

Ανάλυση των ρημάτων κίνησης στα εγχειρίδια νεοελληνικής γλώσσας και λογοτεχνίας Δημοτικού με τη χρήση Σωμάτων Κειμένων και Τεχνητής Νοημοσύνης

Αλεξάνδρα Φιωτάκη

Δρ. Γλωσσολογίας, Πανεπιστήμιο Ιωαννίνων

alexandra.fiotaki@gmail.com

Νίκος Μαθιουδάκης

Ξένος Λέκτορας, Πανεπιστήμιο Βελιγραδίου

nikosmathioudakis@gmail.com

Περίληψη

Η παρούσα μελέτη ασχολείται με το σημασιολογικό πεδίο της κίνησης, όπως αυτό αναπτύσσεται είτε ως ρήματα κίνησης είτε ως γεγονότα κίνησης, στα εγχειρίδια νεοελληνικής γλώσσας και νεοελληνικής λογοτεχνίας Ε' και ΣΤ' Δημοτικού. Σκοπός μας είναι αφενός η δημιουργία ενός σώματος κειμένων, αξιοποιώντας τα αυθεντικά κείμενα που υπάρχουν στα σχολικά εγχειρίδια, και αφετέρου η ανάλυση των κειμενικών δεδομένων αναφορικά με το σημασιολογικό πεδίο της κίνησης. Ως βασικό εργαλείο της έρευνάς μας αποτελεί το σώμα κειμένων που αφορά το "γεγονός κίνησης", το οποίο επισημειώνεται με στόχο την κωδικοποίηση των συστατικών του γεγονότος κίνησης. Η επισημείωση έγινε τόσο ως προς τα ρήματα κίνησης (π.χ. *φτάνω, έρχομαι, πηγαίνω*) όσο και ως προς τα συντακτικά και σημασιολογικά χαρακτηριστικά των ρημάτων που βοηθούν στη διάκριση σημασιών και στην ομαδοποίηση τους (επιρρηματικές και προθετικές εκφράσεις). Το σχήμα επισημείωσης σχεδιάστηκε σύμφωνα με την τυπολογία του Talmy, ο οποίος υποστηρίζει ότι η κίνηση ως γεγονός προσδιορίζεται από έξι σημασιολογικά χαρακτηριστικά: (α) την οντότητα που επιτελεί την κίνηση, (β) την ίδια την κίνηση ως γεγονός, (γ) τη διαδρομή, (δ) την οντότητα αναφοράς (ορόσημο, χαρακτηριστικό στοιχείο, τοποθεσία αναφορικά με την αρχική οντότητα που κινείται), (ε) τον τρόπο και (στ) την αιτία. Μέσα από τη διαδικασία επισημείωσης και εν συνεχεία από τη μελέτη του σώματος κειμένων, κατανοούμε τις σημασιολογικές αποχρώσεις, εμβαθύνοντας στην κατανόηση των διαφορετικών ειδών ρημάτων κίνησης με προοπτική τη δημιουργία ενός ψηφιακού γλωσσαρίου για το δημοτικό. Συμπληρωματικά, η μελέτη αυτή προορίζεται όχι μόνο για να φωτίσει τις λεπτομέρειες του σημασιολογικού πεδίου της κίνησης, αλλά και να αξιοποιηθεί για τη διαμόρφωση παιδαγωγικών εργαλείων που προωθούν τη γλωσσική κατανόηση του πεδίου από τα παιδιά δημοτικού.

Λέξεις-κλειδιά: Σώμα κειμένων, ρήματα κίνησης, εγχειρίδια δημοτικού, αυθεντικά κείμενα, νεοελληνική γλώσσα.

Summary

The present study focuses on the semantic field of motion, as it is developed either as motion verbs or motion events, in the textbooks of Modern Greek Language and Modern Greek Literature for fifth and sixth grade of primary school. Our aim is, on the one hand, the creation of a corpus utilizing the authentic texts found in the school textbooks, and on the other hand, the analysis of the textual data regarding the semantic field of motion. The main tool of our research is the “motion” corpus, which is annotated with the goal of codifying the components of the motion event. The corpus was annotated both with respect to motion verbs (e.g., arrive, come, go) and with respect to the syntactic and semantic characteristics of the verbs that help in distinguishing meanings and grouping them (adverbial and prepositional expressions). The annotation scheme was designed according to Talmy’s typology, which posits that motion as an event is defined by six semantic characteristics: (a) the entity performing the motion, (b) the motion itself as an event, (c) the path, (d) the referential entity (landmark, characteristic element, location relative to the initial moving entity), (e) the manner, and (f) the cause. Through the annotation process and subsequent study of the corpus, we understand the semantic nuances, delving into the comprehension of different types of motion verbs with the perspective of creating a digital glossary for primary school. Additionally, this study is intended not only to shed light on the details of the semantic field of motion, but also to be utilized for the development of pedagogical tools that promote children’s linguistic understanding of the field in primary education.

Keywords: Corpora, motion verbs, primary school textbooks, authentic texts, modern Greek language

1. Εισαγωγή

Η μελέτη της κίνησης αποτελεί ένα σημαντικό σημασιολογικό πεδίο στη γλωσσολογία, καθώς συνδέεται άρρηκτα με τον τρόπο που οι γλώσσες περιγράφουν και αναπαριστούν τις κινήσεις στον χώρο και τον χρόνο. Η σημασιολογική ανάλυση των ρημάτων κίνησης προσφέρει μια βαθύτερη κατανόηση των γλωσσικών δομών και των εννοιολογικών σχέσεων που διέπουν τις φράσεις που εκφράζουν κίνηση.

Στην παρούσα μελέτη πρόκειται να μελετήσουμε το σημασιολογικό πεδίο της κίνησης, όπως αυτό αναπτύσσεται είτε ως ρήματα κίνησης είτε ως γεγονότα κίνησης, στα εγχειρίδια νεοελληνικής γλώσσας και νεοελληνικής λογοτεχνίας Ε΄ και ΣΤ΄ Δημοτικού, ακολουθώντας την τυπολογία του Talmy (1985), με τη χρήση των Σωμάτων Κειμένων και της Τεχνητής Νοημοσύνης.

2. Θεωρητικό πλαίσιο

2.1. Η Γλωσσολογία Σωμάτων Κειμένων

Η Γλωσσολογία Σωμάτων Κειμένων (Corpus Linguistics) αποτελεί έναν σύγχρονο κλάδο της γλωσσολογίας, ο οποίος επικεντρώνεται στη μελέτη της γλώσσας μέσα από μεγάλα σύνολα αυθεντικών κειμένων, που τα ονομάζουμε *σώματα κειμένων* (corpora). Η χρήση σωμάτων κειμένων επιτρέπει την εξαγωγή δεδομένων που αντανακλούν τη γλωσσική πραγματικότητα, με ιδιαίτερη έμφαση στην ανάλυση των γλωσσικών φαινομένων μέσω αυτοματοποιημένων μεθόδων επεξεργασίας (Sinclair, 1991, 2004). Όπως είναι αναμενόμενο, ο κλάδος αυτός έχει

επηρεαστεί σημαντικά από την εξέλιξη των νέων τεχνολογιών και τη χρήση ηλεκτρονικών υπολογιστών, οι οποίοι παρέχουν τη δυνατότητα αποθήκευσης και ανάλυσης μεγάλων ποσοτήτων γλωσσικού υλικού, καθιστώντας δυνατές τις αναλύσεις που θα ήταν αδύνατον να πραγματοποιηθούν με μη αυτοματοποιημένο τρόπο.

Σύμφωνα με τους Γούτσο & Φραγκάκη (2015), τα σώματα κειμένων είναι συλλογές αυθεντικών γλωσσικών αποσπασμάτων που επιλέγονται βάσει συγκεκριμένων γλωσσολογικών κριτηρίων, με στόχο να αποτελέσουν ένα αντιπροσωπευτικό δείγμα μιας γλώσσας ή μιας ποικιλίας της. Τα σώματα αυτά μπορεί να είναι είτε μονόγλωσσα είτε πολύγλωσσα, ενώ διακρίνονται σε συγχρονικά και διαχρονικά, ανάλογα με το αν περιλαμβάνουν στοιχεία μιας συγκεκριμένης χρονικής περιόδου ή αν καλύπτουν ένα μεγαλύτερο χρονικό εύρος. Επίσης, τα σώματα κειμένων ταξινομούνται με βάση διάφορα κριτήρια, όπως το εύρος γλωσσικής κάλυψης, η χρονική περίοδος και η μορφή της γλώσσας. Αφενός, υπάρχουν γενικά σώματα κειμένων που εκπροσωπούν τη γλώσσα σε ευρεία κλίμακα και εξυπηρετούν γενικούς σκοπούς, όπως είναι οι συχνές χρήσεις της γλώσσας. Αφετέρου, υπάρχουν και τα ειδικά σώματα κειμένων που εστιάζουν στη μελέτη συγκεκριμένων λειτουργιών ή χρήσεων της γλώσσας (Γούτσος & Φραγκάκη, 2015). Οπότε, ο σχεδιασμός, η κατασκευή και ανάλυση των σωμάτων κειμένων βασίζεται στην *αρχή της αντιπροσωπευτικότητας* (representativeness), η οποία διασφαλίζει ότι τα δεδομένα που συλλέγονται αντικατοπτρίζουν επαρκώς τη γλωσσική ποικιλία που μελετάται.

Επομένως, η χρήση των σωμάτων κειμένων επιτρέπει στους γλωσσολόγους να παρατηρούν τη συχνότητα χρήσης λεξικών τύπων, το συμφραστικό τους περιβάλλον και τις μεταβολές τους διαχρονικά (Sinclair, 1991). Ιδιαίτερα στη λεξικογραφία, η ανάλυση σωμάτων κειμένων προσφέρει πλούσια δεδομένα για τη συμπεριφορά των λέξεων, αναδεικνύοντας τις αυθεντικές χρήσεις τους. Αυτό είναι σημαντικό για την εξαγωγή αξιόπιστων συμπερασμάτων για τη γλώσσα και την αποφυγή της χρήσης κατασκευασμένων/μη αυθεντικών παραδειγμάτων που δεν λαμβάνουν υπόψη τα πραγματικά συμφραζόμενα (Γούτσος et al., 1995).

Με την εξέλιξη της τεχνολογίας, η χρήση ψηφιακών σωμάτων κειμένων έχει διευκολύνει τη γλωσσολογική έρευνα. Τα ειδικά προγράμματα ανάλυσης προσφέρουν τη δυνατότητα αυτόματης ταξινόμησης και μέτρησης γλωσσικών φαινομένων, όπως οι συμφραστικοί πίνακες και οι λίστες συχνότητων, τα οποία επιτρέπουν στους ερευνητές να εντοπίζουν γρήγορα τις πιο συχνές και αντιπροσωπευτικές χρήσεις των λέξεων και φράσεων (Μικρός, 2004).

Οπότε, η γλωσσολογία σωμάτων κειμένων προσφέρει πολύτιμα εργαλεία για τη μελέτη της γλώσσας, παρέχοντας τη δυνατότητα ανάλυσης σε διάφορα επίπεδα, όπως τη γραμματική, τη σημασιολογία και την κειμενογλωσσολογία. Τα σώματα κειμένων, με τη χρήση της τεχνολογίας, ανοίγουν νέους δρόμους για την εξαγωγή συμπερασμάτων που βασίζονται σε αυθεντικό γλωσσικό υλικό, συμβάλλοντας σε σημαντικό βαθμό ευρύτερα στη γλωσσική έρευνα.

2.2. Τυπολογία Talmy για το σημασιολογικό πεδίο της κίνησης

Ο L. Talmy –όπως και άλλοι γνωστικοί γλωσσολόγοι– υποστηρίζει ότι η γραμματική είναι άμεσα καθορισμένη από τη σημασία και αποτελεί σημασιολογικό σύστημα, πράγμα που επιβεβαιώνεται και από τη Levin (1993: 1): «Η συμπεριφορά του ρήματος, ιδιαίτερα σε

σχέση με την έκφραση και την ερμηνεία των ορισμάτων του, εξαρτάται σε μεγάλο βαθμό από τη σημασία του».

Συγκεκριμένα, για τα ρήματα κίνησης, ο Talmy έχει συνεισφέρει σημαντικά στη μελέτη της σημασίας των ρημάτων, κυρίως καταγράφοντας πώς αυτά κωδικοποιούνται σε διάφορες γλώσσες. Οπότε, διακρίνει τις γλώσσες σε δύο κατηγορίες βάσει του στοιχείου της κατεύθυνσης (Talmy, 1985, 2000), το οποίο μπορεί να εκφράζεται είτε από το κύριο ρήμα (verb-framed) είτε από ένα εξωρηματικό στοιχείο (satellite-framed).

Από τη μία πλευρά έχουμε τις *δορυφορικές γλώσσες* (satellite-framed languages), οι οποίες διαθέτουν τις πληροφορίες τους στα στοιχεία που συνοδεύουν το ρήμα, που ονομάζονται «δορυφόροι» (satellites). Σε αυτές τις γλώσσες, το ρήμα εκφράζει συνήθως τον τρόπο της κίνησης (manner), ενώ η κατεύθυνση (path) εκφράζεται από τον δορυφόρο, δηλαδή ένα στοιχείο εκτός του κύριου ρήματος (π.χ. μια πρόθεση ή πρόθεμα). Στα αγγλικά (1), για παράδειγμα, ο τρόπος εκφράζεται με το κύριο ρήμα και η κατεύθυνση με εξωρηματικά στοιχεία, καθιστώντας τα αγγλικά “satellite-framed” γλώσσα (Strömqvist et.al. 2004).

(1) *The cat ran into the house.* (= Η γάτα έτρεξε μέσα στο σπίτι.)

- Το ρήμα *ran* δηλώνει την κίνηση και τον τρόπο (δηλ. τρέξιμο).
- Η πρόθεση *into* δηλώνει την κατεύθυνση (δηλ. μέσα στο σπίτι).

Σε αυτό το παράδειγμα, το ρήμα (*ran*) περιγράφει το πώς (manner) της κίνησης, δηλαδή η γάτα κινήθηκε τρέχοντας, ενώ το στοιχείο *into* λειτουργεί ως δορυφόρος που εκφράζει την κατεύθυνση (path), δηλώνοντας ότι η γάτα μπήκε μέσα στο σπίτι.

Από την άλλη πλευρά έχουμε τις *ρηματικές γλώσσες* (verb-framed languages), στις οποίες η πορεία της κίνησης λεξικοποιείται ως σημασιολογικό συστατικό του ίδιου του ρήματος κίνησης. Με άλλα λόγια, το ρήμα εκφράζει τόσο την κίνηση όσο και την κατεύθυνση ταυτόχρονα, ενώ ο τρόπος της κίνησης (manner) μπορεί να δηλωθεί με διαφορετικούς τρόπους, όπως ένα επίρρημα ή μια φράση. Στα ισπανικά (2), και γενικά στις λατινογενείς γλώσσες, η κατεύθυνση κωδικοποιείται στο ρήμα και ο τρόπος περιφερειακά, καθιστώντας τα ισπανικά “verb-framed” γλώσσα (Strömqvist et.al. 2004):

(2) *El gato entró a la casa corriendo.* (= Η γάτα μπήκε στο σπίτι τρέχοντας.)

- Το ρήμα *entró* εκφράζει την κίνηση και την κατεύθυνση ταυτόχρονα, δηλώνοντας τη μετάβαση στο εσωτερικό του σπιτιού (δηλ. μπήκε).
- Η λέξη *corriendo* προστίθεται για να περιγράψει τον τρόπο (δηλ. τρέχοντας).

Σύμφωνα με τον Talmy, η ελληνική γλώσσα παρουσιάζει ένα υβριδικό σύστημα συγχώνευσης, στο οποίο τα γεγονότα κίνησης μπορούν να εκφραστούν και με τους δύο τρόπους (Talmy, 2000: 66). Αυτό σημαίνει ότι η ίδια έννοια της κίνησης μπορεί να αποδοθεί είτε δορυφορικά (3) είτε ρηματικά (4), επιτρέποντας εναλλακτικές επιλογές

ανάλογα με το πώς ενσωματώνεται ο τρόπος και η κατεύθυνση στην πρόταση. Στην ελληνική γλώσσα, μπορούμε να διακρίνουμε δύο χαρακτηριστικές δομές:

(3) *Έτρεξα μέσα στο σπίτι.*

○ Σε αυτή τη δομή, το ρήμα *έτρεξα* εκφράζει τον τρόπο της κίνησης (δηλ. τρέξιμο), ενώ η πρόθεση *μέσα* λειτουργεί ως δορυφόρος, προσδιορίζοντας την κατεύθυνση (δηλ. είσοδος στο σπίτι). Πρόκειται για έναν δορυφορικό τρόπο κωδικοποίησης της κίνησης.

(4) *Μπήκα τρέχοντας στο σπίτι.*

○ Εδώ το ρήμα *μπήκα* εκφράζει την κίνηση και την κατεύθυνση ταυτόχρονα (δηλ. είσοδος στο σπίτι), ενώ το *τρέχοντας* προστίθεται για να δηλώσει τον τρόπο (δηλ. τρέξιμο). Αυτή η δομή αντιστοιχεί στη ρηματική συγχώνευση, όπου η κατεύθυνση ενσωματώνεται στο ρήμα και ο τρόπος εκφράζεται με επίρρημα.

Η ύπαρξη αυτών των δύο γλωσσικών δομών καταδεικνύει την ευελιξία της ελληνικής γλώσσας, καθώς επιτρέπει την εναλλαγή μεταξύ διαφορετικών στρατηγικών για την έκφραση της κίνησης, γεγονός που την καθιστά ιδιαίτερα ενδιαφέρουσα από σημασιολογική και τυπολογική άποψη. Συνολικά, η ελληνική γλώσσα δεν έχει κατηγοριοποιηθεί με σαφήνεια σε καμία από τις δύο κύριες κατηγορίες από την πλειονότητα των ερευνητών, μιας και οι υπάρχουσες μελέτες δεν παρέχουν επαρκή στοιχεία για τη διατύπωση ασφαλών συμπερασμάτων. Παρά ταύτα, έχουν προταθεί διαφορετικές απόψεις. Για παράδειγμα, οι Parafraç et al. (2002) υποστήριξε ότι η ελληνική γλώσσα ανήκει στις λεγόμενες «ρηματικές γλώσσες» και το απέδειξε μέσω πειραμάτων που περιλάμβαναν περιγραφές εικόνων. Ένα από τα επιχειρήματά της ήταν πως οι Έλληνες ομιλητές χρησιμοποιούν σημαντικά μικρότερο αριθμό ρημάτων τρόπου σε σύγκριση με την αγγλική γλώσσα. Ωστόσο, το συγκεκριμένο ζήτημα δεν θα αποτελέσει αντικείμενο εξέτασης στην παρούσα έρευνα, ούτε θα υιοθετηθεί κάποια συγκεκριμένη θέση σχετικά με την κατηγοριοποίηση της ελληνικής γλώσσας.

Ωστόσο, αν και η θεωρία του Talmy παρουσιάζει ιδιαίτερο ενδιαφέρον, έχει αποτελέσει αντικείμενο κριτικής. Συγκεκριμένα, οι Hickmann et al. (2022) και ο Slobin (2004) προτείνουν μια πιο ρευστή προσέγγιση, τοποθετώντας τις γλώσσες σε ένα συνεχές (cline), όπου τα άκρα αντιπροσωπεύουν τις κατηγορίες του Talmy. Παρά τις αντιρρήσεις αυτές, το έργο του Talmy έχει ασκήσει βαθιά επίδραση στη μελέτη της σημασίας και στον τρόπο προσέγγισης των γλωσσικών φαινομένων, ιδίως στο σημασιολογικό πεδίο της κίνησης.

Η θεμελιώδης αρχή της τυπολογίας του Talmy είναι η έννοια της «συγχώνευσης γεγονότων» (event conflation). Ένα κατευθυνόμενο γεγονός κίνησης στη γλώσσα παρουσιάζεται συχνά σε συνδυασμό με ένα άλλο ταυτόχρονο γεγονός (Berman & Slobin, 1994). Πρόκειται για τη διαδικασία κατά την οποία ένα ρήμα κίνησης ενσωματώνει περισσότερα από ένα σημασιολογικά συστατικά του γεγονότος κίνησης σε μία μόνο λέξη ή φράση. Με βάση αυτή την ιδέα, ο Talmy προτείνει την ανάλυση των διαφορετικών σημασιολογικών στοιχείων που απαρτίζουν ένα γεγονός κίνησης, τα οποία αντιστοιχούν

σε επιφανειακά συντακτικά στοιχεία μιας πρότασης. Ξεκινά με τον προσδιορισμό των στοιχείων του βασικού σχήματος που υποκρύπτει ένα γεγονός κίνησης (Talmy, 1985: 61). Τα έξι σημασιολογικά στοιχεία που εντοπίζει είναι τα εξής: (1) Εικόνα, (2) Ορόσημο/Τοποθεσία, (3) Πορεία/Μονοπάτι, (4) Κίνηση, (5) Τρόπος και (6) Αιτία.

(1) Εικόνα/Φιγούρα: το κινούμενο αντικείμενο

(2) Ορόσημο/Τοποθεσία: οντότητες/χαρακτηριστικά στοιχεία που κινούνται σε σχέση με κάτι άλλο

(3) Πορεία/Μονοπάτι: η πορεία που ακολουθείται από τη φιγούρα

(4) Κίνηση: η παρούσα κίνηση που δίδεται, απεικονίζεται

(5) Τρόπος: ο τρόπος που η κίνηση διεκτυλίσσεται

(6) Αιτία: ποια είναι η αιτία της κινήσεως

Στον Πίνακα 1, δίνεται ένα χαρακτηριστικό παράδειγμα του Talmy (1985) που δείχνει πώς κωδικοποιούνται τα στοιχεία της κίνησης. Στην αγγλική γλώσσα, ο τρόπος της κίνησης (π.χ. *rolled* = κυλίστηκε/γλίστρησε) εκφράζεται μέσω του ρήματος, ενώ η πορεία (π.χ. *off* = μακριά από) δηλώνεται μέσω προθέσεων. Αντίστοιχα, στην ελληνική γλώσσα, ο τρόπος εκφράζεται με το ρήμα «γλίστρησε» και η πορεία με το επίρρημα «μακριά» ή τη φράση «μακριά από».

The pencil	rolled	off	the table.
Το μολύβι	γλίστρησε	μακριά	από το τραπέζι.
Εικόνα	Κίνηση Τρόπος	Πορεία	Ορόσημο/Τοποθεσία

Πίνακας 1. Χαρακτηριστικό παράδειγμα του Talmy

Με βάση αυτή τη θεωρητική υποδομή, θα εξετάσουμε πώς η ελληνική γλώσσα ενσωματώνει τα δομικά στοιχεία της κίνησης, εστιάζοντας στη χρήση των ρημάτων και στις στρατηγικές που υιοθετούν οι ομιλητές/τριες για να εκφράσουν αυτά τα στοιχεία. Μέσω αυτής της ανάλυσης, θα επιδιώξουμε να διαφωτίσουμε τις γλωσσικές τάσεις που χαρακτηρίζουν την ελληνική, δίνοντας έμφαση στη συστηματική μελέτη των ρημάτων κίνησης. Η κατηγοριοποίηση αυτή θα μας επιτρέψει όχι μόνο να κατανοήσουμε πιο βαθιά τη γραμματική και τη σημασιολογία των ρημάτων κίνησης στην ελληνική γλώσσα, αλλά και να συνεισφέρουμε στην ευρύτερη κατανόηση της γλωσσικής δομής.

2.3. Μεγάλα Γλωσσικά Μοντέλα

Η εμφάνιση των *Μεγάλων Γλωσσικών Μοντέλων* (Large Language Models – LLMs) έχει αλλάξει το τοπίο τόσο στην έρευνα και όσο και στη διδασκαλία. Τα γλωσσικά μοντέλα έχουν εξελιχθεί σημαντικά από τα προηγούμενα μοντέλα Μηχανικής Μάθησης, τα οποία

βασίζονταν στην τεχνική αναζήτησης δέσμης για την πρόβλεψη της επόμενης λέξης. Με την εισαγωγή των μετασχηματιστών (Vaswani et al., 2017), σημειώθηκε μεγάλη πρόοδος, ενώ ακολούθησαν ακόμα πιο προηγμένα και σύνθετα μοντέλα, όπως το *RoBERTa* (Liu et al., 2019), το *T5* (Raffel et al., 2020) και τα *GPT* μοντέλα, όπως το *GPT-2* (Radford et al., 2019).

Παρόλο που η επιστημονική κοινότητα είχε ήδη αρκετές γνώσεις για τέτοια μοντέλα, το ευρύτερο κοινό δεν ήταν ενημερωμένο και φυσικά ούτε εξοικειωμένο, πράγμα που άλλαξε τον Δεκέμβριο του 2022, όταν η OpenAI κυκλοφόρησε δημόσια και δωρεάν το *ChatGPT* (OpenAI, 2022), ένα Μεγάλο Μοντέλο Γλώσσας ειδικά εκπαιδευμένο για τη λειτουργία ως chatbot. Αξιοσημείωτο είναι ότι το *ChatGPT* έγινε μία από τις πλατφόρμες με την ταχύτερη υιοθέτηση όλων των εποχών, φτάνοντας το ένα εκατομμύριο χρήστες μέσα σε πέντε ημέρες. Για σύγκριση, το *Instagram* χρειάστηκε δυόμισι μήνες για να φτάσει το ένα εκατομμύριο λήψεις, ενώ το *Netflix* έφτασε το ένα εκατομμύριο χρήστες τρεισήμισι χρόνια μετά την κυκλοφορία του.

Τα LLMs είναι προχωρημένα νευρωνικά δίκτυα, τα οποία έχουν εκπαιδευτεί πάνω σε τεράστιες ποσότητες κειμένου. Η εκπαίδευση αυτή τους παρέχει τη δυνατότητα να κατανοούν και να παράγουν ανθρώπινη γλώσσα, προσομοιώνοντας την κατανόηση και τις γλωσσικές ικανότητες ενός ανθρώπινου νου. Επιπλέον, είναι ικανά να αποκωδικοποιούν την πολυπλοκότητα της φυσικής γλώσσας, αναγνωρίζοντας τα συμφραζόμενα και την έννοια πίσω από τις λέξεις και τις φράσεις.

Το μέγεθος τους μπορεί να ποικίλλει σημαντικά, από μικρά μοντέλα, όπως το *LLaMA 7B*, έως πολύ μεγάλα μοντέλα, όπως το *GPT-4*, το οποίο διαθέτει δισεκατομμύρια παραμέτρους. Το μέγεθος του μοντέλου έχει άμεση επίδραση στις δυνατότητές του και στην πολυπλοκότητα των προβλημάτων που μπορεί να επιλύσει. Τα μεγαλύτερα μοντέλα, παρόμοια με αυτά των δισεκατομμυρίων παραμέτρων, έχουν μεγαλύτερη ικανότητα γενίκευσης και παρέχουν πιο ακριβείς και συναφείς απαντήσεις. Συγκεκριμένα, τα LLMs έχουν τη δυνατότητα να χειρίζονται τον πλούτο και την πολυπλοκότητα της γλώσσας, αναγνωρίζοντας τη σύνταξη, τη σημασιολογία και το πρίσμα των συναισθημάτων που περιέχονται στα κείμενα, γεγονός που τους επιτρέπει να αποδώσουν με ακρίβεια τις λεπτομέρειες και τις αποχρώσεις της ανθρώπινης επικοινωνίας.

Επιπλέον, τα LLMs βρίσκουν εφαρμογή σε ένα ευρύ φάσμα τομέων. Όπως:

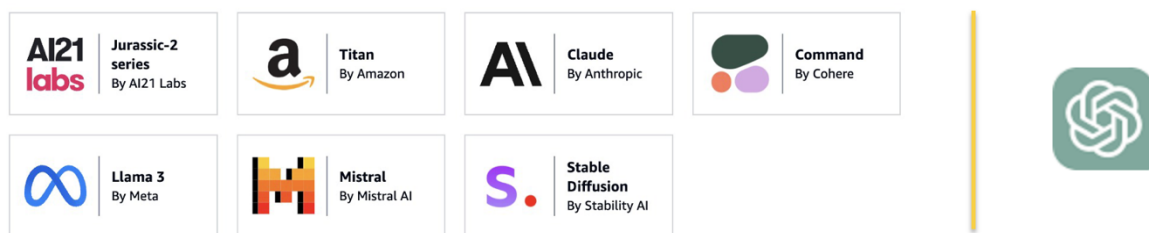
- (α) Στην αυτόματη μετάφραση, τα μοντέλα μπορούν να μεταφράζουν κείμενα από μία γλώσσα σε άλλη με υψηλό επίπεδο ακρίβειας, διευκολύνοντας την κατανόηση μεταξύ ανθρώπων διαφορετικών γλωσσών.
- (β) Στην Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing – NLP), χρησιμοποιούνται για αναγνώριση οντοτήτων, ανάλυση συναισθημάτων και εξαγωγή πληροφορίας από κείμενα, καθιστώντας δυνατή την ανάλυση μεγάλων όγκων δεδομένων με ακρίβεια και ταχύτητα.
- (γ) Στα αυτόματα συστήματα ομιλίας, τα LLMs διευκολύνουν τη φυσική επικοινωνία μεταξύ ανθρώπων και μηχανών, βελτιώνοντας την αλληλεπίδραση χρήστη-υπολογιστή και καθιστώντας τις συσκευές πιο φιλικές και προσαρμοσμένες στις ανάγκες των χρηστών.

Συνολικά, μπορούμε να συμπεράνουμε ότι τα μεγάλα γλωσσικά μοντέλα συνεισφέρουν σημαντικά στην τεχνολογική πρόοδο και στην ενίσχυση της αποδοτικότητας σε διάφορους

τομείς, συμπεριλαμβανομένης της γλωσσικής ανάλυσης, που αποτελεί και το αντικείμενο της παρούσας μελέτης.

Στο σημείο αυτό πρέπει να αναφερθούμε σε ορισμένα διάσημα μοντέλα τεχνητής νοημοσύνης που κυριαρχούν στον τομέα της επεξεργασίας φυσικής γλώσσας και των εφαρμογών της, παρουσιάζοντας διαφορετικές προσεγγίσεις και χαρακτηριστικά, τα οποία συμβάλλουν στην πρόοδο της έρευνας και της τεχνολογίας. Ενδεικτικά:

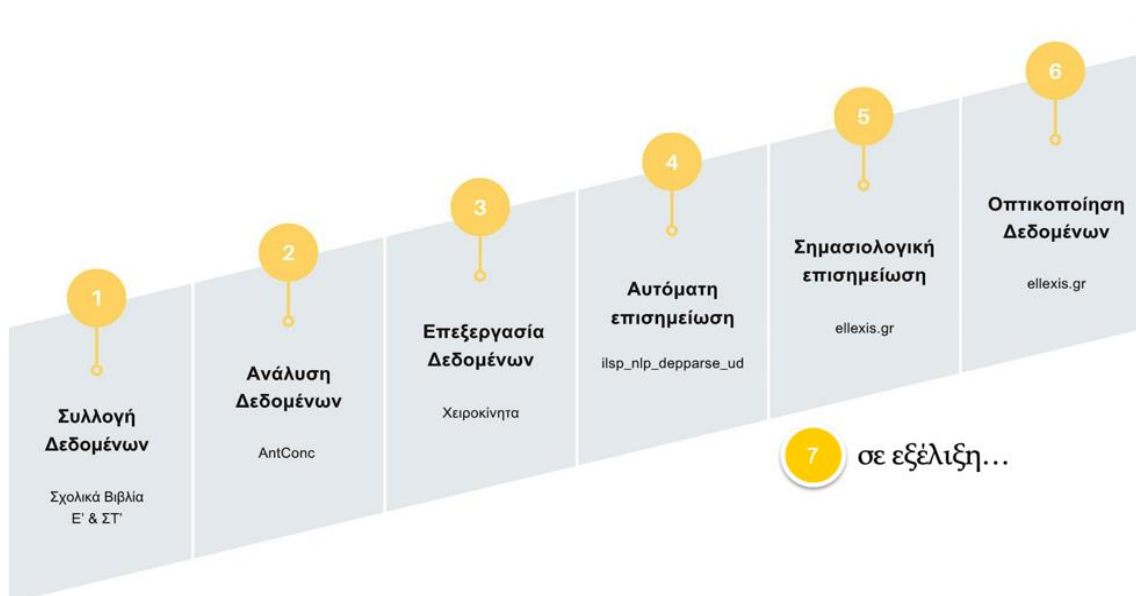
- Το *GPT-3.5* και η μεταγενέστερη έκδοση *GPT-4*, που προέρχονται από την OpenAI, αποτελούν εξελιγμένα γλωσσικά μοντέλα με δισεκατομμύρια παραμέτρους. Αυτά τα μοντέλα είναι ευρέως γνωστά για την ικανότητά τους να προσαρμόζονται σε ποικίλες γλωσσικές εργασίες με λίγη ή καθόλου επιμέρους εκπαίδευση.
- Το *Claude*, μοντέλο της εταιρείας Anthropic, εστιάζει στην ασφαλή και ανθρωποκεντρική αλληλεπίδραση. Η ανάπτυξή του δίνει ιδιαίτερη έμφαση στην ασφάλεια, την ηθική και την υπευθυνότητα στη χρήση της τεχνητής νοημοσύνης.
- Το *LLaMA* (Touvron et al., 2023), που αναπτύχθηκε από τη Meta, προορίζεται κυρίως για ερευνητικούς σκοπούς. Το μοντέλο αυτό διατίθεται σε διάφορα μεγέθη, επιτρέποντας την προσαρμογή του σε διαφορετικά ερευνητικά και πρακτικά περιβάλλοντα.



Εικόνα 1. Διάσημα μοντέλα τεχνητής νοημοσύνης

3. Δημιουργία Σώματος Κειμένων και Σημασιολογική Επισημείωση

Στην παρούσα ερευνητική εργασία, η μελέτη του σημασιολογικού πεδίου της κίνησης βασίζεται στη μελέτη του Σώματος Κειμένων. Η διαδικασία ανάλυσης δεδομένων περιγράφεται λεπτομερώς και περιλαμβάνει τη χρήση προηγμένων εργαλείων λογισμικού και μεθόδων επεξεργασίας. Η πολυπλοκότητα και η ακρίβεια των βημάτων που ακολουθήθηκαν, επιτρέπουν την εξαγωγή συμπερασμάτων από τα δεδομένα, συμβάλλοντας έτσι στην αποτύπωση και κατανόηση των πληροφοριών που αυτά εμπεριέχουν. Η ανάλυση πραγματοποιήθηκε σε έξι διαδοχικά βήματα, με την εφαρμογή διαφόρων εργαλείων και μεθόδων επεξεργασίας δεδομένων, ξεκινώντας από τη συλλογή των δεδομένων μέχρι την τελική οπτικοποίηση και ερμηνεία των αποτελεσμάτων. Παρακάτω, παρουσιάζονται αναλυτικά τα έξι βασικά στάδια της διαδικασίας που ακολουθήθηκε: (1) Συλλογή Δεδομένων, (2) Ανάλυση Δεδομένων, (3) Επεξεργασία Δεδομένων, (4) Αυτόματη Επισημείωση, (5) Σημασιολογική Επισημάνση, και (6) Οπτικοποίηση Δεδομένων. (Η μελέτη συνεχίζεται με ένα έβδομο στάδιο που βρίσκεται σε εξέλιξη, όπου αναμένονται επιπλέον βελτιώσεις και ευρήματα.



Εικόνα 2. Στάδια συλλογής και επεξεργασίας δεδομένων

3.1. Συλλογή Δεδομένων

Το Σώμα Κειμένων (ΣΚ) για την παρούσα μελέτη βασίστηκε σε κείμενα από τα σχολικά βιβλία της Νεοελληνικής Γλώσσας και της Νεοελληνικής Λογοτεχνίας που χρησιμοποιούνται στην Ε' και ΣΤ' τάξη του Δημοτικού. Δημιουργήθηκε, λοιπόν, ένα ειδικό και μονόγλωσσο ΣΚ, μεγέθους 89.580 λέξεων. Τα εγχειρίδια αυτά αποτέλεσαν την πρωτογενή πηγή δεδομένων, παρέχοντας αυθεντικά παραδείγματα γραπτού λόγου στα ελληνικά σε εκπαιδευτικό πλαίσιο. Μέσω της ανάλυσης αυτών των κειμένων, επιχειρούμε να διερευνήσουμε τη χρήση των ρημάτων κίνησης, εξετάζοντας τις γλωσσικές στρατηγικές που προτείνονται για το σημασιολογικό πεδίο της κίνησης. Η επιλογή των συγκεκριμένων βιβλίων κρίθηκε κατάλληλη λόγω του ότι περιλαμβάνουν διαφορετικούς κειμενικούς τύπους και ποικίλα είδη λόγου, καθιστώντας το δείγμα αντιπροσωπευτικό της γλωσσικής έκφρασης που διδάσκεται στους/τις μαθητές/τριες.

3.2. Ανάλυση Δεδομένων

Η αρχική ανάλυση των συλλεχθέντων δεδομένων πραγματοποιήθηκε με τη χρήση του λογισμικού *AntConc* (<https://www.laurenceanthony.net/software.html>). Το *AntConc* παρέχει στους ερευνητές ένα ευρύ φάσμα εργαλείων για τη λεπτομερή ανάλυση κειμένων, επιτρέποντας τη συγκεντρωτική και συγκριτική εξέταση των γλωσσικών δεδομένων. Με τις δυνατότητές του, όπως η δημιουργία λιστών συχνότητας λέξεων, η ανίχνευση λέξεων-κλειδιών και η ανάλυση των συμφραζομένων, καθίσταται δυνατή η ενδελεχής διερεύνηση των γλωσσικών μοτίβων και των λεξιλογικών δομών που εμφανίζονται σε ένα ΣΚ. Στην παρούσα έρευνα, αξιοποιήσαμε το *AntConc* για την εξαγωγή και ανάλυση των ρημάτων από το συγκεκριμένο σώμα κειμένων. Μέσω της χρήσης αυτού του λογισμικού, εντοπίσαμε και καταγράψαμε 5.268 μοναδικούς ρηματικούς τύπους, μαζί με τις συγκεκριμένες προτάσεις στις οποίες εντάσσονται. Η διαδικασία αυτή μας επέτρεψε να χαρτογραφήσουμε με ακρίβεια τη χρήση των ρημάτων στο ΣΚ με στόχο στο επόμενο στάδιο επεξεργασίας να απομονώσουμε τις προτάσεις που εκφράζουν κίνηση με σκοπό να τις αναλύσουμε.

3.3. Επεξεργασία Δεδομένων

Σε αυτό το στάδιο, πραγματοποιήθηκαν διορθώσεις, καθαρισμός και μορφοποίηση των δεδομένων, ώστε να είναι κατάλληλα για περαιτέρω επεξεργασία και ανάλυση. Έτσι, δημιουργήσαμε το τελικό ΣΚ που περιλαμβάνει όλες τις προτάσεις που εκφράζουν κίνηση, επιτρέποντας τη λεπτομερή ανάλυση των ρημάτων κίνησης. Το μέγεθός του είναι 12.678 λέξεις. Το σύνολο των ρημάτων κίνησης που εντοπίστηκαν στο ΣΚ είναι 36 λήμματα, εκ των οποίων ενδεικτικά παρατίθενται τα εξής: *έρχομαι, περπατάω, τρέχω, πηδάω, πετάω, κολυμπάω, σέρνομαι, κουτσαίνω, πέφτω, σκαρφαλώνω, ταξιδεύω, γλιστρώ, βουτάω, φεύγω, μπαίνω*.

Η εφαρμογή αυτής της μεθοδολογίας μάς επιτρέπει να εξετάσουμε σε βάθος τις σημασιολογικές και γραμματικές λειτουργίες των ρημάτων κίνησης, καθώς και τις στρατηγικές που χρησιμοποιούνται για την έκφραση της κίνησης στην ελληνική γλώσσα, προσφέροντας σημαντικές γλωσσολογικές πληροφορίες για το υπό μελέτη φαινόμενο. Η επεξεργασία των δεδομένων πραγματοποιήθηκε χειροκίνητα, εξασφαλίζοντας την ακρίβεια στον εντοπισμό και την ανάλυση των ρημάτων κίνησης.

Ωστόσο, κατά τη διαδικασία συλλογής δεδομένων, προέκυψαν ορισμένοι περιορισμοί. Ένα από τα προβλήματα ήταν η πολλαπλή ερμηνεία ορισμένων ρημάτων, τα οποία, αν και περιμέναμε να εκφράζουν κίνηση (5), χρησιμοποιούνται μεταφορικά ή σε μη κυριολεκτικές εκφράσεις (6).

(5) *Ο Σπόρος έρχεται να μας ενημερώσει.*

○ Ρήμα κίνησης (*έρχομαι*): κυριολεκτική χρήση, εκφράζει την κίνηση προς ένα σημείο/πρόσωπο.

(6) *Έτσι μού έρχεται να κλαίω.*

○ Ρήμα κίνησης (*έρχομαι*): μεταφορική χρήση, εκφράζει την αίσθηση της ανάγκης ή της επιθυμίας να κάνει κάποιος κάτι, χωρίς φυσική κίνηση.

Επομένως, ορισμένες προτάσεις που εκφράζουν μεταφορικά ή μη κυριολεκτικά γεγονότα απομακρύνθηκαν από το ΣΚ, ώστε να εξασφαλιστεί ότι το υλικό που θα εξεταστεί είναι απαλλαγμένο από μεταφορικές/μη κυριολεκτικές χρήσεις, διασφαλίζοντας έτσι την ακρίβεια της ανάλυσης των ρημάτων κίνησης. Στον Πίνακα 2, παρουσιάζονται ενδεικτικά ορισμένες από τις προτάσεις που αναλύθηκαν.

	Δεδομένα	Ρήμα Κίνησης
1	<i>Μάθαμε ότι έρχεται ο Ιμπραΐμ με τους Άραβες να βοηθήσει τον Κιουταχή.</i>	έρχομαι
2	<i>Ο Ένας και ο Άλλος χαμογέλασαν επιτέλους και αγκαλιασμένοι πήγαν παράμερα να παίξουν μια παρτίδα.</i>	πηγαίνω

3	Το παιδί έπεσε πάνου στο βιβλίο με τα μούτρα.	πέφτω
4	Μπαίνει ο ιστορικός Α' κρατώντας ένα μεγάλο τετράδιο.	μπαίνω
5	Σκαρφάλωσε στη βιβλιοθήκη του ως το πάνω πάνω ράφι κατέβασε μια εγκυκλοπαίδεια.	σκαρφαλώνω

Πίνακας 2. Ενδεικτικά παραδείγματα προτάσεων

3.4. Αυτόματη Επισημείωση

Για την αυτόματη επισήμανση των δεδομένων, χρησιμοποιήθηκε το εργαλείο *ilsp_nlp_depparse_ud*, προκειμένου να γίνει γραμματική ανάλυση και συντακτική επεξεργασία των δεδομένων μου, καθώς είναι το μόνο δωρεάν εργαλείο διαθέσιμο online για την ελληνική γλώσσα. Στο πρώτο στάδιο, το *ilsp_nlp_depparse_ud* ανιχνεύει τα όρια των λέξεων, προτάσεων και παραγράφων, ενώ παράγει ετικέτες μέρους του λόγου και λεξιλογικές αναφορές για κάθε λέξη. Το σχήμα ανάλυσης που χρησιμοποιείται βασίζεται κυρίως στο ILSP_PAROLE_Tagset (http://nlp.ilsp.gr/nlp/tagset_examples/tagset_en/). Το ILSP_PAROLE_Tagset, το οποίο είναι μια προσαρμογή του PAROLE Tagset, είναι ένα αυθεντικό και ολοκληρωμένο σχήμα ανάλυσης όσον αφορά τα λεξιλογικά και συντακτικά χαρακτηριστικά της ελληνικής γλώσσας (Labropoulou et al., 1996). Τα επισημειωμένα δεδομένα από το *ilsp_nlp_depparse_ud* επιστρέφονται σε ένα έγγραφο της μορφής CoNLL 2007 (βλ. λεπτομερώς Buchholz & Marsi, 2006, καθώς και Nivre et al. 2007).

Στην Εικόνα 3, παρουσιάζεται η έξοδος *ilsp_nlp_depparse_ud* της πρότασης «Το παιδί βγήκε στον κήπο τρέχοντας μέσα από το γεφυράκι» σε μορφή CoNLL. Η πρώτη στήλη αντιστοιχεί με το ID (A/A): Πρόκειται για τον αύξοντα αριθμό που δείχνει τη θέση της λέξης μέσα στην πρόταση. Η δεύτερη στήλη αντιστοιχεί με το Form (Μορφή): Αυτή η στήλη περιέχει τη λέξη όπως εμφανίζεται στην πρόταση. Η τρίτη στήλη αντιστοιχεί με το Lemma (Λήμμα): Το λήμμα είναι η βασική μορφή της λέξης, όπως θα εμφανιζόταν σε ένα λεξικό. Η τέταρτη στήλη αντιστοιχεί με το CPOSTAG (Γενική κατηγορία λέξης): Αυτή η στήλη δείχνει το μέρος του λόγου (όπως ουσιαστικό, ρήμα). Η πέμπτη στήλη αντιστοιχεί με το POSTAG (Λεπτομερής κατηγορία λέξης): Εξειδικευμένη κατηγορία λέξης που παρέχει περισσότερες πληροφορίες για την κατηγορία. Η έκτη στήλη αντιστοιχεί με το FEATS (Χαρακτηριστικά λέξης): Περιέχει μορφολογικά χαρακτηριστικά, όπως γένος, αριθμός, χρόνος. Η έβδομη στήλη αντιστοιχεί με το HEAD (Αναφορά στον επικεφαλής): Δείχνει τον αριθμό ID της λέξης-κεφαλής με την οποία η λέξη συνδέεται συντακτικά. Η όγδοη στήλη αντιστοιχεί με το DEPREL (Σχέση εξάρτησης): Περιγράφει τη συντακτική σχέση εξάρτησης ανάμεσα στη λέξη και την κεφαλή της, όπως υποκείμενο, αντικείμενο κ.λπ.

```
# sent_id = o_outputout1
# text = Το παιδί βγήκε στον κήπο τρέχοντας μέσα από το γεφυράκι.
1 Το ο DET AtDf Case=Nom|Definite=Def|Gender=Neut|Number=Sing|PronType=Art 2 det
2 παιδί παιδί NOUN NoCm Case=Nom|Gender=Neut|Number=Sing 3 nsubj
3 βγήκε βγαίνω VERB VbMn Aspect=Perf|Mood=Ind|Number=Sing|Person=3|Tense=Past|Voice=Act 0 root
4 στον σπου ADP AsPpPa Case=Acc|Gender=Masc|Number=Sing 5 case
5 κήπο κήπος NOUN NoCm Case=Acc|Gender=Masc|Number=Sing 3 obl
6 τρέχοντας τρέχω VERB VbMn Aspect=Imp|VerbForm=Part|Voice=Act 3 advcl
7 μέσα μέσα ADV Ad _ 6 advmod
8 από από ADP AsPpSp _ 10 case
9 το ο DET AtDf Case=Acc|Definite=Def|Gender=Neut|Number=Sing|PronType=Art 10 det
10 γεφυράκι γεφυράκι NOUN NoCm Case=Acc|Gender=Neut|Number=Sing 7 obl
11 . . PUNCT PTERM_P PunctType=PTERM_P 3 punct
```

Εικόνα 3. Παράδειγμα πρότασης σε μορφή CoNLL

3.5. Σημασιολογική Επισήμανση

Για τη δημιουργία ενός επισημειωμένου ΣΚ που θα αποδίδει με ακρίβεια τις σημασιολογικές πληροφορίες σχετικά με το πεδίο της κίνησης, ήταν απαραίτητο να σχεδιαστεί ένα κατάλληλο σχήμα επισημείωσης. Η ανάγκη για ένα τέτοιο σχήμα προέκυψε από την απαίτηση να κατηγοριοποιηθούν και να αναλυθούν συστηματικά οι διαφορετικές πτυχές της κίνησης, όπως εκφράζονται στη γλώσσα. Βασισμένοι στο θεωρητικό πλαίσιο του Talmy για τη γλωσσική έκφραση της κίνησης, αναπτύξαμε ένα μοντέλο επισημείωσης που αντανακλά τις βασικές συνιστώσες του φαινομένου της κίνησης, παρέχοντας μια πλήρη χαρτογράφηση των στοιχείων που εμπλέκονται σε αυτή τη διαδικασία. Οι ετικέτες επισημείωσης περιλαμβάνουν τις εξής κατηγορίες:

- **Figure** (Εικόνα): Η φιγούρα ή το αντικείμενο που βρίσκεται σε κίνηση.
- **Motion** (Κίνηση): Το ρήμα που περιγράφει την κίνηση.
- **GoalLoc** (Τοποθεσία Προορισμού): Η τοποθεσία ή ο προορισμός της κίνησης.
- **Manner** (Τρόπος): Ο τρόπος με τον οποίο πραγματοποιείται η κίνηση.
- **Path** (Μονοπάτι): Η διαδρομή ή η πορεία που ακολουθείται κατά την κίνηση.

Εφαρμόζοντας το προαναφερθέν σχήμα επισημείωσης στην πρόταση «Το παιδί βγήκε στον κήπο τρέχοντας μέσα από το γεφυράκι», προκύπτει η εξής ανάλυση:

- **Το παιδί** → *SemMotion=Figure*: Το ουσιαστικό «παιδί» αναφέρεται στη φιγούρα που εκτελεί την κίνηση.
- **βγήκε** → *SemMotion=Motion*: Το ρήμα «βγήκε» περιγράφει την κίνηση του παιδιού από έναν εσωτερικό σε έναν εξωτερικό χώρο.
- **στον κήπο** → *SemMotion=GoalLoc*: Η φράση «στον κήπο» εκφράζει την τοποθεσία προορισμού, δηλαδή τον τελικό προορισμό της κίνησης.
- **τρέχοντας** → *SemMotion=Manner*: Ο ρηματικός τύπος «τρέχοντας» εκφράζει τον τρόπο με τον οποίο πραγματοποιήθηκε η κίνηση, δηλαδή το γεγονός ότι το παιδί κινήθηκε με γρήγορο βήμα.
- **μέσα από το γεφυράκι** → *SemMotion=Path*: Η φράση «μέσα από το γεφυράκι» περιγράφει την πορεία ή το μονοπάτι που ακολούθησε το παιδί για να φτάσει στον κήπο.

Η σημασιολογική επισήμανση των δεδομένων πραγματοποιήθηκε μέσω της πλατφόρμας *ellexis* (ellexis.gr), η οποία επιτρέπει τη σημασιολογική ανάλυση και επισήμανση των δεδομένων, εξάγοντας παράλληλα πολύτιμες πληροφορίες. Η πλατφόρμα αυτή θα παρουσιαστεί αναλυτικά στο επόμενο στάδιο της έρευνας, καθώς συνδέεται άμεσα με τη διαδικασία οπτικοποίησης των αποτελεσμάτων.

3.6. Οπτικοποίηση Δεδομένων

Σε αυτό το στάδιο της μελέτης, προχωρούμε στην οπτικοποίηση των δεδομένων με τη χρήση *ellexis* (Fiotaki & Tsakalis, 2025 *forthcoming*). Ειδικότερα, θα παρουσιάσουμε την οπτικοποίηση των σημασιολογικών ετικετών, εφαρμόζοντάς τες στο παράδειγμα της πρότασης που αναλύθηκε και επισημειώθηκε στο προηγούμενο στάδιο. Στην Εικόνα 4, παρουσιάζεται το παράδειγμα με τις σημασιολογικές ετικέτες που αφορούν την κίνηση.

Το παιδί βγήκε στον κήπο τρέχοντας μέσα από το γεφυράκι.						
1	1	Το	ο	DET	AtDf	Case=Nom Definite=Def Gender=Neut Number=Sing PronType=Art
2	2	παιδί	παιδί	NOUN	NoCm	Case=Nom Gender=Neut Number=Sing SemMotion=Figure
3	3	βγήκε	βγαίνω	VERB	VbMn	Aspect=Perf Mood=Ind Number=Sing Person=3 Tense=Past Voice=Act SemMotion
4	4	στον	στου	ADP	AsPpPa	Case=Acc Gender=Masc Number=Sing
5	5	κήπο	κήπος	NOUN	NoCm	Case=Acc Gender=Masc Number=Sing SemMotion=GoalLoc
6	6	τρέχοντας	τρέχω	VERB	VbMn	Aspect=Imp VerbForm=Part Voice=Act SemMotion=Manner
7	7	μέσα	μέσα	ADV	Ad	
8	8	από	από	ADP	AsPpSp	
9	9	το	ο	DET	AtDf	Case=Acc Definite=Def Gender=Neut Number=Sing PronType=Art
10	10	γεφυράκι	γεφυράκι	NOUN	NoCm	Case=Acc Gender=Neut Number=Sing SemMotion=Path
11	11	.	.	PUNCT	PTermP	PunctType=PTermP

Εικόνα 4. Σημασιολογική και συντακτική επισημείωση της πρότασης μελέτης στο περιβάλλον *ellexis*

Το *ellexis* δεν οπτικοποιεί απλώς τα δεδομένα, αλλά δίνει τη δυνατότητα στον χρήστη να κάνει αναζήτηση με βάση συγκεκριμένα κριτήρια, όπως λέξεις, μέρη του λόγου κ.ά. Επιπλέον, δίνει τη δυνατότητα για περιγραφική στατιστική ανάλυση των δεδομένων, όπως η σχετική συχνότητα εμφάνισης της λέξης, με ποιες λέξεις συντάσσεται ο όρος αναζήτησης κ.ά. Παρακάτω παρατίθενται τα περιγραφικά αποτελέσματα του ΣΚ που δημιουργήσαμε για την παρούσα μελέτη.

Number of sentences	Number of Unique Lemmas	Number of Unique Tokens
N	N	N
1057	7903	16903

Πίνακας 3. Περιγραφικά αποτελέσματα

Ο Πίνακας 3 περιλαμβάνει τις μετρήσεις του υποσώματος που μελετάμε, όπως αυτά παράχθηκαν αυτόματα από το *ellexis*.

- **Αριθμός προτάσεων (Number of sentences):** Ο συνολικός αριθμός των προτάσεων είναι **1.057**.
- **Αριθμός μοναδικών λημμάτων (Number of Unique Lemmas):** Υπάρχουν **7.903** μοναδικά λήμματα.
- **Αριθμός μοναδικών λέξεων (Number of Unique Tokens):** Ο αριθμός των μοναδικών λέξεων ανέρχεται σε **16.903**.

Παράλληλα με τις ενδεικτικές λίστες που περιλαμβάνουν τις πιο συχνές λέξεις και λήμματα (τις οποίες δεν παραθέτουμε εδώ, βλ. Πίνακα 4), η ανάλυση του ΣΚ παρέχει επίσης πληροφορίες σχετικά με τις διμερείς και τριμερείς συνδυαστικές φράσεις/συνάψεις λέξεων (collocations).

2-Word Collocation			3-Word Collocation		
two-word Collocations		count	three-word Collocations		count
1	δε φτάνει	7	1	σαν είναι να	2
2	είχαν έρθει	5	2	έρχεται να μας	2
3	να πάω	18	3	κι άλλο κι	2
4	στη θάλασσα	6	4	είπε κι άνοιξε	2
5	δρόμο όπου	3	5	που ήρθαν στη	2
6	φορές ημέρα	3	6	να πηγαίνει να	3
7	ακόμα πιο	3	7	περνάει να σου	2
8	γυρίζει πίσω	3	8	σα να ήρθε	2
9	όπου πήγαινε	3	9	να σου φτάνει	2
10	σαν ήρθε	5	10	έτσι ρχεται να	2

Πίνακας 4. Διμερείς και τριμερείς συνάψεις λέξεων

Η προσέγγιση που υιοθετήσαμε διασφαλίζει την ακρίβεια και την επαναληψιμότητα της έρευνας, δίνοντας παράλληλα μια ολοκληρωμένη εικόνα των δεδομένων και των πληροφοριών που περιέχουν. Ένα τέτοιο επισημειωμένο σώμα κειμένων συνιστά μια

πολύτιμη πηγή δεδομένων για τη μελέτη των γλωσσικών δομών που σχετίζονται με το σημασιολογικό πεδίο της κίνησης.

4. Σημασιολογικό πεδίο της Κίνησης στα Μεγάλα Γλωσσικά Μοντέλα

Ως επιπρόσθετη μελέτη περίπτωσης στην παρούσα εργασία, αποπειραθήκαμε να εξετάσουμε πώς τα μεγάλα γλωσσικά μοντέλα μπορούν να συμβάλουν στην ανάλυση των γλωσσικών δομών του σημασιολογικού πεδίου της κίνησης. Για να προχωρήσουμε σε αυτή τη μελέτη, επιλέξαμε να τρέξουμε το παρακάτω prompt (το οποίο συνήθως μεταφράζεται ως προτροπή) στα αγγλικά, αντί στα ελληνικά, προκειμένου να αξιοποιήσουμε τις καλύτερες δυνατότητες των γλωσσικών μοντέλων. Η επιλογή αυτή επιτρέπει την αξιολόγηση και τη σύγκριση των αποτελεσμάτων από διαφορετικά γλωσσικά μοντέλα, όπως το *GPT* και το *Claude*.

- **Prompt:** *Consider the Talmy's typology and give the necessary semantic annotation for the following clause in Greek: Το παιδί βγήκε στον κήπο τρέχοντας μέσα από το γεφυράκι.*

ChatGPT3.5-turbo

In this sentence, we have motion through a path contained within a surrounding object. Therefore, the semantic annotation for this clause in Greek would be:

Το παιδί (theme) βγήκε (V) στον κήπο (goal) τρέχοντας (manner) μέσα από το γεφυράκι (path)

ChatGPT4

In Talmy's typology, the Greek language is considered **as satellite-framed**. The core schema of the event expressed in the clause is: 'The child went out in the garden running through the bridge'.

The semantic annotation would look like this:

1. **[Figure: Το παιδί]** – 'The child', is the moving entity or theme of the motion event.
2. **[Motion verb: βγήκε]** – 'went out', describes the act or state of motion of the figure.
3. **[Path: στον κήπο]** – 'in the garden', indicates the spatial course or trajectory that the figure follows.
4. **[Manner: τρέχοντας]** – 'running', describes the way or mode in which the motion is carried out.
5. **[Path: μέσα από το γεφυράκι]** – 'through the bridge', indicates the specific trajectory the motion event follows.

In Greek, the path of motion is usually expressed outside the verb root, typically

	using prepositional phrases or adverbs, which are known as ‘satellites’ in Talmy’s typology. This is evident in this example where ‘στον κήπο’ (in the garden) and ‘μέσα από το γεφυράκι’ (through the bridge) express the path of the moving entity ‘το παιδί’ (the child). The motion verb ‘βγήκε’ (went out) depicts the motion itself while ‘τρέχοντας’ (running) describes the manner of motion.
--	---

ChatGPT4-turbo

Talmy’s typology focuses on how different languages encode the relationship between motion, path, and manner in verb phrases and sentences. In languages like Greek, which can be categorized as **verb-framed** or **satellite-framed** depending on the elements involved, the detailed semantics of motion events are often expressed through verb prefixes/suffixes (satellites) that indicate path or particles and other verb complements that express manner or cause.

Here’s the breakdown for the Greek clause: Το παιδί βγήκε στον κήπο τρέχοντας μέσα από το γεφυράκι.

- Το παιδί = The child (subject)
- βγήκε = went out, emerged (main verb, indicates the motion)
- στον κήπο = to the garden (path, destination of the motion)
- τρέχοντας = running (manner of the motion; gerund/participle form expressing manner)
- μέσα από το γεφυράκι = through the bridge (path detail, further specifies the trajectory of the motion through a constrained path)

Let’s annotate it semantically according to Talmy’s typology:

- Figure: Το παιδί (the child)
- Motion: βγήκε (went out, emerged)
- Path: στον κήπο (to the garden), μέσα από το γεφυράκι (through the bridge)
- Manner: τρέχοντας (running)

Complete Semantic Annotation:

- **Figure (F):** “Το παιδί” - **Motion Verb (M):** “βγήκε” - **Path (P):** “στον κήπο”, “μέσα από το γεφυράκι” - **Manner (Mnr):** “τρέχοντας”

Συγκρίνοντας τις τρεις απαντήσεις από τις εκδόσεις του ChatGPT, παρατηρούμε σημαντικές διαφορές στην ανάλυση και την πληρότητα της παρουσίασης. Αρχικά, στην

απάντηση του *ChatGPT 3.5-turbo*, η ανάλυση είναι σαφής και περιεκτική, αλλά περιορίζεται σε μια σύντομη απάντηση χωρίς να εξετάσει το πλαίσιο της τυπολογίας του Talmy. Αυτό μειώνει την κατανόηση της δομής της ελληνικής γλώσσας και των στοιχείων που συμμετέχουν. Έπειτα, στην απάντηση του *ChatGPT 4*, η ανάλυση είναι πιο αναλυτική και οργανωμένη, προσφέροντας μια λεπτομερή περιγραφή των στοιχείων της πρότασης. Επίσης, εξετάζει τη διάκριση μεταξύ “satellite-framed” και “verb-framed” γλωσσών, κάτι που ενισχύει τη δυνατότητα κατανόησης της τυπολογίας.

Τέλος, στην απάντηση του *ChatGPT 4-turbo*, παρέχεται μια ολοκληρωμένη ανάλυση που αναδεικνύει τη σημασία της τυπολογίας του Talmy στην κατανόηση της κίνησης και της διαδρομής στην ελληνική γλώσσα. Η διάρθρωση της ανάλυσης είναι εξαιρετικά σαφής, με κατηγορίες όπως Figure, Motion, Path και Manner να παρατίθενται λεπτομερώς. Επίσης, υπογραμμίζει τη χρήση των προθέσεων σε σχέση με το κυρίως ρήμα, κάτι που προσθέτει βάθος στην ερμηνεία της πρότασης. Ωστόσο, πρέπει να τονίσουμε ότι οι απαντήσεις τόσο του *ChatGPT 4* όσο και του *ChatGPT 4-turbo* δίνουν εσφαλμένα ότι “στον κήπο” είναι Μονοπάτι (Path), ενώ θα έπρεπε να χαρακτηρίζεται ως Τοποθεσία Προορισμού. Συγκριτικά, η απάντηση του *ChatGPT 4-turbo* είναι η πιο ολοκληρωμένη, συνδυάζοντας την ανάλυση των στοιχείων με τις θεωρητικές πτυχές της τυπολογίας. Ακολουθεί η απάντηση του *ChatGPT 4*, η οποία είναι επίσης αρκετά καλή αλλά λιγότερο λεπτομερής. Τέλος, η απάντηση του *ChatGPT 3.5-turbo* είναι χρήσιμη, αλλά δεν παρέχει την απαιτούμενη σε βάθος θεωρητική αναφορά.

Αντίστοιχα, συγκρίνοντας τις απαντήσεις από τις εκδόσεις *Claude 2* και *Claude 3*, παρατηρούμε διαφορές τόσο στην ανάλυση όσο και την παρουσίαση της σημασιολογικής δομής της πρότασης.

Claude 2

Here is the semantic annotation for the Greek clause according to Talmy's typology: Το παιδί (Figure) βγήκε (Motion verb) στον κήπο (Ground) τρέχοντας (Manner) μέσα από το γεφυράκι (Path)
The key elements:

- Το παιδί (to paidi) = Figure (the moving object) - βγήκε (vghike) = Motion verb expressing the core schema of Motion - στον κήπο (ston kipo) = Ground (the reference object) - τρέχοντας (trechontas) = Manner (the way the motion occurs) - μέσα από το γεφυράκι (mesa apo to gefyraki) = Path (the trajectory of the motion) So this clause exhibits the typological pattern of a **satellite-framed language** like Greek, where the Path is encoded in a satellite and the Manner is expressed in the main

Claude 3

According to Talmy's typology, languages can be classified into two main types based on how they encode motion events: **verb-framed languages and satellite-framed languages**.

Greek is a verb-framed language, which means that the core schema or Path of motion is typically encoded in the main verb, while other co-event components, such as Manner and Cause, are expressed through adjuncts or subordinate elements.

In the given Greek clause, “Το παιδί βγήκε στον κήπο τρέχοντας μέσα από το γεφυράκι” (The child exited into the garden running through the small bridge), the semantic annotation based on Talmy's typology would be as follows:

verb. The Motion verb expresses the core schema, while Path and Manner are encoded in adjuncts.	1. Main Event: [Motion (βγήκε ‘exited’)] - The main verb “βγήκε” (exited) encodes the Path of motion, which is the core schema in this verb-framed language. 2. Co-Event 1: [Manner (τρέχοντας ‘running’)] - The gerund “τρέχοντας” (running) expresses the Manner of motion, which is a co-event component. 3. Co-Event 2: [Path (μέσα από το γεφυράκι ‘through the small bridge’)] - The prepositional phrase “μέσα από το γεφυράκι” (through the small bridge) further specifies the Path of motion, which is a co-event component. 4. Ground: [Ground (στον κήπο ‘into the garden’)] - The prepositional phrase “στον κήπο” (into the garden) denotes the Ground or reference object in relation to the Path of motion. 5. Figure: [Figure (Το παιδί ‘The child’)] - “Το παιδί” (The child) is the Figure or moving entity.
--	---

Στο *Claude 2*, η απάντηση παρέχει μια σαφή και περιεκτική ανάλυση, χρησιμοποιώντας κατηγορίες όπως Figure, Motion verb, Ground, Manner και Path. Ωστόσο, η αναφορά στο “Ground” για τη φράση “στον κήπο” δεν είναι ακριβής σύμφωνα με την τυπολογία του Talmy, καθώς θα έπρεπε να χαρακτηριστεί ως Τοποθεσία Προορισμού. Από την άλλη, στο *Claude 3*, προσφέρεται μια πιο λεπτομερή ανάλυση, επισημαίνοντας τη διάκριση μεταξύ “verb-framed” και “satellite-framed” γλωσσών. Αναφέρει ότι η ελληνική γλώσσα είναι “verb-framed”, κάτι που δεν είναι ακριβές, καθώς πρέπει να αναγνωριστεί η “satellite-framed” φύση της. Παρ’ όλα αυτά, η ανάλυση είναι πιο οργανωμένη, περιγράφοντας κάθε στοιχείο της πρότασης με αναφορά σε Main Event, Co-Event (1, 2) και Ground. Η διάρθρωση της ανάλυσης είναι σαφής και κατανοητή, προσφέροντας ένα πιο στέρεο θεωρητικό πλαίσιο. Συμπερασματικά, η απάντηση του *Claude 3* είναι πιο ολοκληρωμένη και οργανωμένη, προσφέροντας μια πιο δομημένη ανάλυση των στοιχείων της πρότασης. Αν και το *Claude 2* παρέχει μια κατανοητή και περιεκτική ανάλυση, η αναφορά στο “Ground” και η έλλειψη θεωρητικής βάσης περιορίζουν την αποτελεσματικότητα της παρουσίασης. Η απάντηση του *Claude 3*, αν και δεν είναι απολύτως ακριβής στην κατηγοριοποίηση της ελληνικής γλώσσας, προσφέρει αναλύει καλύτερα τις σχέσεις μεταξύ των στοιχείων της κίνησης.

Επομένως, αν συγκρίνουμε όλα τα μοντέλα, το *ChatGPT 4-turbo* φαίνεται να επιστρέφει τα καλύτερα αποτελέσματα συνολικά, λόγω της πληρότητας, της σαφήνειας και της αναλυτικής προσέγγισης. Ακολουθεί το *ChatGPT 4* ως μια επίσης ισχυρή επιλογή. Το

Claude 3 προσφέρει καλή ανάλυση, αλλά με μια ανακρίβεια στη γλωσσική κατηγοριοποίηση. Το *ChatGPT 3.5-turbo* και το *Claude 2* είναι λιγότερο αναλυτικά και δεν προσφέρουν την απαιτούμενη θεωρητική βάση.

5. Τα LLMs στην εκπαίδευση

Η αξιοποίηση των LLMs στην εκπαίδευση προσφέρει σημαντικές ευκαιρίες για τη βελτίωση της μάθησης και της διδασκαλίας. Ωστόσο, τα πειράματα (όπως παρουσιάστηκαν προηγουμένως), που πραγματοποιήθηκαν αναφορικά με τα ρήματα κίνησης στα γλωσσικά μοντέλα, αναδεικνύουν την ανάγκη για αξιολόγηση της ορθότητας των αποτελεσμάτων από τον ίδιο τον χρήστη ανά περίπτωση.

Παρά τις προόδους στις γλωσσικές τεχνολογίες, κανένα από τα γλωσσικά μοντέλα που δοκιμάστηκαν δεν παρήγαγε απόλυτα σωστή ανάλυση των ρηματικών δομών, ενώ παρατηρήθηκαν ελλείψεις και ανακρίβειες στις αναφορές τους. Οπότε, υπογραμμίζεται emphatically η σημασία της ανθρώπινης κρίσης στην αξιολόγηση των αποτελεσμάτων που παρέχουν αυτά τα μοντέλα, ιδίως σε περιπτώσεις που αφορούν πολύπλοκες γλωσσικές κατηγορίες.

Συνεπώς, η αποτελεσματική χρήση αυτών των μοντέλων απαιτεί προσοχή, καθώς είναι απαραίτητη η ανάπτυξη της κριτικής σκέψης, η ορθή καθοδήγηση κατά τη χρήση τους και την υιοθέτηση σύγχρονων εκπαιδευτικών εργαλείων.

Πιο συγκεκριμένα, η διδασκαλία δεξιοτήτων που προάγουν την κριτική σκέψη είναι θεμελιώδης για την αποτελεσματική αξιοποίηση των LLMs. Μέσω της ανάπτυξης αναλυτικής και ανεξάρτητης σκέψης, οι μαθητές/τριες μαθαίνουν να αξιολογούν την ποιότητα και την αξιοπιστία των πληροφοριών που προκύπτουν από αυτά. Η ικανότητα αυτή είναι κρίσιμη σε έναν κόσμο όπου οι πληροφορίες είναι άμεσα διαθέσιμες, αλλά η ακρίβεια τους ποικίλλει και είναι αμφισβητήσιμη. Ενθαρρύνοντας τους/τις μαθητές/τριες να σκέφτονται κριτικά, προάγουμε την υπεύθυνη χρήση της τεχνολογίας.

Επιπλέον, η καθοδήγηση στη χρήση των LLMs είναι καίριας σημασίας για την αποτελεσματική αλληλεπίδραση με αυτά. Η διδασκαλία τεχνικών και στρατηγικών για τον σχεδιασμό αποδοτικών προτροπών (prompts) επιτρέπει στους/τις μαθητές/τριες να εκμεταλλεύονται πλήρως τις δυνατότητες των LLMs. Οι προτροπές αποτελούν τον κρίσιμο σύνδεσμο ανάμεσα στους χρήστες και τα μοντέλα, και η ικανότητα δημιουργίας κατάλληλων ερωτήσεων ή εντολών μπορεί να καθορίσει την ποιότητα των απαντήσεων. Εξοπλίζοντας τους/τις μαθητές/τριες με αυτές τις γνώσεις, ενισχύουμε την ικανότητά τους να αξιοποιούν την τεχνολογία με καινοτόμο και αποδοτικό τρόπο.

Απαραίτητη κρίνεται και η χρήση σύγχρονων εκπαιδευτικών εργαλείων που αξιοποιούν τα Μεγάλα Γλωσσικά Μοντέλα για εκπαιδευτικούς σκοπούς, προσφέροντας νέες δυνατότητες στη διαδικασία της μάθησης. Αυτά τα εργαλεία παρέχουν διαδραστική μάθηση και ευκαιρίες για πειραματισμό με γλωσσικά δεδομένα, διευκολύνοντας έτσι την ανάπτυξη δεξιοτήτων και τη δημιουργική σκέψη. Μέσω της αλληλεπίδρασης με τα γλωσσικά μοντέλα, οι μαθητές/τριες μπορούν να εξερευνούν διάφορες πτυχές της γλώσσας και να συμμετέχουν σε δραστηριότητες που ενθαρρύνουν την εξερεύνηση και την καινοτομία. Αυτή η εμπειρία δεν μόνο ενισχύει την κατανόηση των γλωσσικών εννοιών, αλλά και προάγει τη χρήση της τεχνολογίας ως εργαλείου υποστήριξης της μάθησης.

Τέλος, σοβαρές ανησυχίες ανακύπτουν σε ηθικό επίπεδο, ιδίως όσον αφορά την εκπαίδευση, καθώς η χρήση συστημάτων τεχνητής νοημοσύνης για αυτόματη λήψη αποφάσεων, όπως η αξιολόγηση μαθητών/τριών ή η επιλογή εκπαιδευτικών προγραμμάτων, εγείρει ερωτήματα σχετικά με τη λογοδοσία, την ευθύνη και τη διαφάνεια. Είναι κρίσιμο να υπάρχουν σαφείς κατευθυντήριες γραμμές και έλεγχος για να διασφαλιστεί ότι οι αποφάσεις λαμβάνονται με δίκαιο και ηθικό τρόπο. Οι ανησυχίες αυτές απαιτούν από τους εκπαιδευτικούς και τους ερευνητές να ενσωματώνουν διαφανείς και υπεύθυνες πρακτικές στον σχεδιασμό και την ανάπτυξη των εκπαιδευτικών συστημάτων τεχνητής νοημοσύνης.

6. Συμπεράσματα

Η μελέτη του σημασιολογικού πεδίου της κίνησης παρουσιάζει ιδιαίτερο ενδιαφέρον, καθώς η κίνηση, είτε φυσική είτε μεταφορική, διαδραματίζει κεντρικό ρόλο στη σύλληψη και την έκφραση των εννοιών και των ιδεών μας. Βασικό στοιχείο προς αυτή την κατεύθυνση είναι η προσθήκη κατάλληλου υλικού στα σχολικά εγχειρίδια, το οποίο να καλύπτει αυτά τα σημασιολογικά πεδία με τρόπο που να είναι προσιτός και κατανοητός στους/τις μαθητές/τριες. Η χρήση γλωσσικών μοντέλων για την κατανόηση των σημασιολογικών πεδίων προσφέρει νέες δυνατότητες στην ερευνητική και εκπαιδευτική κοινότητα.

Η αξιοποίηση, λοιπόν, αυτών των τεχνολογιών στην εκπαίδευση μπορεί να βοηθήσει τους/τις μαθητές/τριες να αναπτύξουν μια πλουσιότερη και επαρκέστερη κατανόηση της γλώσσας και του κόσμου γύρω τους. Επιπλέον, η δημιουργία σύγχρονων παιδαγωγικών εργαλείων που βασίζονται σε αυτές τις τεχνολογίες μπορεί να ενισχύσει τη διδακτική διαδικασία, καθιστώντας την ευέλικτη και άρρηκτα συνδεδεμένη με τις ανάγκες των μαθητών/τριων.

Τέλος, η ενσωμάτωση αυτών των στοιχείων στην εκπαίδευση απαιτεί μια ολιστική προσέγγιση, όπου τα εκπαιδευτικά προγράμματα, τα εγχειρίδια διδασκαλίας και τα παιδαγωγικά εργαλεία συνεργάζονται αρμονικά για τη δημιουργία ενός περιβάλλοντος μάθησης που προωθεί την κριτική σκέψη στη σύγχρονη εποχή. Με την αξιοποίηση των νέων τεχνολογιών και την παροχή πλούσιου και σχετικού εκπαιδευτικού υλικού, μπορούμε να διασφαλίσουμε ότι οι μαθητές/τριες θα είναι καλύτερα προετοιμασμένοι/ες, ώστε να ανταποκριθούν στις προκλήσεις και τις απαιτήσεις της σύγχρονης κοινωνίας.

Ευχαριστίες

Τα πειράματα στα γλωσσικά μοντέλα πραγματοποιήθηκαν με την ευγενική υποστήριξη της Omilia Ltd, η οποία παρείχε πρόσβαση στους απαιτούμενους πόρους και υποδομές.

Βιβλιογραφία

Berman, R.A. & Slobin, D.I. (1994). *Relating events in narrative: A cross linguistic Development study*. Lawrence Erlbaum.

- Buchholz, S. & Marsi, E. (2006). CoNLL-X shared task on multilingual dependency parsing. *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL-X)*, 149-164.
- Γούτσος, Δ., King, P. & Χατζηδάκη, P. (1995). Η χρήση των corpus στη λεξικογραφία και περιγραφή της Νέας Ελληνικής. Στο *Μελέτες για την Ελληνική Γλώσσα. Πρακτικά της 15ης ετήσιας συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Αριστοτέλειου Πανεπιστημίου Θεσσαλονίκης, 11-14 Μαΐου 1994*. Αφοί Κυριακίδη, 843-854.
- Γούτσος, Δ. & Φραγκάκη, Γ. 2015. *Εισαγωγή στη γλωσσολογία σωμάτων κειμένων*. Κάλλιπος.
- Fiotaki, A. & Tsakalis, V. (2025). Design and Development of a Linguistic Resource: ellexis. *Proceedings of the 7th European Congress of Modern Greek Studies*. European Society of Modern Greek Studies (*forthcoming*).
- Hickmann, M., Hendriks, H., Demagny, A-C, Engemann, H., Ochsenbauer, A-K., Iakovleva, T., Ji, Y. & **Soroli, E.** (2022). *Spatial Representation - experimental and cross-linguistic studies. Volume I: Coding Manual*. University of Paris 8 & CNRS.
- Labropoulou, P., Mantzari, E. & Gavrilidou, M. (1996). *Lexicon-Morphosyntactic Specifications: Language Specific Instantiation* (Greek), PP-PAROLE, MLAP: 63-386.
- Levin, B. (1993). *English Verb Classes and Alternations: a Preliminary Investigation*. University of Chicago Press.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. In: CoRR abs/1907.11692. arXiv: 1907.11692. url: <http://arxiv.org/abs/1907.11692>.
- Μικρός, Γ. (2004). Ηλεκτρονικά σώματα κειμένων και ορολογία. Στο Μ. Κατσογιάννου & Ε. Ευθυμίου (επιμ.). *Ελληνική ορολογία: Έρευνα και εφαρμογές*. Καστανιώτης, 139-164.
- Nivre, J., Hall, J., Kübler, S., McDonald, R., Nilsson, J., Riedel, S. & Yuret, D. (2007). The CoNLL 2007 Shared Task on Dependency Parsing. *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL*.
- Sinclair J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
- Slobin, D.I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömquist & L. Verhoeven (eds.), *Relating events in narrative. Typological and contextual perspectives*. Lawrence Erlbaum Associates, 219-257.
- Strömquist, S. & Verhoeven, L. (2004). *Relating Events in Narrative, Volume 2: Typological and Contextual Perspectives*. Routledge.
- Talmy, L. (1985). Lexicalization patterns. Semantic, Structure in lexical form. In T. Shopen (ed.), *Language typology and syntactic description*. Cambridge University Press.
- Talmy, L. (2000). *Toward a cognitive semantics. Volume 1: Concept structuring systems. Volume 2: Typology and process in concept structuring*. MIT Press.
- OpenAI. (2022). *Introducing ChatGPT*. url: <https://openai.com/blog/chatgpt>.

Papafragou, A., Massey, C. & Gleitman, L. (2002). Shake, rattle, ‘n’ roll: The representation of motion in language and cognition. *Cognition*, 84(2), 189-219.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. url: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. & Liu, P.J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21.140: 1-67. url: <http://jmlr.org/papers/v21/20-074.html>.

Touvron, H., Martin, L., Stone, K. et al. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. In: arXiv: 2307.09288 [cs.CL]. url: <https://arxiv.org/pdf/2307.09288.pdf>.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett (eds.), *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc.