Beyond the Case: The Logics and Practices of Comparative Ethnography* (σελ. 209–237). Oxford University Press. <https://dx.doi.org/10.1093/oso/9780190608484.003.0009>
- Beuving, J. J. (2020). Ethnography's future in the big data era. *Information, Communication & Society*, 23(11), 1625–1639. <https://doi.org/10.1080/1369118X.2019.1602664>
- Bharwani, S. (2006). Understanding Complex Behavior and Decision Making Using Ethnographic Knowledge Elicitation Tools (KnETs). *Social Science Computer Review*, 24(1), 78–105. <https://doi.org/10.1177/0894439305282346>
- Bhuyan, B. P., Ramdane-Cherif, A., Tomar, R., & Singh, T. P. (2024). Neuro-symbolic artificial intelligence: A survey. *Neural Computing and Applications*, 36(21), 12809–12844. <https://doi.org/10.1007/s00521-024-09960-z>
- Birx, J. H. (Επιμ.). (2010). *21st Century Anthropology: A Reference Handbook*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412979283>
- Blair, D. (2006). *Wittgenstein, Language and Information: 'Back to the Rough Ground!'* (τ. 10). Springer.
- Boellstorff, T. (2015). *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human* (New edition). Princeton University Press.
- Borgatti, S. P. (1996). *ANTHROPAC 4.0*. Natick, MA: Analytic Technologies. [https://ay14-15.moodle.wisc.edu/prod/pluginfile.php/59450/mod\\_resource/content/0/APGUIDE6.doc](https://ay14-15.moodle.wisc.edu/prod/pluginfile.php/59450/mod_resource/content/0/APGUIDE6.doc)
- Boroditsky, L., & Gaby, A. (2010). Remembrances of Times East: Absolute Spatial Representations of Time in an Australian Aboriginal Community. *Psychological Science*, 21(11), 1635–1639. <https://doi.org/10.1177/0956797610386621>
- Borofsky, R. (2002). The Four Subfields: Anthropologists as Mythmakers. *American Anthropologist*, 104(2), 463–480. <https://doi.org/10.1525/aa.2002.104.2.463>
- Bryda, G. (2020). From CAQDAS to Text Mining. The Domain Ontology as a Model of Knowledge Representation About Qualitative Research Practices. Στο A. P. Costa, L. P. Reis, & A. Moreira (Επιμ.), *Computer Supported Qualitative Research* (σελ. 72–88). Springer International Publishing. [https://doi.org/10.1007/978-3-030-31787-4\\_6](https://doi.org/10.1007/978-3-030-31787-4_6)

- Burns, P., McCormack, T., Jaroslawska, A. J., O'Connor, P. A., & Caruso, E. M. (2019). Time Points: A Gestural Study of the Development of Space–Time Mappings. *Cognitive Science*, *43*(12), e12801. <https://doi.org/10.1111/cogs.12801>
- Burton, M. L. (1970). Computer applications in cultural anthropology. *Computers and the Humanities*, *5*(1), 37–45. <https://doi.org/10.1007/BF02404256>
- Burton, M. L. (1973). Recent computer applications in cultural anthropology. *Computers and the Humanities*, *7*(6), 337–341. <https://doi.org/10.1007/BF02395108>
- Caliandro, A. (2017). Digital Methods for Ethnography: Analytical Concepts for Ethnographers Exploring Social Media Environments. *Journal of Contemporary Ethnography*, in press. <https://doi.org/10.1177/0891241617702960>
- Campos, M. L. de A., & Gomes, H. E. (2017). Ontology: Several Theories on the Representation of Knowledge Domains. *Knowledge Organization*, *44*(3), 178–186. <https://doi.org/10.5771/0943-7444-2017-3-178>
- Carlson, S., & Anderson, B. (2007). What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. *Journal of Computer-Mediated Communication*, *12*(2), 635–651. <https://doi.org/10.1111/j.1083-6101.2007.00342.x>
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, *63*, 67–80. <https://doi.org/10.1016/j.dss.2013.08.008>
- Chansanam, W., Nguyen, L. T., Li, C., & Guan, C. K. S. (2025). Semantic Knowledge Graphs for Intercultural and Ethnographic Diversity in the Greater Mekong Subregion. *Journal of Intercultural Communication*, *25*(3), 11–24. <https://doi.org/10.36923/jicc.v25i3.1161>
- Chantas, G., Karavarsamis, S., Nikolopoulos, S., & Kompatsiaris, I. (2018). A Probabilistic, Ontological Framework for Safeguarding the Intangible Cultural Heritage. *Journal on Computing and Cultural Heritage*, *11*(3), 12:1-12:29. <https://doi.org/10.1145/3131610>
- Charles, V., & Gherman, T. (2018). Big Data Analytics and Ethnography: Together for the Greater Good. Στο *Big Data for the Greater Good* (σελ. 19–33). Springer. [https://doi.org/10.1007/978-3-319-93061-9\\_2](https://doi.org/10.1007/978-3-319-93061-9_2)
- Chi, Y.-L., Sung, H.-Y., & Lien, Y.-Y. (2020). Towards the Ethnic Understanding of Taiwanese Indigenous Peoples: A Mashup Based on Semantic Web and Open Data. Στο P.-L. P. Rau (Επιμ.), *Cross-Cultural Design. User Experience of Products, Services, and Intelligent Environments* (σελ. 287–297). Springer International Publishing. [https://doi.org/10.1007/978-3-030-49788-0\\_21](https://doi.org/10.1007/978-3-030-49788-0_21)
- Christin, A. (2020a). Algorithmic ethnography, during and after COVID-19. *Communication and the Public*, 2057047320959850. <https://doi.org/10.1177/2057047320959850>
- Christin, A. (2020b). The ethnographer and the algorithm: Beyond the black box. *Theory and Society*. <https://doi.org/10.1007/s11186-020-09411-3>
- Chung, H. (2006). *GlobalMind: Bridging the gap between different cultures and languages with common-sense computing* [Thesis, Massachusetts Institute of Technology]. <http://hdl.handle.net/1721.1/37388>
- Chung, H., & Lieberman, H. (2007). GlobalMind: Automated analysis of cultural contexts with multicultural common-sense computing. *International Journal on Semantic Web and Information Systems*, *3*(1), 65–95. <https://doi.org/10.4018/jswis.2007010104>

- Cioffi-Revilla, C. (2016). *Bigger Computational Social Science: Data, Theories, Models, and Simulations -- Not Just Big Data* (SSRN Scholarly Paper No. ID 2784278; σελ. 1–5). Social Science Research Network. <http://ssrn.com/abstract=2784278>
- Clarke, S. D. (2001). Thesaural Relationships. Στο C. A. Bean & R. Green (Επιμ.), *Relationships in the Organization of Knowledge* (σελ. 37–52). Springer Netherlands. [https://doi.org/10.1007/978-94-015-9696-1\\_11](https://doi.org/10.1007/978-94-015-9696-1_11)
- Clifford, J., & Marcus, G. E. (Επιμ.). (1986). *Writing culture: The poetics and politics of ethnography*. University of California Press.
- Cliggett, L. (2013). Qualitative Data Archiving in the Digital Age: Strategies for Data Preservation and Sharing. *The Qualitative Report*, 18(24), 1–11.
- Combi, M. (1992). The imaginary, the computer, artificial intelligence: A cultural anthropological approach. *AI & Society*, 6(1), 41–49. <https://doi.org/10.1007/BF02472768>
- Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology* (2nd ed). University of Chicago Press.
- Cukier, K., & Mayer-Schoenberger, V. (2013). The Rise of Big Data: How it's Changing the Way We Think about the World. *Foreign Affairs*, 92, 28–40.
- Curran, J. (2013). Big Data or 'Big Ethnographic Data'? Positioning Big Data within the ethnographic space. *Ethnographic Praxis in Industry Conference Proceedings*, 2013(1), 62–73. <https://doi.org/10.1111/j.1559-8918.2013.00006.x>
- Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies. *Current Anthropology*, 60(1), 47–69. <https://doi.org/10.1086/701478>
- Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a Knowledge Representation? *AI Magazine*, 14(1), 17–33.
- De Giorgis, S., Gangemi, A., & Damiano, R. (2022). Basic Human Values and Moral Foundations Theory in ValueNet Ontology. Στο O. Corcho, L. Hollink, O. Kutz, N. Troquard, & F. J. Ekaputra (Επιμ.), *Knowledge Engineering and Knowledge Management* (σελ. 3–18). Springer International Publishing. [https://doi.org/10.1007/978-3-031-17105-5\\_1](https://doi.org/10.1007/978-3-031-17105-5_1)
- de Koning, M., Meyer, B., Moors, A., & Pels, P. (2019). Guidelines for anthropological research: Data management, ethics and integrity. *Ethnography*, 20(2), 170–174. <https://doi.org/10.1177/1466138119843312>
- Decoteau, C. L. (2017). The AART of Ethnography: A Critical Realist Explanatory Research Model. *Journal for the Theory of Social Behaviour*, 47(1), 58–82. <https://doi.org/10.1111/jtsb.12107>
- Delli Paoli, A., & D'Auria, V. (2025). The Digitalization of Ethnography: A Scoping Review of Methods in Netnography. *Journal of Contemporary Ethnography*, [in press]. <https://doi.org/10.1177/08912416251342795>
- Doerr, M. (2001). Semantic Problems of Thesaurus Mapping. *Journal of Digital Information*, 1(8). <http://journals.tdl.org/jodi/article/viewArticle/31/32>
- Doerr, M. (2003). The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3), 75. <https://doi.org/10.1609/aimag.v24i3.1720>

- Doerr, M. (2005). Modelling Learning Subjects as Relationships. Στο G. Grieser & Y. Tanaka (Επιμ.), *Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets* (τ. 3359, σελ. 201–214). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-32279-5\\_14](https://doi.org/10.1007/978-3-540-32279-5_14)
- Doerr, M. (2025). Identifiable Individuals and Reality: Describing the Past by Formal Propositions. *Journal on Computing and Cultural Heritage*, 3715011. <https://doi.org/10.1145/3715011>
- Doerr, M., & Iorizzo, D. (2008). The dream of a global knowledge network: A new approach. *Journal on Computing and Cultural Heritage*, 1(1), 5:1-5:23. <https://doi.org/10.1145/1367080.1367085>
- Dohan, D., & Sánchez-Jankowski, M. (1998). Using Computers to Analyze Ethnographic Field Data: Theoretical and Practical Considerations. *Annual Review of Sociology*, 24(1), 477–498. <https://doi.org/10.1146/annurev.soc.24.1.477>
- Dourish, P., & Gómez Cruz, E. (2018). Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society*, 5(2), 1–10. <https://doi.org/10.1177/2053951718784083>
- Edwards, J., & Kaimal, G. (2016). Using meta-synthesis to support application of qualitative methods findings in practice: A discussion of meta-ethnography, narrative synthesis, and critical interpretive synthesis. *The Arts in Psychotherapy*, 51, 30–35. <https://doi.org/10.1016/j.aip.2016.07.003>
- Ekbia, H. R., & Nardi, B. A. (2018). From Form to Content. *Cultural Anthropology*, 33(3), 360–367. <https://doi.org/10.14506/ca33.3.02>
- Ember, C. R. (2007). Using the HRAF Collection of Ethnography in Conjunction With the Standard Cross-Cultural Sample and the Ethnographic Atlas. *Cross-Cultural Research*, 41(4), 396–427. <https://doi.org/10.1177/1069397107306593>
- Emeneau, M. B. (Murray B. (1941). Language And Social Forms: A Study Of Toda Kinship And Dual Descent. Στο *Language, Culture And Personality: Essays In Memory Of Edward Sapir* (σελ. 158–179). Sapir Memorial Publication Fund. <https://ehrafworldcultures.yale.edu/document?id=aw60-012>
- Emmelhainz, C. (2014). Controlled Vocabulary Standards for Anthropological Datasets. *International Journal of Digital Curation*, 9(1), 185–192. <https://doi.org/10.2218/ijdc.v9i1.290>
- Engerer, V. (2017). Control and syntagmatization: Vocabulary requirements in information retrieval thesauri and natural language lexicons. *Journal of the Association for Information Science and Technology*, 68(6), 1480–1490. <https://doi.org/10.1002/asi.23783>
- Everett, D. L. (2009). *Don't Sleep, There Are Snakes: Life and Language in the Amazonian Jungle*. Vintage.
- Fauconnier, G., & Turner, M. (2003). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.
- Feyerabend, P. (2009). *Against method* (3rd ed.). Verso.
- Fischer, M. (2004). Integrating Anthropological Approaches to the Study of Culture: The “hard” and the “soft”. *Cybernetics and Systems*, 35(2–3), 147–162. <https://doi.org/10.1080/01969720490426830>
- Fischer, M. D. (χ.χ.). *Representing Anthropological Knowledge: Calculating Kinship [Prolog]*. Ανακτήθηκε 14 Απρίλιος 2025, από <https://era.anthropology.ac.uk/Kinship/prolog.html>
- Fischer, M. D. (1994). *Applications in Computing for Social Anthropologists* (edition published in the Taylor&Francis e-Library, 2005). Routledge.

- Fischer, M. D. (2006). Introduction: Configuring Anthropology. *Social Science Computer Review*, 24(1), 3–14. <https://doi.org/10.1177/0894439305282575>
- Fischer, M. D., & Ember, C. (2018). Big Data and Research Opportunities Using HRAF Databases. Στο C. Shu-Heng (Επιμ.), *Big Data in Computational Social Science and Humanities*. Springer. [http://dx.doi.org/10.1007/978-3-319-95465-3\\_17](http://dx.doi.org/10.1007/978-3-319-95465-3_17)
- Fischer, M. D., & Finkelstein, A. (1991). Social knowledge representation: A case study. Στο N. Fielding & R. Lee (Επιμ.), *Using computers in qualitative research* (σελ. 119–135). SAGE.
- Fischer, M. D., Lyon, S. M., Sosna, D., & Henig, D. (2013). Harmonizing Diversity Tuning Anthropological Research to Complexity. *Social Science Computer Review*, 31(1), 3–15. <https://doi.org/10.1177/0894439312455311>
- Fortun, M., Fortun, K., & Marcus, G. E. (2017). Computers in/and anthropology: The poetics and politics of digitization. Στο L. Hjorth, H. A. Horst, A. Galloway, & G. Bell (Επιμ.), *The Routledge companion to digital ethnography* (σελ. 11–20). Routledge, Taylor & Francis Group. <https://www.routledge.com/The-Routledge-Companion-to-Digital-Ethnography-1st-Edition/Hjorth-Horst-Galloway-Bell/p/book/9781138940918>
- Foucault, M. (1986). *Οι λέξεις και τα πράγματα: Μια αρχαιολογία των επιστημών του ανθρώπου*. Γνώση.
- Frické, M. (2019). The Knowledge Pyramid: The DIKW Hierarchy. *Knowledge Organization*, 46(1), 33–46. <https://doi.org/10.5771/0943-7444-2019-1-33>
- Fuentes, A., & Wiessner, P. (2016). Reintegrating Anthropology: From Inside Out: An Introduction to Supplement 13. *Current Anthropology*, 57(S13), S3–S12. <https://doi.org/10.1086/685694>
- Galeano, E. (2008). *Καθρέφτες: Μια σχεδόν παγκόσμια ιστορία* (Ι. Κανσή, Μτφ.). Πάπυρος Εκδοτικός Οργανισμός.
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (2002). Sweetening Ontologies with DOLCE. Στο A. Gómez-Pérez & V. R. Benjamins (Επιμ.), *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web* (σελ. 166–181). Springer. [https://doi.org/10.1007/3-540-45810-7\\_18](https://doi.org/10.1007/3-540-45810-7_18)
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). *Retrieval-Augmented Generation for Large Language Models: A Survey* (No. arXiv:2312.10997). arXiv. <https://doi.org/10.48550/arXiv.2312.10997>
- Geertz, C. (2003). *Η ερμηνεία των πολιτισμών* (Θ. Παραδέλλης, Μτφ.). Αλεξάνδρεια.
- Geiger, R. S., & Ribes, D. (2011). Trace Ethnography: Following Coordination through Documentary Practices. *2011 44th Hawaii International Conference on System Sciences (HICSS)*, 1–10. <https://dx.doi.org/10.1109/HICSS.2011.455>
- Goguen, J. A. Jr. (1974). Concept representation in natural and artificial languages: Axioms, extensions and applications for fuzzy sets. *International Journal of Man-Machine Studies*, 6(5), 513–561. [https://doi.org/10.1016/S0020-7373\(74\)80017-9](https://doi.org/10.1016/S0020-7373(74)80017-9)
- Gorichanaz, T., & Latham, K. F. (2016). Document phenomenology: A framework for holistic analysis. *Journal of Documentation*, 72(6), 1114–1133. <https://doi.org/10.1108/JD-01-2016-0007>
- Graeber, D., & Wengrow, D. (2022). *The Dawn of Everything: A New History of Humanity*. Penguin.

- Green, R. (2001). Relationships in the Organization of Knowledge: An Overview. Στο C. A. Bean & R. Green (Επιμ.), *Relationships in the Organization of Knowledge* (σελ. 3–18). Springer Netherlands. [https://doi.org/10.1007/978-94-015-9696-1\\_1](https://doi.org/10.1007/978-94-015-9696-1_1)
- Green, R., & Panzer, M. (2011). Relations in the Notational Hierarchy of the Dewey Decimal Classification. Στο A. Slavic & E. Civalero (Επιμ.), *Classification & ontology: Formal approaches and access to knowledge: Proceedings of the international UDC seminar 19-20 September 2011, The Hague, the Netherlands, organized by UDC Consortium, The Hague* (σελ. 161–176). Ergon.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5–6), 907–928. <https://doi.org/10.1006/ijhc.1995.1081>
- Gruber, T. R. (2007). Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems*, 3(1), 1–11.
- Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, 43(5–6), 625–640. <https://doi.org/10.1006/ijhc.1995.1066>
- Hakken, D. (1993). Computing and Social Change: New Technology and Workplace Transformation, 1980-1990. *Annual Review of Anthropology*, 22, 107–132.
- Haron, H., & Hamiz, M. (2014a). An Ontological Framework to Preserve Malay Indigenous Health Knowledge. *Advanced Science Letters*, 20(1), 226–230. <https://doi.org/10.1166/asl.2014.5261>
- Haron, H., & Hamiz, M. (2014b). An Ontological Model for Indigenous Knowledge of Malay Confinement Dietary. *Journal of Software*, 9(5), 1302–1312. <https://doi.org/10.4304/jsw.9.5.1302-1312>
- Harper, C. A. (2006). Encoding Library of Congress Subject Headings in SKOS: Authority control for the Semantic Web. *Proceedings of the 2006 International Conference on Dublin Core and Metadata Applications*. International Conference on Dublin Core and Metadata Applications. <http://dcpapers.dublincore.org/pubs/article/view/842>
- Harris, M. (1976). History and significance of the emic/etic distinction. *Annual Review of Anthropology*, 5(1), 329–350. <https://doi.org/10.1146/annurev.an.05.100176.001553>
- Headland, T. N., Pike, K., & Harris, M. (Επιμ.). (1990). *Emics and Etics: The Insider/Outsider Debate*. SAGE Publications, Inc.
- Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool.
- Heiner-Freiling, M. (2000). Survey on Subject Heading Languages Used in National Libraries and Bibliographies. *Cataloging & Classification Quarterly*, 29(1–2), 189–198. [https://doi.org/10.1300/J104v29n01\\_13](https://doi.org/10.1300/J104v29n01_13)
- Henley, J. (2019, Ιούλιος 20). The Norwegian island that abolished time: ‘You can cut the lawn at 4am’. *The Guardian*. <https://www.theguardian.com/world/2019/jun/20/sommaroy-island-norway-attempt-create-first-time-free-zone>
- Hine, C. M. (2000). *Virtual Ethnography*. SAGE Publications Ltd.
- Hitzler, P., Eberhart, A., Ebrahimi, M., Sarker, M. K., & Zhou, L. (2022). Neuro-symbolic approaches in artificial intelligence. *National Science Review*, 9(6), nwac035. <https://doi.org/10.1093/nsr/nwac035>

- Hjørland, B. (1992). The Concept of 'Subject' in Information Science. *Journal of Documentation*, 48(2), 172–200. <https://doi.org/10.1108/eb026895>
- Hjørland, B. (2009). Concept Theory. *Journal of the American Society for Information Science and Technology*, 60(8), 1519–1536. <https://doi.org/10.1002/asi.21082>
- Hjørland, B. (2012). Is Classification Necessary After Google? *Journal of Documentation*, 68(3), 299–317. <https://doi.org/10.1108/00220411211225557>
- Hjørland, B., & Pedersen, K. N. (2005). A Substantive Theory of Classification for Information Retrieval. *Journal of Documentation*, 61(5), 582–597. <https://doi.org/10.1108/00220410510625804>
- Hoeppe, G. (2015). Representing Representation. *Science, Technology & Human Values*, 40(6), 1077–1092. <https://doi.org/10.1177/0162243915594025>
- Holbraad, M., & Pedersen, M. A. (2017). *The Ontological Turn: An Anthropological Exposition*. Cambridge University Press.
- Hymes, D. (Επιμ.). (1965). *The Use of Computers in Anthropology* (First Edition edition). Mouton.
- Ingold, T. (2017). Anthropology contra ethnography. *HAU: Journal of Ethnographic Theory*, 7(1), 21–26. <https://doi.org/10.14318/hau7.1.005>
- ISO. (2006). *ISO 21127: Information and documentation: A reference ontology for the interchange of cultural heritage information*. ISO.
- ISO. (2011). *ISO 25964-1: Information and documentation—Thesauri and interoperability with other vocabularies—Part 1: Thesauri for information retrieval* (No. ISO 25964-1:2011(E); σελ. 160). International Organization for Standardization (ISO).
- ISO. (2023). *ISO 21127: Information and documentation: A reference ontology for the interchange of cultural heritage information*. ISO. <https://www.iso.org/standard/85100.html>
- Jacob, E. K. (2004). Classification and Categorization: A Difference that Makes a Difference. *Library Trends*, 52(3), 515–540.
- Jain, P., Hitzler, P., Yeh, P. Z., Verma, K., & Sheth, A. (2010). Linked Data is merely more data. Στο D. Brickley, V. K. Chaudhri, H. Halpin, & D. McGuinness (Επιμ.), *Linked Data Meets Artificial Intelligence* (σελ. 82–86). AAAI Press. [http://knoesis.wright.edu/library/publications/linkedai2010\\_submission\\_13.pdf](http://knoesis.wright.edu/library/publications/linkedai2010_submission_13.pdf)
- Johansson, I. (2008). Four Kinds of Is\_a Relation. Στο K. Munn & B. Smith (Επιμ.), *Applied Ontology: An Introduction* (σελ. 235–254). De Gruyter.
- Jupp, S., Bechhofer, S., & Stevens, R. (2009). *SKOS with OWL: Don't be full-ish!* 432. Scopus.
- Jürgens, P., Stark, B., & Magin, M. (2019). Two Half-Truths Make a Whole? On Bias in Self-Reports and Tracking Data. *Social Science Computer Review*, 0894439319831643. <https://doi.org/10.1177/0894439319831643>
- Karpouzis, K. (2025). *AI, Digital Humanities, and the Legacies of Colonial Power* (No. 2025021823). Preprints. <https://doi.org/10.20944/preprints202502.1823.v1>
- Kelder, J.-A. (2005). Using Someone Else's Data: Problems, Pragmatics and Provisions. *Forum: Qualitative Social Research*, 6(1). <http://www.qualitative-research.net/index.php/fqs/article/view/501>
- Kemper, J., & Kolkman, D. (2018). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 0(0), 1–16. <https://doi.org/10.1080/1369118X.2018.1477967>

- Keuschnigg, M., Lovsjö, N., & Hedström, P. (2017). Analytical sociology and computational social science. *Journal of Computational Social Science*, 1–12. <https://doi.org/10.1007/s42001-017-0006-5>
- Kim, Y., & Adler, M. (2015). Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management*, 35(4), 408–418. <https://doi.org/10.1016/j.ijinfomgt.2015.04.007>
- Kippen, J. (1988). On the Uses of Computers in Anthropological Research. *Current Anthropology*, 29(2), 317–320.
- Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., Leeher, D., Low, B. S., McCarter, J., Divale, W., & Gavin, M. C. (2016). D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. *PLOS ONE*, 11(7), e0158391. <https://doi.org/10.1371/journal.pone.0158391>
- Kless, D., Jansen, L., & Milton, S. (2016). A content-focused method for re-engineering thesauri into semantically adequate ontologies using OWL. *Semantic Web*, 7(5), 543–576. <https://doi.org/10.3233/SW-150194>
- Kless, D., Lindenthal, J., Milton, S., & Kazmierczak, E. (2011). Interoperability of knowledge organization systems with and through ontologies. Στο A. Slavic & E. Civalero (Επιμ.), *Classification & ontology: Formal approaches and access to knowledge: Proceedings of the international UDC seminar 19-20 September 2011, The Hague, the Netherlands, organized by UDC Consortium, The Hague* (σελ. 55–74). Ergon.
- Kless, D., Milton, S., Kazmierczak, E., & Lindenthal, J. (2015). Thesaurus and ontology structure: Formal and pragmatic differences and similarities. *Journal of the Association for Information Science and Technology*, 66(7), 1348–1366. <https://doi.org/10.1002/asi.23268>
- Kohne, J. (2014). Ontology, Its Origins and Its Meaning in Information Science. Στο R. Hagengruber & U. Riss (Επιμ.), *Philosophy, computing and information science* (σελ. 85–89). Pickering & Chatto.
- Kopytoff, I. (1964). Family And Lineage Among The Suku Of The Congo. Στο *The Family Estate In Africa* (σελ. 83–116). Routledge & K. Paul. <https://ehrafworldcultures.yale.edu/document?id=fo46-002>
- Kuhn, T. (2014). A Survey and Classification of Controlled Natural Languages. *Computational Linguistics*, 40(1), 121–170. [https://doi.org/10.1162/COLI\\_a\\_00168](https://doi.org/10.1162/COLI_a_00168)
- Kuhn, T. S. (1970). *The structure of scientific revolutions* ([2d ed., enl]). University of Chicago Press.
- Kuper, A. (1994). *Ανθρωπολογία και ανθρωπολόγοι: η σύγχρονη βρετανική σχολή* (Χ. Μιχαλοπούλου - Βεΐκου, Μτφ.). Καστανιώτης.
- Kwaśnik, B. H. (2011). Approaches to Providing Context in Knowledge Representation Structures. Στο A. Slavic & E. Civalero (Επιμ.), *Classification & ontology: Formal approaches and access to knowledge: Proceedings of the international UDC seminar 19-20 September 2011, The Hague, the Netherlands, organized by UDC Consortium, The Hague* (σελ. 9–23). Ergon.
- Lachal, J., Revah-Levy, A., Orri, M., & Moro, M. R. (2017). Metasynthesis: An Original Method to Synthesize Qualitative Literature in Psychiatry. *Frontiers in Psychiatry*, 8, 269. <https://doi.org/10.3389/fpsy.2017.00269>

- Lakoff, G. (1990). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.
- Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By*. University of Chicago Press.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Alstytne, M. V. (2009). Computational Social Science. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>
- Lazer, D., & Radford, J. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology*, 43(1), 19–39. <https://doi.org/10.1146/annurev-soc-060116-053457>
- Leach, E. R. (1993). Πολιτισμός και επικοινωνία: Η λογική της διαπλοκής των συμβόλων: Μια εισαγωγή στη χρήση της στρουκτουραλιστικής ανάλυσης στην κοινωνική ανθρωπολογία (Θ. Παραδέλλης, Επιμ.; Έ. Σουρράπα, Μτφ.). Καστανιώτης.
- Leetaru, K. H., Perkins, T., & Rewerts, C. (2014). Cultural Computing at Literature Scale: Encoding the Cultural Knowledge of Tens of Billions of Words of Academic Literature. *D-Lib Magazine*, 20(9/10). <https://doi.org/10.1045/september2014-leetaru>
- L’Homme, M.-C., & Bernier-Colborne, G. (2012). Terms as labels for concepts, terms as lexical units: A comparative analysis in ontologies and specialized dictionaries. *Applied Ontology*, 7(4), 387–400. <https://doi.org/10.3233/AO-2012-0116>
- Lioma, C., & Ounis, I. (2008). A syntactically-based query reformulation technique for information retrieval. *Information Processing & Management*, 44(1), 143–162. <https://doi.org/10.1016/j.ipm.2006.12.005>
- Lippolis, A. S., Saeedizade, M. J., Keskiarkka, R., Gangemi, A., Blomqvist, E., & Nuzzolese, A. G. (2025). *Assessing the Capability of Large Language Models for Domain-Specific Ontology Generation* (No. arXiv:2504.17402). arXiv. <https://doi.org/10.48550/arXiv.2504.17402>
- Lippolis, A. S., Saeedizade, M. J., Keskiarkka, R., Zuppiroli, S., Ceriani, M., Gangemi, A., Blomqvist, E., & Nuzzolese, A. G. (2025). *Ontology Generation using Large Language Models* (No. arXiv:2503.05388). arXiv. <https://doi.org/10.48550/arXiv.2503.05388>
- Loizos, P. (1998). Ο θάνατος του θετικισμού είναι υπερβολή (P. Αστρινάκη, Μτφ.). Στο *Ανθρωπολογική θεωρία και Εθνογραφία: Σύγχρονες τάσεις* (4η έκδ., σελ. 273–295). Ελληνικά Γράμματα.
- Lowrie, I. (2018). Algorithms and Automation: An Introduction. *Cultural Anthropology*, 33(3), 349–359. <https://doi.org/10.14506/ca33.3.01>
- Lyon, S. M. (2013). Networks and Kinship Formal Models of Alliance, Descent, and Inheritance in a Pakistani Punjabi Village. *Social Science Computer Review*, 31(1), 45–55. <https://doi.org/10.1177/0894439312453275>
- Lyon, S. M., & Magliveras, S. S. (2006). Kinship, Computing, and Anthropology. *Social Science Computer Review*, 24(1), 30–42. <https://doi.org/10.1177/0894439305281494>
- Machado, L. M. O., Souza, R. R., & Simões, M. da G. (2019). Semantic web or web of data? A diachronic study (1999 to 2017) of the publications of Tim Berners-Lee and the world wide web consortium. *Journal of the Association for Information Science and Technology*, 70(7), 701–714. <https://doi.org/10.1002/asi.24111>
- Mader, C., Haslhofer, B., & Isaac, A. (2012). Finding Quality Issues in SKOS Vocabularies. Στο P. Zaphiris, G. Buchanan, E. Rasmussen, & F. Loizides (Επιμ.), *Theory and Practice of Digital Libraries*. Springer. [https://doi.org/10.1007/978-3-642-33290-6\\_25](https://doi.org/10.1007/978-3-642-33290-6_25)

- Mai, J.-E. (1999). A Postmodern Theory of Knowledge Organization. Στο L. Woods (Επιμ.), *ASIS '99: Proceedings of the 62nd ASIS Annual Meeting, Washington, DC, October 31-November 4, 1999: Knowledge, Creation, Organization and Use* (σελ. 547–556). Information Today.
- Mai, J.-E. (2011). The modernity of classification. *Journal of Documentation*, 67(4), 710–730. <https://doi.org/10.1108/002204111111145061>
- Malinowski, B. (2014). *Argonauts of the Western Pacific*. Routledge.
- Maltese, V., & Farazi, F. (2011). Towards the integration of knowledge organization systems with the linked data cloud. Στο A. Slavic & E. Civalero (Επιμ.), *Classification & ontology: Formal approaches and access to knowledge: Proceedings of the international UDC seminar 19-20 September 2011, The Hague, the Netherlands, organized by UDC Consortium, The Hague* (σελ. 75–90). Ergon.
- Manganaro, M. (Επιμ.). (1990). *Modernist Anthropology: From Fieldwork to Text*. Princeton University Press.
- Marcus, G. E. (1998). The once and future ethnographic archive. *History of the Human Sciences*, 11(4), 49–63. <https://doi.org/10.1177/095269519801100404>
- Marcus, G. E. (2007). Ethnography Two Decades after Writing Culture: From the Experimental to the Baroque. *Anthropological Quarterly*, 80(4), 1127–1145. JSTOR.
- Margolis, E., & Laurence, S. (2011). Concepts. Στο E. N. Zalta & U. Nodelman (Επιμ.), *The Stanford Encyclopedia of Philosophy* (Spring 2014). <http://plato.stanford.edu/archives/spr2014/entries/concepts/>
- Martínez, E., Mollica, F., & Gibson, E. (2022). Poor writing, not specialized concepts, drives processing difficulty in legal language. *Cognition*, 224, 105070. <https://doi.org/10.1016/j.cognition.2022.105070>
- Mastora, A., Peponakis, M., & Kapidakis, S. (2017). SKOS Concepts and Natural Language Concepts: An Analysis of Latent Relationships in KOSs. *Journal of Information Science*, 43(4), 492–508. <https://doi.org/10.1177/0165551516648108>
- Mazzocchi, F. (2017). Relations in KOS: Is it possible to couple a common nature with different roles? *Journal of Documentation*, 73(2), 368–383. <https://doi.org/10.1108/JD-05-2016-0063>
- McKinnon, S. (2009). *Νεοφιλελεύθερη γενετική: Μύθοι και ηθικού περιεχομένου ιστορίες της εξελικτικής ψυχολογίας* (Θ. Παραδέλλης, Μτφ.). Εκδόσεις Εικοστού Πρώτου.
- McNamara, P. (2006). Deontic logic. Στο D. M. Gabbay & J. Woods (Επιμ.), *Handbook of the History of Logic* (τ. 7, σελ. 197–288). North-Holland. [https://doi.org/10.1016/S1874-5857\(06\)80029-4](https://doi.org/10.1016/S1874-5857(06)80029-4)
- McNamara, P., & Van De Putte, F. (2022). Deontic Logic. Στο E. N. Zalta & U. Nodelman (Επιμ.), *The Stanford Encyclopedia of Philosophy* (Fall 2022). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2022/entries/logic-deontic/>
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014), 176–182. <https://doi.org/10.1126/science.1199644>
- Mitchell, J. S., Zeng, M. L., & Žumer, M. (2014). Modeling Classification Systems in Multicultural and Multilingual Contexts. *Cataloging & Classification Quarterly*, 52(1), 90–101. <https://doi.org/10.1080/01639374.2013.845620>

- Moore, R. J., Smith, R., & Liu, Q. (2020). Using computational ethnography to enhance the curation of real-world data (RWD) for chronic pain and invisible disability use cases. *ACM SIGACCESS Accessibility and Computing*, 127, 4:1-4:7. <https://doi.org/10.1145/3412836.3412840>
- Moretti, F. (2013). *Distant Reading*. Verso Books.
- Mueller, A. (2016). Beyond ethnographic scriptocentrism: Modelling multi-scalar processes, networks, and relationships. *Anthropological Theory*, 16(1), 98–130. <https://doi.org/10.1177/1463499615626621>
- Murdock, G. P. (1967). *Ethnographic Atlas*. University of Pittsburgh Press.
- Murdock, G. P., & White, D. R. (1969). Standard Cross-Cultural Sample. *Ethnology*, 8(4), 329–369. <https://doi.org/10.2307/3772907>
- Murillo, L. F. R. (2018). What Does “Open Data” Mean for Ethnographic Research? *American Anthropologist*, 120(3), 577–582. <https://doi.org/10.1111/aman.13088>
- Murphy, G. L. (2002). *The Big Book of Concepts*. MIT Press.
- Nafus, D. (2018). Exploration or Algorithm? The Undone Science Before the Algorithms. *Cultural Anthropology*, 33(3), 368–374. <https://doi.org/10.14506/ca33.3.03>
- Needham, R. (2004). Remarks on the analysis of kinship and marriage. Στο *Rethinking Kinship and Marriage* (σελ. 1–34). Routledge.
- Nelson, L. K. (2017). Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research*. <https://doi.org/10.1177/0049124117729703>
- Noack, R. (2015, Φεβρουάριος 24). Out of fear of racism, Sweden changes the names of bird species. *The Washington Post*. <http://www.washingtonpost.com/blogs/worldviews/wp/2015/02/24/out-of-fear-of-racism-sweden-changes-the-names-of-bird-species/>
- Noakes, T., Harpur, P., & Uys, C. (2023). Noteworthy Disparities With Four CAQDAS Tools: Explorations in Organising Live Twitter Data. *Social Science Computer Review*, 08944393231204163. <https://doi.org/10.1177/08944393231204163>
- Novak, L., & Harris, J. (2020). Anthropology and Informatics in Health Care. Στο L. Novak & J. Harris, *Oxford Research Encyclopedia of Anthropology*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190854584.013.6>
- Nowroozi, M., Mirzabeigi, M., & Sotudeh, H. (2018). The comparison of thesaurus and ontology: Case of ASIS&T web-based thesaurus and designed ontology. *Library Hi Tech*. <https://doi.org/10.1108/LHT-03-2017-0060>
- O’Kane, P. (2020). Demystifying CAQDAS: A Series of Dilemmas. Στο T. Russell Crook, J. Lê, & A. D Smith (Επιμ.), *Advancing Methodological Thought and Practice* (τ. 12, σελ. 133–152). Emerald Publishing Limited. <https://doi.org/10.1108/S1479-838720200000012020>
- Oldman, D., Doerr, M., & Gradmann, S. (2016). Zen and the Art of Linked Data. Στο S. Schreibman, R. Siemens, & J. Unsworth (Επιμ.), *A New Companion to Digital Humanities* (σελ. 251–273). John Wiley & Sons, Ltd. <http://onlinelibrary.wiley.com/doi/10.1002/9781118680605.ch18/summary>
- Olson, H. A. (2000). Difference, Culture and Change: The Untapped Potential of LCSH. *Cataloging & Classification Quarterly*, 29(1–2), 53–71. [https://doi.org/10.1300/J104v29n01\\_04](https://doi.org/10.1300/J104v29n01_04)
- Olson, H. A. (2001). The Power to Name: Representation in Library Catalogs. *Signs*, 26(3), 639–668.

- O'Neil, D. (1987). Anthropological Computing in the Mid-1980s. *Social Science Microcomputer Review*, 5(4), 452–465. <https://doi.org/10.1177/089443938700500403>
- Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 36(7), 3580–3599. <https://doi.org/10.1109/TKDE.2024.3352100>
- Panzer, M. (2008). Cool URIs for the DDC: Towards Web-scale Accessibility of a Large Classification System. Στο J. Greenberg & K. Wolfgang (Επιμ.), *Metadata for Semantic and Social Applications: Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications* (σελ. 183–190). <http://dcpapers.dublincore.org/pubs/article/view/932>
- Papakyriakopoulos, O., & Mboya, A. M. (2022). Beyond Algorithmic Bias: A Socio-Computational Interrogation of the Google Search by Image Algorithm. *Social Science Computer Review*, 08944393211073169. <https://doi.org/10.1177/08944393211073169>
- Pattueli, M. C., Provo, A., & Thorsen, H. (2015). Ontology Building for Linked Open Data: A Pragmatic Perspective. *Journal of Library Metadata*, 15(3–4), 265–294. <https://doi.org/10.1080/19386389.2015.1099979>
- Pels, P., Boog, I., Florusbosch, J. H., Kripe, Z., Minter, T., Postma, M., Sleeboom-Faulkner, M., Simpson, B., Dilger, H., Schönhuth, M., Poser, A. von, Castillo, R. C. A., Lederman, R., & Richards-Rissetto, H. (2018). Data management in anthropology: The next phase in ethics governance? *Social Anthropology*, 26(3), 391–413. <https://doi.org/10.1111/1469-8676.12526>
- Peponakis, M. (2016). In the Name of the Name: RDF literals, ER Attributes and the Potential to Rethink the Structures and Visualizations of Catalogs. *Information Technology and Libraries*, 35(2), 19–38. <https://doi.org/10.6017/ital.v35i2.8749>
- Peponakis, M., Kapidakis, S., Doerr, M., & Tountasaki, E. (2024). From calculations to reasoning: History, trends and the potential of Computational Ethnography and Computational Social Anthropology. *Social Science Computer Review*, 42(1), 84–102. <https://doi.org/10.1177/08944393231167692>
- Peponakis, M., Mastora, A., Kapidakis, S., & Doerr, M. (2019). Expressiveness and machine processability of Knowledge Organization Systems (KOS): An analysis of concepts and relations. *International Journal on Digital Libraries*, 20(4), 433–452. <https://doi.org/10.1007/s00799-019-00269-0>
- Perec, G. (2005). *Σκέψη—Ταξινόμηση* (Λ. Τσιριμώκου, Μτφ.). Άγρα.
- Peters, I., & Weller, K. (2008). Paradigmatic and syntagmatic relations in knowledge organization systems. *Information Wissenschaft und Praxis*, 59(2), 100–107.
- Phefo, O. S. D., Kefitiley, N., & Hlomani, H. (2015). Towards the Cultural Knowledge Ontology. *2015 IEEE International Conference on Information Reuse and Integration (IRI)*, 526–533. <https://doi.org/10.1109/IRI.2015.85>
- Pink, S., & Lanzeni, D. (2018). Future Anthropology Ethics and Datafication: Temporality and Responsibility in Research. *Social Media + Society*, 4(2), 1–9. <https://doi.org/10.1177/2056305118768298>
- Pinker, S. (2007). *The Language Instinct: How the Mind Creates Language* (third edition). Harper Perennial Modern Classics.
- Pirkola, A. (2001). Morphological typology of languages for IR. *Journal of Documentation*, 57(3), 330–348. <https://doi.org/10.1108/EUM0000000007085>

- Poincaré, H. (1952). *Science and Hypothesis*. Courier Corporation.
- Pokraev, S., Reichert, M., Steen, M. W. A., & Wieringa, R. J. (2005). *Semantic and pragmatic interoperability: A model for understanding*. 160. Scopus. <http://ceur-ws.org/Vol-160/paper21.pdf>
- Polanyi, M. (1997). *Personal knowledge: Towards a post-critical philosophy*. Routledge.
- Poli, R., & Obrst, L. (2010). The Interplay Between Ontology as Categorical Analysis and Ontology as Technology. Στο R. Poli, M. Healy, & A. Kameas (Επιμ.), *Theory and Applications of Ontology: Computer Applications* (σελ. 1–26). Springer Netherlands. [http://link.springer.com/chapter/10.1007/978-90-481-8847-5\\_1](http://link.springer.com/chapter/10.1007/978-90-481-8847-5_1)
- Pool, R. (2017). The verification of ethnographic data. *Ethnography*, 18(3), 281–286. <https://doi.org/10.1177/1466138117723936>
- Portner, P. H. (2005). *What is Meaning?: Fundamentals of Formal Semantics*. Blackwell.
- Prokopidis, P., Georgantopoulos, B., & Papageorgiou, H. (2011). A suite of NLP tools for Greek. *Proceedings of the 10th International Conference of Greek Linguistics*. [http://nlp.ilsp.gr/nlp/ICGL2011\\_Prokopidis\\_etal.pdf](http://nlp.ilsp.gr/nlp/ICGL2011_Prokopidis_etal.pdf)
- Purzycki, B. G., & Jamieson-Lane, A. (2017). AnthroTools: An R Package for Cross-Cultural Ethnographic Data Analysis. *Cross-Cultural Research*, 51(1), 51–74. <https://doi.org/10.1177/1069397116680352>
- Read, D., Fischer, M., & Leaf, M. (2013). What Are Kinship Terminologies, and Why Do We Care? A Computational Approach to Analyzing Symbolic Domains. *Social Science Computer Review*, 31(1), 16–44. <https://doi.org/10.1177/0894439312455914>
- Read, D. W. (1990). Anthropology and computers: Promise and potential. *Social Science Computer Review*, 8(4), 503–514. <https://doi.org/10.1177/089443939000800402>
- Read, D. W. (2001). Formal analysis of kinship terminologies and its relationship to what constitutes kinship. *Anthropological Theory*, 1(2), 239–267. <https://doi.org/10.1177/146349960100100205>
- Read, D. W. (2006). Kinship Algebra Expert System (KAES) A Software Implementation of a Cultural Theory. *Social Science Computer Review*, 24(1), 43–67. <https://doi.org/10.1177/0894439305282372>
- Read, D. W. (2007). Kinship Theory: A Paradigm Shift. *Ethnology*, 46(4), 329–364.
- Read, D. W., & Behrens, C. A. (1990). KAES: An expert system for the algebraic analysis of kinship terminologies. *Journal of Quantitative Anthropology*, 2(4), 353–393.
- Reyes, V. (2017). Three models of transparency in ethnographic research: Naming places, naming people, and sharing data. *Ethnography*, 1466138117733754. <https://doi.org/10.1177/1466138117733754>
- Ribes, D., & Bowker, G. C. (2009). Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization*, 19(4), 199–217. <https://doi.org/10.1016/j.infoandorg.2009.04.001>
- Rips, L. J., Smith, E. E., & Medin, D. L. (2013). Concepts and Categories: Memory, Meaning, and Metaphysics. Στο K. J. Holyoak & R. G. Morrison (Επιμ.), *The Oxford Handbook of Thinking and Reasoning* (σελ. 177–209). Oxford University Press.
- Rohner, R. P., Berg, D. S., & Rohner, E. C. (1982). Data Quality Control in the Standard Cross-Cultural Sample: Cross-Cultural Codes. *Ethnology*, 21(4), 359–369. <https://doi.org/10.2307/3773766>

- Romele, A. (2020). The datafication of the worldview. *AI & SOCIETY*.  
<https://doi.org/10.1007/s00146-020-00989-x>
- Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as Consensus: A Theory of Culture and Informant Accuracy. *American Anthropologist*, 88(2), 313–338.  
<https://doi.org/10.1525/aa.1986.88.2.02a00020>
- Sahlins, M. (2002). *Waiting for Foucault, Still*. Prickly Paradigm Press.
- Salah, A. A., Gao, C., Suchecki, K., Scharnhorst, A., & Smiraglia, R. P. (2012). The evolution of classification systems: Ontogeny of the UDC. Στο A. Neelameghan & K. S. Raghavan (Επιμ.), *Categories, Contexts and Relations in Knowledge organization: Proceedings of the Twelfth International ISKO Conference, 6-9 August 2012, Mysore, India* (σελ. 51–57). Ergon.
- Santucci, J.-F., Doja, A., & Capocchi, L. (2020). A Discrete-Event Simulation of Claude Lévi-Strauss' Structural Analysis of Myths Based on Symmetry and Double Twist Transformations. *Symmetry*, 12(10), Article 10. <https://doi.org/10.3390/sym12101706>
- Schmidt, S., & Schweizer, T. (1990). Comparative Anthropological Research Today: The World Cultures Electronic Database and New Substantive Developments. *Anthropos*, 85(1/3), 195–201.
- Schneider, D. M. (1995). *Schneider on Schneider: The Conversion of the Jews and Other Anthropological Stories*. Duke University Press Books.
- Schroeder, R. (2019). Big data and cumulation in the social sciences. *Information, Communication & Society*, 1–15. <https://doi.org/10.1080/1369118X.2019.1594334>
- Schwarz, U., & Smith, B. (2008). Ontological Relations. Στο K. Munn & B. Smith (Επιμ.), *Applied Ontology An Introduction*. De Gruyter.  
<http://www.degruyter.com/view/books/9783110324860/9783110324860.219/9783110324860.219.xml>
- Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 1–12. <https://doi.org/10.1177/2053951717738104>
- Seaver, N. (2018). What Should an Anthropology of Algorithms Do? *Cultural Anthropology*, 33(3), 375–385. <https://doi.org/10.14506/ca33.3.04>
- Seidel, J. V., & Clark, J. A. (1984). THE ETHNOGRAPH: A computer program for the analysis of qualitative data. *Qualitative Sociology*, 7(1), 110–125.  
<https://doi.org/10.1007/BF00987111>
- Seth, A. (Διευθυντής). (2017). *Your brain hallucinates your conscious reality* [Video recording]. TED Talks. <https://youtu.be/lyu7v7nWzfo>
- Shannon, C. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.
- Shiri, A. (2014). Linked Data Meets Big Data: A Knowledge Organization Systems Perspective. *Advances in Classification Research Online*, 24(1), 16–20.  
<https://doi.org/10.7152/acro.v24i1.14672>
- Smiraglia, R., & Heuvel, C. V. den. (2013). Classifications and Concepts: Towards an Elementary Theory of Knowledge Interaction. *Journal of Documentation*, 69(3), 360–383.  
<https://doi.org/10.1108/JD-07-2012-0092>
- Smiraglia, R. P. (2014). *The Elements of Knowledge Organization*. Springer.  
<http://link.springer.com/10.1007/978-3-319-09357-4>

- Smith, B. (2015). *Basic Formal Ontology 2.0: Specification and user's guide* (σελ. 102). <https://raw.githubusercontent.com/BFO-ontology/BFO/master/docs/bfo2-reference/BFO2-Reference.pdf>
- Snow, C. P. (1961). *The two cultures and the scientific revolution*. Cambridge University Press.
- Sosna, D., Galeta, P., Šmejda, L., Sladek, V., & Bruzek, J. (2013). Burials and Graphs Relational Approach to Mortuary Analysis. *Social Science Computer Review*, 31(1), 56–70. <https://doi.org/10.1177/0894439312453277>
- Sowa, J. F. (2009). *Knowledge representation: Logical, philosophical, and computational foundations*. Brooks/Cole.
- Sperberg-McQueen, C. M. (2004). Classification and its Structures. Στο S. Schreibman, R. Siemens, & J. Unsworth (Επιμ.), *A Companion to Digital Humanities*. Blackwell. <http://www.digitalhumanities.org/companion/>
- Spero, S. E. (2008). LCSH is to Thesaurus as Doorbell is to Mammal: Visualizing Structural Problems in the Library of Congress Subject Headings. *Metadata For Semantic and Social Applications: Proceedings of the International Conference On Dublin Core and Metadata Applications*, 203. <http://dcpapers.dublincore.org/pubs/article/view/937>
- Starn, O. (2022). Anthropology and the misery of writing. *American Anthropologist*, 124(1), 187–197. <https://doi.org/10.1111/aman.13677>
- Stellato, A. (2012). Dictionary, Thesaurus or Ontology? Disentangling Our Choices in the Semantic Web Jungle. *Journal of Integrative Agriculture*, 11(5), 710–719. [https://doi.org/10.1016/S2095-3119\(12\)60060-4](https://doi.org/10.1016/S2095-3119(12)60060-4)
- Stone, A. T. (2000). The LCSH Century: A Brief History of the Library of Congress Subject Headings, and Introduction to the Centennial Essays. *Cataloging & Classification Quarterly*, 29(1–2), 1–15. [https://doi.org/10.1300/J104v29n01\\_01](https://doi.org/10.1300/J104v29n01_01)
- Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. *Handbook of qualitative research*, 273–285.
- Stubbersfield, J., & Tehrani, J. (2013). Expect the Unexpected? Testing for Minimally Counterintuitive (MCI) Bias in the Transmission of Contemporary Legends A Computational Phylogenetic Approach. *Social Science Computer Review*, 31(1), 90–102. <https://doi.org/10.1177/0894439312453567>
- Stump, G. T. (2001). *Inflectional morphology: A theory of paradigm structure*. Cambridge Univ. Press.
- Summers, E., Isaac, A., Redding, C., & Krech, D. (2008). LCSH, SKOS and Linked Data. Στο J. Greenberg & K. Wolfgang (Επιμ.), *Metadata for Semantic and Social Applications: Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications* (σελ. 25–33). Dublin Core Metadata Initiative. <http://dcpapers.dublincore.org/pubs/article/view/916/912>
- Svenonius, E. (2000). LCSH: Semantics, Syntax and Specificity. *Cataloging & Classification Quarterly*, 29(1–2), 17–30. [https://doi.org/10.1300/J104v29n01\\_02](https://doi.org/10.1300/J104v29n01_02)
- Szostak, R. (2011). Complex concepts into basic concepts. *Journal of the American Society for Information Science and Technology*, 62(11), 2247–2265. <https://doi.org/10.1002/asi.21635>

- Tallyn, E., Fried, H., Gianni, R., Isard, A., & Speed, C. (2018). The Ethnobot: Gathering Ethnographies in the Age of IoT. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 604:1-604:13. <https://doi.org/10.1145/3173574.3174178>
- TEDx Talks (Διευθυντής). (2011, Ιούλιος 14). *A picture is worth 500 billion words: Erez Lieberman Aiden and Jean-Baptiste Michel at TEDxBoston* [Video recording]. <https://youtu.be/WtJ50v7qByE>
- Tobin, J. (1990). The HRAF as Radical Text? *Cultural Anthropology*, 5(4), 473–487. <https://doi.org/10.1525/can.1990.5.4.02a00090>
- Totaro, P., & Ninno, D. (2014). The Concept of Algorithm as an Interpretative Key of Modern Rationality. *Theory, Culture & Society*, 31(4), 29–49. <https://doi.org/10.1177/0263276413510051>
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind: A quarterly review of psychology and philosophy*, 59(236), 433–460.
- Turner, A. (2019). Using Data from Git and GitHub in Ethnographies of Software Development. Στο C. Öhman & D. Watson (Επιμ.), *The 2018 Yearbook of the Digital Ethics Lab* (σελ. 35–49). Springer International Publishing. [https://doi.org/10.1007/978-3-030-17152-0\\_4](https://doi.org/10.1007/978-3-030-17152-0_4)
- Van Der Leeuw, S. E. (2004). Why Model? *Cybernetics and Systems*, 35(2–3), 117–128. <https://doi.org/10.1080/01969720490426803>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, 30. <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
- W3C. (2004). *OWL Web Ontology Language Reference*. <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>
- W3C. (2009a). *SKOS eXtension for Labels (SKOS-XL) Namespace Document—HTML Variant*. <http://www.w3.org/TR/2009/REC-skos-reference-20090818/skos-xl.html>
- W3C. (2009b). *SKOS Simple Knowledge Organization System Reference*. <http://www.w3.org/TR/skos-reference/>
- W3C. (2014a). *RDF 1.1 Concepts and Abstract Syntax: W3C Recommendation*. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>
- W3C. (2014b). *RDF Schema 1.1*. <https://www.w3.org/TR/2014/REC-rdf-schema-20140225/>
- Wacholder, N., & Liu, L. (2006). User preference: A measure of query-term quality. *Journal of the American Society for Information Science and Technology*, 57(12), 1566–1580. <https://doi.org/10.1002/asi.20315>
- Wagner, R. A. (1989). The Rise of Computing in Anthropology: Hammers and Nails. *Social Science Computer Review*, 7(4), 418–430. <https://doi.org/10.1177/089443938900700403>
- Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: A literature review. *Journal of Advanced Nursing*, 50(2), 204–211. <https://doi.org/10.1111/j.1365-2648.2005.03380.x>
- Warters, B. (2005). Software Review: Review of TAMS Analyzer (Macintosh Version). *Field Methods*, 17(3), 321–328. <https://doi.org/10.1177/1525822X05277533>
- Weber, F. (2017). Towards a digital architecture of reflexive ethnographic data. *Ethnography*, 18(3), 287–294. <https://doi.org/10.1177/1466138117724482>

- Weingart, S., & Jorgensen, J. (2013). Computational analysis of the body in European fairy tales. *Literary and Linguistic Computing*, 28(3), 404–416. <https://doi.org/10.1093/llc/fqs015>
- Weinstein, M. (2006). TAMS Analyzer: Anthropology as Cultural Critique in a Digital Age. *Social Science Computer Review*, 24(1), 68–77. <https://doi.org/10.1177/0894439305281496>
- Wesch, M. (Διευθυντής). (2007, Ιανουάριος 31). *Web 2.0 ... The Machine is Us/ing Us* [Video recording]. <https://youtu.be/6gmP4nk0EOE>
- Wesch, M. (2008, Ιούλιος 29). An anthropological introduction to YouTube. *Digital Ethnography*. <http://mediatedcultures.net/youtube/an-anthropological-introduction-to-youtube-presented-at-the-library-of-congress/>
- White, D. R. (1986). The World Cultures Database. *World Cultures eJournal*, 1(1). <https://escholarship.org/uc/item/3r21c3xb>
- Wilf, E. (2013). Toward an Anthropology of Computer-Mediated, Algorithmic Forms of Sociality. *Current Anthropology*, 54(6), 716–739. <https://doi.org/10.1086/673321>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), Article 1. <https://doi.org/10.1038/sdata.2016.18>
- Wilmont, I., Hengeveld, S., Barendsen, E., & Hoppenbrouwers, S. (2013). Cognitive Mechanisms of Conceptual Modelling. Στο W. Ng, V. C. Storey, & J. C. Trujillo (Επιμ.), *Conceptual Modeling* (σελ. 74–87). Springer. [http://link.springer.com/chapter/10.1007/978-3-642-41924-9\\_7](http://link.springer.com/chapter/10.1007/978-3-642-41924-9_7)
- Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>
- Wittgenstein, L. (1977). *Φιλοσοφικές έρευνες* (Π. Χριστοδουλίδης, Μτφ.). Εκδόσεις Παπαζήση.
- Wittgenstein, L. (1978). *Tractatus Logico-Philosophicus* (Θ. Κιτσόπουλος, Μτφ.). Εκδόσεις Παπαζήση.
- Wolfe, A. W. (1978). The rise of network thinking in anthropology. *Social Networks*, 1(1), 53–64. [https://doi.org/10.1016/0378-8733\(78\)90012-6](https://doi.org/10.1016/0378-8733(78)90012-6)
- Yoon, A., Kim, J., & Donaldson, D. R. (2025). Big data curation framework: Curation actions and challenges. *Journal of Information Science*, 51(1), 205–223. <https://doi.org/10.1177/01655515221133528>
- Zavalina, O. L. (2012). Subject Access: Conceptual Models, Functional Requirements, and Empirical Data. *Journal of Library Metadata*, 12(2–3), 140–163. <https://doi.org/10.1080/19386389.2012.699829>
- Zeng, M. L., & Mayr, P. (2019). Knowledge Organization Systems (KOS) in the Semantic Web: A multi-dimensional review. *International Journal on Digital Libraries*, 20(3), 209–230. <https://doi.org/10.1007/s00799-018-0241-2>
- Zheng, K., Hanauer, D. A., Weibel, N., & Agha, Z. (2015). Computational Ethnography: Automated and Unobtrusive Means for Collecting Data In Situ for Human–Computer Interaction Evaluation Studies. Στο V. Patel, T. Kannampallil, & D. Kaufman (Επιμ.), *Cognitive Informatics for Biomedicine* (σελ. 111–140). Springer, Cham. [https://doi.org/10.1007/978-3-319-17272-9\\_6](https://doi.org/10.1007/978-3-319-17272-9_6)

## Γλωσσάρι

Στο γλωσσάρι που ακολουθεί παρουσιάζεται το πώς έχει αποδοθεί ο αγγλικός όρος στην ελληνική γλώσσα εντός του κειμένου. Στις περιπτώσεις που ο όρος μπορεί να έχει διαφορετικό νόημα, ανάλογα με το επιστημονικό πεδίο στο οποίο απαντάται, τότε ακολουθείται από μια παρένθεση η οποία τον εντάσσει σε συγκεκριμένα συμφραζόμενα.

Αγγλικός όρος	Ελληνικός όρος
<b>artificial intelligence</b>	τεχνητή νοημοσύνη
<b>attribute</b>	χαρακτηριστικό
<b>cardinality</b>	πληθικότητα
<b>cenogamy ή group marriage</b>	κοινογαμία
<b>class</b>	κλάση
<b>cluster</b>	συστάδα
<b>computational anthropology</b>	υπολογιστική ανθρωπολογία
<b>computational ethnography</b>	υπολογιστική εθνογραφία
<b>concept</b>	έννοια
<b>conceptualization</b>	ενοιολόγηση
<b>context</b>	συμφραζόμενα
<b>dataset</b>	σύνολο δεδομένων
<b>deontic logic</b>	δεοντική λογική
<b>digital ethnography</b>	ψηφιακή εθνογραφία
<b>first order logic</b>	λογική πρώτου βαθμού
<b>genitor</b>	γεννήτωρ
<b>grounded theory</b>	θεμελιωμένη θεωρία
<b>informant</b>	πληροφορητής
<b>information integration</b>	ολοκλήρωση της πληροφορίας
<b>information science</b>	επιστήμη της πληροφορίας
<b>instance</b>	στιγμιότυπο
<b>intersubjectivity</b>	διυποκειμενικότητα
<b>individual</b>	άτομο
<b>knowledge graph</b>	γράφος γνώσης
<b>knowledge organization system</b>	σύστημα οργάνωσης της γνώσης

<b>knowledge representation</b>	αναπαράσταση γνώσης
<b>linked data</b>	διασυνδεδεμένα δεδομένα
<b>machine learning</b>	μηχανική μάθηση
<b>multi-site ethnography</b>	πολυτοπική εθνογραφία
<b>ontology</b>	οντολογία
<b>ontological turn</b>	οντολογική στροφή
<b>patterns</b>	μοτίβα
<b>polyandry</b>	πολυανδρία
<b>polygamy</b>	πολυγαμία
<b>polygyny</b>	πολυγυνία
<b>precision (information retrieval)</b>	ακρίβεια (ανάκτηση πληροφορίας)
<b>probabilistic</b>	πιθανοκρατικό
<b>property</b>	ιδιότητα
<b>query (information system)</b>	επερώτηση
<b>reasoning</b>	συλλογιστική ή συμπερασμός
<b>recall (information retrieval)</b>	ανάκληση (ανάκτηση πληροφορίας)
<b>reflexive property (ontology)</b>	ανακλαστική ιδιότητα
<b>reflexivity (anthropology)</b>	αναστοχασμός
<b>reflexivity (information science)</b>	ανακλαστικότητα
<b>relation or relationship</b>	σχέση
<b>semantic drift</b>	σημασιολογική μετατόπιση
<b>semantic web</b>	σημασιολογικός ιστός
<b>semantics</b>	σημασιολογία
<b>sociotechnical</b>	κοινωνικοτεχνικό
<b>string (information science)</b>	συμβολοσειρά
<b>symmetric property (ontology)</b>	συμμετρική ιδιότητα
<b>textualization</b>	κειμενοποίηση
<b>transitive property (ontology)</b>	μεταβατική ιδιότητα
<b>virtual ethnography</b>	εικονική εθνογραφία
<b>web of data</b>	ιστός των δεδομένων