

14 Ποσοτική ανάλυση διαφορών μεταξύ ελληνικής και γερμανικής νομικής γλώσσας: Η περίπτωση της μετάφρασης του Αστικού Κώδικα προς τα Γερμανικά

Χρήστος Γιαννούτσος, Δημήτρης Μπουμπάρης

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας μελέτης είναι η εξέταση των κυριότερων χαρακτηριστικών της ελληνικής και της γερμανικής νομικής γλώσσας καθώς και των λειτουργικών μεταφραστικών μονάδων, αντλώντας παραδείγματα από τον ελληνικό Αστικό Κώδικα (ΑΚ) (1940) και τον γερμανικό (2023). Πραγματοποιήσαμε ποσοτική ανάλυση των διαφορών μεταξύ ελληνικής και γερμανικής νομικής γλώσσας, επικεντρώνοντας το ενδιαφέρον μας στη μορφολογία, τη γραμματική και τη σύνταξη κάνοντας χρήση του spaCy και των Universal Dependencies. Επιπλέον, συγκρίναμε τις μεταφράσεις των άρθρων του ελληνικού ΑΚ (1940), που είναι όμοια στο περιεχόμενο με αντίστοιχα του γερμανικού ΑΚ, προς τη γερμανική γλώσσα από (α) τον Δημήτριο Γώγο (1951), (β) τα συστήματα νευρωνικής αυτόματης μετάφρασης DeepL και Google, και (γ) το μεγάλο γλωσσικό μοντέλο GPT-4 στο περιβάλλον του ChatGPT με οδηγίες για τεκμηριωτική, όμοια με το (α), και ομοιολεειτουργική μετάφραση, όμοια με το (β). Τα αποτελέσματά μας δείχνουν πως οι γλωσσικές διαφορές είναι ανιχνεύσιμες όχι μόνο με βάση την απόλυτη συχνότητα αλλά και τη σημασιολογική ομοιότητα των κειμένων.

Quantitative analysis of the differences between Greek and German legal language: The case of translating the Civil Code into German

Christos Giannoutsos, Dimitris Boumparis

ABSTRACT

The aim of this paper is to examine the main features of the Greek and German legal language along with the functional translation units, using examples from the Greek and the German Civil Codes (1940 and 2023 respectively). We conducted a quantitative analysis of the differences between Greek and German legal language, focusing on morphology, grammar and syntax by using spaCy and Universal Dependencies. We then compared translations of Greek Civil Code articles, which are identical in content to their counterparts in the German one, into German by (a) Dimitrios Gogos (1951), (b) DeepL and Google NMT systems, and (c) GPT-4 in ChatGPT. In the latter case, we provided it with instructions for documentary and equifunctional translations, in line with (a) and (b) respectively. Our results suggest that the linguistic differences are identifiable based on both absolute frequency and text's cosine similarity.

0 Εισαγωγή

Στην ελληνόφωνη βιβλιογραφία υπάρχουν μονογραφίες (Παναρέτου, 2009, Σκυτιώτη, 2022) και διδακτορικές διατριβές (Γιουτίκας, 2016, Κοζομπόλης, 2022) για τη νομική γλώσσα εν

γένει, μονογραφίες (Σταυράκης, 2010, Βλαχόπουλος, 2014) για τις ιδιαιτερότητες της ελληνικής νομικής γλώσσας, καθώς και μονογραφίες που τη συγκρίνουν με την αντίστοιχη γερμανική (Βαλεοντής & Κριμπάς, 2014, Βλαχόπουλος, 2014). Μάλιστα, ο Βλαχόπουλος (2014, σ. 77-80) αναφέρεται στη μετάφραση του ελληνικού Αστικού Κώδικα (ΑΚ) προς τη γερμανική γλώσσα από τον Δημήτριο Γώγο (Gogos, 1951), η οποία πραγματοποιήθηκε για να μελετηθεί η επιρροή του γερμανικού ΑΚ στον αντίστοιχο ελληνικό από γερμανόφωνους νομικούς εμπειρογνώμονες, καθώς «ο γερμανικός Αστικός Κώδικας (BGB) διαμόρφωσε όχι μόνο το περιεχόμενο, αλλά και τη δομή της ελληνικής κωδικοποίησης» (Georgiades, 2000, σ. 493). Στο πεδίο της ελληνόφωνης μεταφρασιολογίας, επικρατεί η προσέγγιση του λειτουργισμού, με τη Nord (1997, 2014) να προτείνει τη χρήση λειτουργικών μεταφραστικών μονάδων για την αξιολόγηση των μεταφράσεων.

Στο πλαίσιο της παρούσας εργασίας, θα μελετηθούν τα κυριότερα χαρακτηριστικά της ελληνικής (Βλαχόπουλος, 1999, Βαλεοντής & Κριμπάς, 2014, Βλαχόπουλος, 2014) και της γερμανικής (Weinrich, 1993, Biskup, 2013) νομικής γλώσσας καθώς και τα κυριότερα χαρακτηριστικά των λειτουργικών μεταφραστικών μονάδων, αντλώντας παραδείγματα από τον ελληνικό ΑΚ (1940) και τον γερμανικό. Έπειτα, μέσω των λειτουργικών μεταφραστικών μονάδων, θα προβούμε στην ποσοτική ανάλυση των διαφορών μεταξύ ελληνικής και γερμανικής νομικής γλώσσας επικεντρώνοντας το ενδιαφέρον μας στη μορφολογία, τη γραμματική και τη σύνταξη με τη χρήση του spaCy (Honribal & Montani, 2017) και των Universal Dependencies (Nivre κ.ά., 2017, Zeman κ.ά., 2022).

Για την ποσοτική ανάλυση, θα προβούμε στη σύγκριση μεταξύ μεταφράσεων των άρθρων του ελληνικού ΑΚ (1940), που είναι όμοια στο περιεχόμενο με αντίστοιχα του γερμανικού, προς τη γερμανική γλώσσα από (α) τον Δημήτριο Γώγο (1951), (β) τα συστήματα νευρωνικής αυτόματης μετάφρασης DeepL και Google, και (γ) το μεγάλο γλωσσικό μοντέλο GPT-4 στο περιβάλλον του ChatGPT με οδηγίες για τεκμηριωτική¹, όμοια με το (α), και ομοιολειτουργική² μετάφραση, όμοια με το (β).

¹ Η τεκμηριωτική μετάφραση (Nord, 2014, σ. 112) έχει ως σκοπό τη δημιουργία στη γλώσσα υποδοχής «ενός είδους τεκμηρίου (ορισμένων πτυχών) μιας επικοινωνιακής διάδρασης, κατά την οποία ο πομπός του πολιτισμού-πηγής επικοινωνεί με τους δέκτες του πολιτισμού-πηγής μέσω του κειμένου-πηγής και κάτω από συνθήκες του πολιτισμού-πηγής» (Nord, 2014, σ. 112). Ουσιαστικά, η τεκμηριωτική μετάφραση υποστηρίζει τη μέθοδο της κατά λέξη μετάφρασης, δηλαδή της πιστής και γραμμικής απόδοσης του κειμένου αφετηρίας στη γλώσσα υποδοχής.

² Η ομοιολειτουργική μετάφραση (Nord, 2014, σ. 118), στην οποία «η λειτουργία του κειμένου-στόχου είναι όμοια με αυτή του κειμένου-πηγής» (Nord, 2014, σ. 118). Παραδείγματα ομοιολειτουργικής

1 Χαρακτηριστικά γερμανικής και ελληνικής νομικής γλώσσας

Σύμφωνα με τον Κουτσιβίτη (1994, σ. 225 στο Βλαχόπουλος, 2014, σ. 46), η γερμανική νομική γλώσσα παρουσιάζει αρκετά όμοια στοιχεία σε σχέση με την ελληνική νομική γλώσσα. Αυτό οφείλεται στο γεγονός ότι και τα δύο νομικά συστήματα ανήκουν στη ρωμανογερμανική οικογένεια δικαίου και η ελληνική νομική γλώσσα αποτελεί δημιούργημα της μετάφρασης από τη γερμανική νομική γλώσσα. Επιπλέον, τόσο η γερμανική (Biskup, 2013, σ. 17) όσο και η ελληνική νομική γλώσσα (Βαλεοντής & Κριμπάς, 2014, σ. 30) περιέχουν εκφράσεις που απορρέουν από τη γενική γλώσσα.

Κύριο χαρακτηριστικό της γερμανικής νομικής ορολογίας, όπως αναφέρει ο Βλαχόπουλος (1999, σ. 146, 2014, σ. 45-46), αποτελεί το ουσιαστικό. Όπως παρατηρούμε στα παρακάτω παραδείγματα, όπου τα κείμενα είναι όμοια ως προς το περιεχόμενο (AK 173, BGB § 133), και στη γερμανική και στην ελληνική νομική γλώσσα το ουσιαστικό χρησιμοποιείται όχι μόνο ως λεξιλογικό (*έρμηνειαν*, *Auslegung*) αλλά και ως συντακτικό φαινόμενο (*τῆς δηλώσεως βουλήσεως*, *Willenserklärung*).

Κατὰ τὴν ἐρμηνείαν τῆς δηλώσεως βουλήσεως ἀναζητεῖται ἡ ἀληθὴς βούλησις ἄνευ προσηλώσεως εἰς τὰς λέξεις. (AK 173).

Bei der Auslegung einer Willenserklärung ist der wirkliche Wille zu erforschen und nicht an dem buchstäblichen Sinne des Ausdrucks zu haften. (BGB § 133)

Επιπλέον, με βάση το γερμανικό παράδειγμα, παρατηρούμε ότι σε επίπεδο ορολογίας τα ουσιαστικά στη γερμανική νομική γλώσσα αποτελούν και σύνθετες λέξεις, χαρακτηριστικό που διαφοροποιεί τη γερμανική από την ελληνική νομική ορολογία (π.χ. *δήλωση βούλησης*, *Willenserklärung*). Όπως αναφέρει ο Weinrich (1993, σ. 924 στο Βλαχόπουλος, 2014, σ. 46-47), «πρόκειται για δομές που συγκροτούνται από επιμέρους λεξιλογικά συστατικά, που το καθένα είναι φορέας σημαντικής έννοιας, και σημασιολογικά συνδέονται μεταξύ τους μέσα από τον περιορισμό της έννοιας του ενός μεμονωμένου ουσιαστικού από το άλλο».

Παρά τις ομοιότητες με τη γερμανική νομική γλώσσα, η ελληνική νομική γλώσσα παρουσιάζει και ιδιαιτερότητες (Βαλεοντής & Κριμπάς, 2014, σ. 45-47) ως προς τη φύση της λόγω της επικράτησης της καθαρεύουσας στο παρελθόν. Αυτό έχει ως αποτέλεσμα άλλοτε τη χρήση βεβιασμένης καθαρεύουσας (βλ. *προσημειούχος/προσημειωσούχος*, βλ. Βαλεοντής & Κριμπάς, 2014, σ. 180-188) και άλλοτε τη χρήση βεβιασμένης δημοτικής (βλ. *αποσβήνεται* αντί *αποσβέννυται*, AK 1169).

μετάφρασης σύμφωνα με την Nord (2014, σ. 119) αποτελούν τα τεχνικά εγχειρίδια, οδηγίες χρήσης, συνταγές κτλ.

Στη συνέχεια, θα μελετήσουμε τις λεξιλογικές μονάδες της ελληνικής νομικής γλώσσας κατά τη μετάφραση προς την αντίστοιχη γερμανική, υπό το πρίσμα του λειτουργισμού (Nord, 1997, 2014), της κυρίαρχης πλέον θεωρητικής προσέγγισης στο πεδίο της μεταφρασιολογίας (Λάμπρου, 2020, σ. 347).

2 Μεθοδολογία της έρευνας

Η Μεθοδολογική προσέγγιση της μελέτης αυτής βασίζεται στην Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing – NLP) μέσω εργαλείων ανάλυσης ηλεκτρονικών σωμάτων κειμένων (corpus linguistics) και εξόρυξης κειμενικής πληροφορίας (text mining). Οι μετρήσεις μας πραγματοποιήθηκαν με χρήση κώδικα Python και στηρίχθηκαν κυρίως – αλλά όχι αποκλειστικά – στο spaCy³, μια βιβλιοθήκη ανοικτού κώδικα (open source library) που περιέχει εργαλεία ανάλυσης φυσικής γλώσσας.

2.1 Δείγμα

Δεδομένου ότι άρθρα του ελληνικού Αστικού Κώδικα έχουν βασιστεί στον γερμανικό, το δείγμα μας περιείχε αποκλειστικά άρθρα ή μέρη αυτών που αποτελούν μετάφραση του δεύτερου. Για μεγαλύτερη διευκόλυνση κατά την ανάλυση καθώς και για εξισορρόπηση του μήκους των περιόδων κατά το δυνατόν, τα άρθρα χωρίστηκαν ανά περιόδους, δηλαδή στις τελείες. Ο μέσος αριθμός λέξεων ανά περίοδο ανέρχεται σε 19,5 και 20 λέξεις για τα ελληνικά και τα γερμανικά αντίστοιχα. Το δείγμα μας χωρίστηκε (α) ανά γλώσσα του κειμένου (DE: γερμανικά και EL: ελληνικά) και (β) ανά έκδοση στην οποία συναντάται (ελλαΚ-1940, ελλαΚ-2023, ελλαΚ-1951 και γεραΚ-2023), ώστε να έχουμε τη δυνατότητα να πραγματοποιήσουμε τις μετρήσεις μας και από τις δύο «οπτικές γωνίες». Στην παρούσα μελέτη, θα επικεντρωθούμε σε παραδείγματα που ποικίλλουν ως προς τις πληροφορίες που εξορύξαμε, ώστε να δώσουμε μια αντιπροσωπευτική εικόνα της ανάλυσης.

2.2 Υπολογιστική ανάλυση φυσικής γλώσσας

Σκοπός της υπολογιστικής ανάλυσης (computational analysis) που διεξάγουμε είναι, πρωταρχικά, να εξετάσουμε αν είναι δυνατή η ποσοτικοποίηση των προς ανάλυση κειμένων και, στη συνέχεια, να εντοπίσουμε τις (στατιστικά σημαντικές) μεταξύ τους διαφορές. Για να το πετύχουμε αυτό, εξαγάγαμε τα βασικά κειμενικά χαρακτηριστικά τους με τη χρήση ημιαυτόματων μεθόδων. Καταρχάς, προβήκαμε σε έναν απλό αλλά σημαντικό «καθαρισμό» (προεπεξεργασία, pre-processing) των κειμένων από ετεροαναφορές (cross-references),

³ <https://spacy.io>.

δηλαδή παραπομπές σε άλλα άρθρα ή παραγράφους του ΑΚ. Επιπλέον, βεβαιωθήκαμε πως κατά την ανάλυση το κείμενο μετατράπηκε σε αμιγώς πεζούς χαρακτήρες, ώστε να αποτραπεί το ενδεχόμενο να αναγνωρίζεται η ίδια λεκτική μονάδα ως ξεχωριστή αν συναντάται με κεφαλαίο και με πεζό. Έτσι, για παράδειγμα, η πρωτότυπη περίοδος:

Bei der Auslegung einer Willenserklärung ist der wirkliche Wille zu erforschen und nicht an dem buchstäblichen Sinne des Ausdrucks zu haften. (BGB § 133)

Μετατρέπεται, κατά την ανάλυσή της από τον αλγόριθμο, σε:

bei der auslegung einer willenserklärung ist der wirkliche wille zu erforschen und nicht an dem buchstäblichen sinne des ausdrucks zu haften.

Ο αλγόριθμος προβαίνει στις εξής μετρήσεις για την ανάλυση κάθε περιόδου:

- **Γλώσσα εντοπισμού:** ο κωδικός ISO 639-1:2002 της γλώσσας που ο αλγόριθμος εντόπισε αυτόματα, ώστε να προβεί στη σωστή γλωσσική ανάλυση παρακάτω.
- **Αριθμός λέξεων:** ο συνολικός αριθμός λέξεων κάθε περιόδου μετά την προεπεξεργασία της.
- **Συχνότητα λέξεων:** η απόλυτη συχνότητα κάθε λέξης της περιόδου.
- **Συχνότητα λημμάτων:** η απόλυτη συχνότητα κάθε λήμματος της περιόδου.
- **Διαφορά συχνότητας λέξεων-λημμάτων:** η συνολική αριθμητική διαφορά των δύο παραπάνω συχνοτήτων.
- **Μέρη του λόγου:** λίστα με όλες τις ετικέτες του μέρους του λόγου για κάθε λέξη της περιόδου όπως ορίζονται από το Universal Dependencies (Nivre κ.ά., 2016).
- **Συχνότητα μερών του λόγου:** η απόλυτη συχνότητα κάθε ετικέτας μερών του λόγου της περιόδου.
- **Μορφολογικά χαρακτηριστικά:** τα μορφολογικά χαρακτηριστικά κάθε λέξης της περιόδου με βάση το μέρος του λόγου στο οποίο ανήκουν.
- **Συντακτικές εξαρτήσεις:** λίστα με τη συντακτική εξάρτηση κάθε λέξης της περιόδου.
- **Λήμματα:** λίστα με τα μοναδικά λήμματα της περιόδου.
- **Διανύσματα λέξεων:** λίστα με τη διανυσματική απεικόνιση (300 διαστάσεις) της κάθε λέξης της περιόδου, υπολογισμένα με το fastText⁴.
- **Μέσο διάνυσμα περιόδου:** ο μέσος όρος των διανυσμάτων της περιόδου.

Επίσης, αξιοποιήσαμε την τεχνολογία του spaCy για οπτικοποίηση διαστημάτων (span

⁴ <https://fasttext.cc>.

visualization⁵) ώστε να εξαγάγουμε με ημιαυτόματο τρόπο τις ορολογικές, γλωσσικές και γραμματικοσυντακτικές δομές που αποτελούν γνωρίσματα της ελληνικής και της γερμανικής νομικής γλώσσας. Σύμφωνα με τη Nord (2014, σ. 160-162), οι λειτουργικές μονάδες στη νομική μετάφραση, και συγκεκριμένα κατά τη μετάφραση του ΑΚ 173 και της μετάφρασης προς τη γερμανική γλώσσα από τον Δημήτριο Γώγο (Gogos, 1951, σ. 19) έχουν ως εξής:

ελλαΚ-1940	Κατὰ τὴν <u>ΕΡΜΗΝΕΙΑΝ ΤΗΣ ΔΗΛΩΣΕΩΣ ΒΟΥΛΗΣΕΩΣ</u> ἀναζητεῖται Ἡ <u>ἈΛΗΘΗΣ ΒΟΥΛΗΣΙΣ</u> ἄνευ προσηλώσεως εἰς τὰς λέξεις.
ελλαΚ-1951	Bei der <u>AUSLEGUNG EINER WILLENSERKLÄRUNG</u> ist DER WIRKLICHE <u>WILLE</u> zu erforschen, ohne an den Worten zu haften .

Παράδειγμα 1: Οι μεταφραστικές μονάδες σύμφωνα με τη Nord (2014, σ. 161) στο άρθρο ΑΚ 173 (έκδοση 1940) και στη μετάφρασή του στα Γερμανικά (Gogos, 1951).

Όπου:

- (1) ΚΕΦΑΛΑΙΑ: θεματική οργάνωση.
 - Υπερθέμα: (ΕΡΜΗΝΕΙΑΝ ΤΗΣ ΔΗΛΩΣΕΩΣ ΒΟΥΛΗΣΕΩΣ, AUSLEGUNG EINER WILLENSERKLÄRUNG),
 - Θέμα 1^ο: (Ἡ ΑΛΗΘΗΣ ΒΟΥΛΗΣΙΣ, DER WIRKLICHE WILLE) κτλ.
- (2) Απλή υπογράμμιση: γλωσσικά χαρακτηριστικά του συγκεκριμένου κειμενικού τύπου, όπως ρηματικές δομές και νομικοί όροι.
- (3) Πλάγια γραφή: ιδιαίτερες απόψεις του πομπού.
- (4) **Έντονη γραφή**: προσανατολισμός στον δέκτη.

Τα χαρακτηριστικά της γερμανικής γλώσσας περιλαμβάνουν (Nord, 2014, σ. 161):

1. Προτίμηση των ονοματικών δομών (*Auslegung, Wille*).
2. Αναλογικό σχηματισμό λέξεων, δηλαδή λατινισμούς και σύνθετα ουσιαστικά (*Willenserklärung*).
3. Σαφείς ενδείκτες θεματικής οργάνωσης λόγω της πολύπλοκης συντακτικής δομής των γερμανικών προτάσεων (π.χ. *ohne ... zu + απαρέμφατο*).

2.3 Σύγκριση ανθρώπινης και αυτόματης μετάφρασης

Εκτός από τις δύο εκδόσεις στα ελληνικά (ελλαΚ-1940, ελλαΚ-2023) και τις αντίστοιχες γερμανικές (ελλαΚ-1951 και γερΑΚ-2023), προβήκαμε στην ίδια ανάλυση και για τη μηχανική μετάφραση των πρωτοτύπων στα γερμανικά. Για τον σκοπό αυτό,

⁵ <https://spacy.io/usage/visualizers#span>.

χρησιμοποιήσαμε δύο συστήματα νευρωνικής αυτόματης μετάφρασης (DeepL Translator και Google Translate) και το μεγάλο γλωσσικό μοντέλο (Large Language Model – LLM) GPT-4 της OpenAI στο περιβάλλον του ChatGPT, δίνοντάς του συγκεκριμένες μεταφραστικές οδηγίες προτού προβεί στη μετάφραση. Τα ερωτήματά μας σε αυτό το μέρος της μελέτης μας είναι:

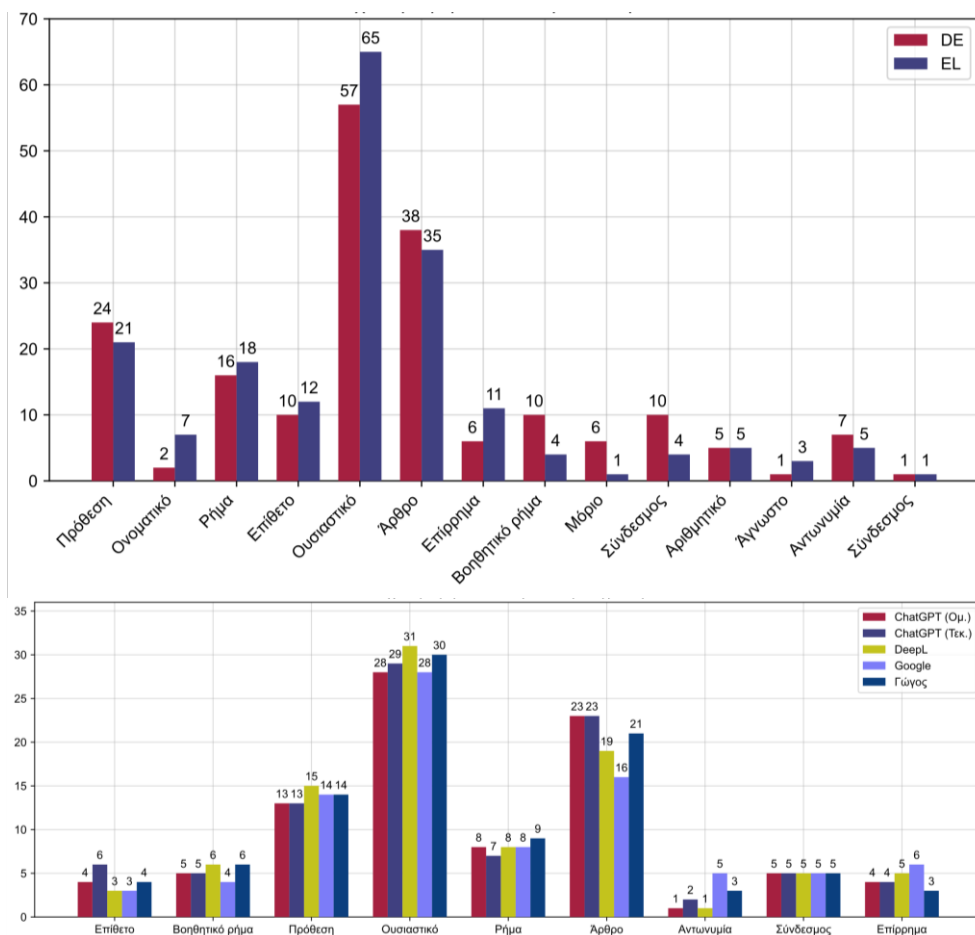
- Διαφέρει η αυτόματη μετάφραση από την ανθρώπινη και, αν ναι, σημαντικά;
- Πλησιάζει, σε επίπεδο νοήματος, η μετάφραση από συστήματα τεχνητής νοημοσύνης την ανθρώπινη αν τους δοθούν συγκεκριμένες μεταφραστικές οδηγίες;

Παρόλο που τα συστήματα αυτόματης νευρωνικής μετάφρασης υπάρχουν εδώ και σχεδόν μία δεκαετία (αρχής γενομένης με τους Bahdanau κ.ά., 2014 και Sutskever κ.ά., 2014) και τα μεγάλα γλωσσικά μοντέλα για περίπου το μισό διάστημα αυτό (βλ. ενδεικτικά Devlin κ.ά., 2017 και Radford κ.ά., 2018), τα δεύτερα έγιναν γνωστά στο ευρύ κοινό στα τέλη του 2022, με τη δημόσια και ελεύθερη διάθεση του ChatGPT της OpenAI⁶. Το ChatGPT αποτελεί προϊόν της εν λόγω εταιρείας που βασίζεται στις δύο τελευταίες εκδόσεις του μεγάλου γλωσσικού της μοντέλου Generative Pre-Trained Transformer (GPT), το GPT-3.5 και το GPT-4, συνδυάζοντας τις δυνατότητές τους για παραγωγή κειμένου σε σχεδόν ανθρώπινο επίπεδο με δυνατότητα παραγωγής διαλόγου με τον χρήστη. Αυτό το ιδιαίτερο γνώρισμα των μοντέλων αυτών, που απορρέει από τη συμπερίληψη ενός τρίτου μεγάλου γλωσσικού μοντέλου (InstructGPT) στην «οικογένεια» των GPT ικανού να ακολουθεί οδηγίες και να παράγει διάλογο (Ouyang κ.ά., 2022), μας επιτρέπει να του αναθέσουμε τον ρόλο του μεταφραστή, παρόλο που δεν έχει εκπαιδευτεί αποκλειστικά στο να μεταφράζει (OpenAI, 2023, σ. 29). Οι παράμετροι που χρησιμοποιήθηκαν είναι: *temperature* = 0.5, *top P* = 1 και *frequency penalty* = *presence penalty* = 0.

Γώγος	Bei der Auslegung einer Willenserklärung wird der wirkliche Wille erforscht, ohne an den Worten zu haften.
DeepL	Bei der Auslegung einer Willenserklärung wird der wahre Wille ohne Bindung an den Wortlaut gesucht.
Google	Bei der Auslegung der Willenserklärung sucht man nach einem wahren Willen ohne Bindung an die Worte.
ChatGPT (Τεκ.)	Bei der Auslegung der Willenserklärung wird die tatsächliche Absicht ohne strikte Bindung an die Worte ermittelt.
ChatGPT (Ομ.)	Bei der Auslegung der Willenserklärung wird die wahre Absicht ohne Bindung an die Worte gesucht.

Οι μεταφράσεις του παραδείγματος στα γερμανικά. (Ομ.: ομοιολεειτουργική, Τεκ.: τεκμηριωτική)

⁶ Αρχαική έκδοση, αποθηκευμένη στο WebArchive (Διαθέσιμο στις 30 Νοεμβρίου 2022): <https://web.archive.org/web/20221130180912/https://openai.com/blog/chatgpt/>.



Γράφημα 1: Απόλυτη συχνότητα των μερών του λόγου στο δείγμα μας ανά νλώσσα (επάνω) και ανά μεταφραστή στα νεογερμανικά (κάτω).

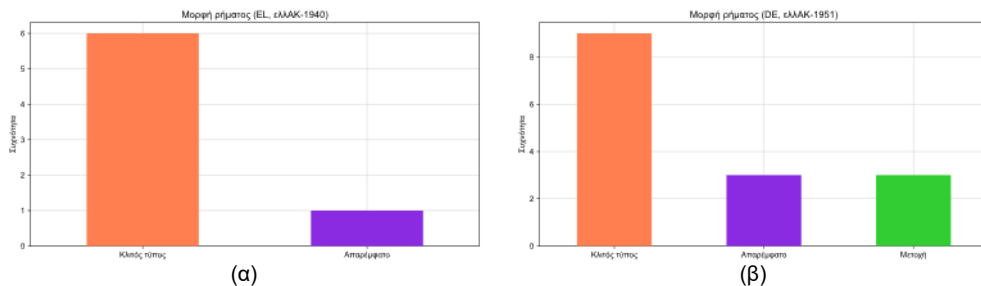
3 Αποτελέσματα

3.1 Μορφολογική, γραμματική και συντακτική ανάλυση

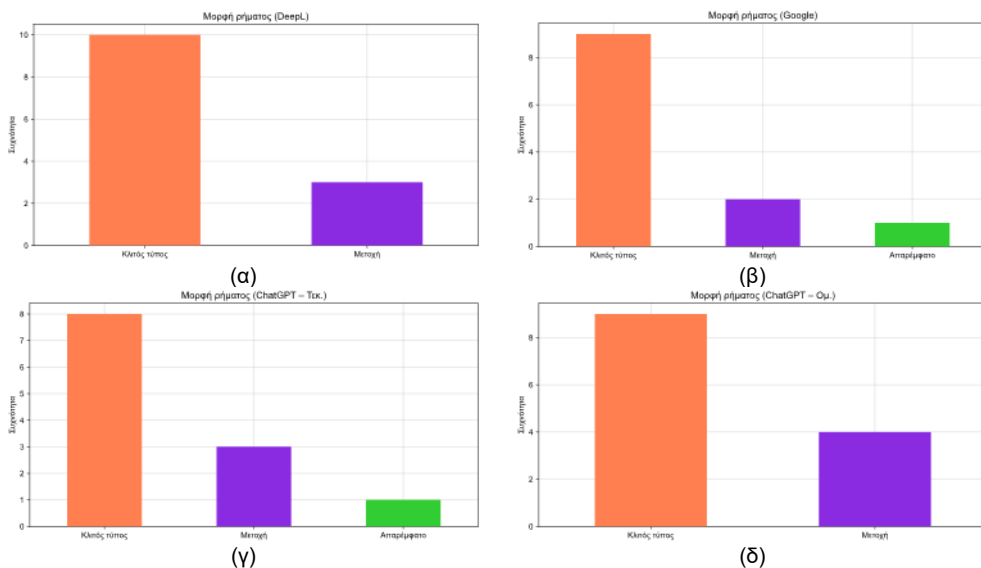
Αν συγκρίνουμε τα κείμενα του δείγματός μας με βάση τη γλώσσα (Γράφημα 1), βλέπουμε πως η χρήση μερών του λόγου δεν διαφέρει στα βασικά σημεία (άρθρα, ουσιαστικά, επίθετα, ρήματα), όμως, εμφανίζει σημαντική διαφορά στη χρήση των βοηθητικών ρημάτων, των συνδέσμων και των μορίων, καθώς στα γερμανικά η χρήση του παρατηρείται πολλαπλάσιες φορές αυτής στα ελληνικά (από 2,5 έως και 6 φορές). Όσον αφορά στα βοηθητικά ρήματα, η χρήση τους στη γερμανική διαφέρει από την ελληνική, καθώς στην

πρώτη χρησιμοποιούνται στη δημιουργία δομών που μπορούν να αποτυπωθούν στα ελληνικά με απλούς ρηματικούς τύπους (π.χ. *ist zu erforschen* = αναζητείται).

Επιπλέον, ως προς τη μορφή των ρηματικών τύπων, παρατηρούμε πως, στην ανθρώπινη μετάφραση (Γράφημα 2), η χρήση του απαρεμφάτου είναι πολύ πιο συχνή στα γερμανικά, δεδομένου πως στα ελληνικά ο τύπος αυτός χρησιμοποιείται πολύ πιο περιορισμένα. Το ίδιο ισχύει και για τις μετοχές, οι οποίες ναι μεν χρησιμοποιούνται πιο συχνά, ειδικά στη νομική γλώσσα χάριν της ουσιαστικοποίησης (Βλαχόπουλος, 2014, σ. 48), ωστόσο, εδώ, δεν συναντώνται στον ελληνικό ΑΚ του 1940 (Γράφημα 2β).



Γράφημα 2: Απόλυτη συχνότητα των ρηματικών τύπων στο δείγμα μας.
(α) στον ΑΚ 1940, (β) στη μετάφρασή του από τον Γώγο.



Γράφημα 3: Απόλυτη συχνότητα των ρηματικών τύπων στις μεταφράσεις.
(α)-(β) DeepL και Google (γ)-(δ) ChatGPT. (Ομ.: ομοιολεειτουργική, Τεκ.: τεκμηριωτική)

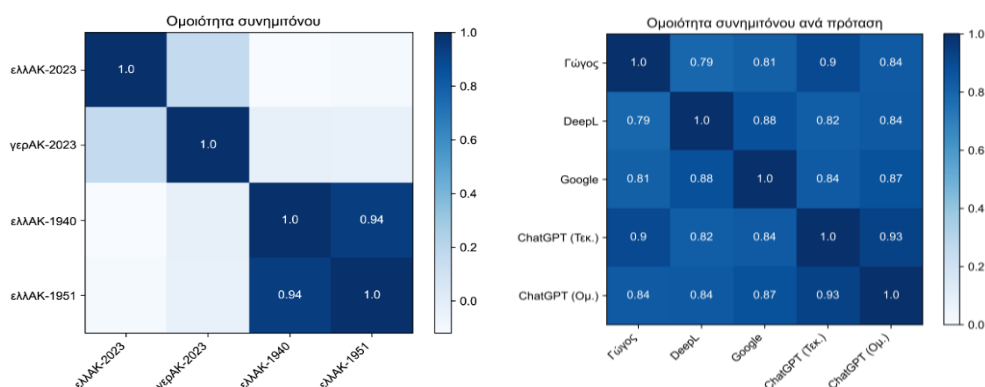
Στην περίπτωση της αυτόματης μετάφρασης (Γράφημα 3), έχουμε τα αντίθετα αποτελέσματα: Όταν δοθούν οδηγίες για ομοιολεειτουργική μετάφραση, το ChatGPT (GPT-4) φαίνεται να αποφεύγει τη χρήση απαρεμφάτων (ή, έστω, να μην τα προτιμάει), όπως ακριβώς και το DeepL, ενώ στην περίπτωση της τεκμηριωτικής μετάφρασης, ισχύει το αντίθετο, μια στρατηγική που ακολουθεί και το σύστημα νευρωνικής μετάφρασης της Google.

3.2 Σημασιολογική ομοιότητα

Η παραπάνω ανάλυση βασίζεται στις λέξεις καθαυτές που συναντώνται στα κείμενα που διαλέξαμε. Πολλές φορές, ωστόσο, κείμενα εμφανίζουν σημασιολογική ομοιότητα ακόμα και όταν χρησιμοποιούν διαφορετικές λέξεις, ειδικά στη δική μας περίπτωση αφού μελετούμε πρωτότυπα κείμενα και τις μεταφράσεις τους. Ένας αξιόπιστος τρόπος να δούμε κατά πόσο

$$\text{συν}(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2 \cdot \sum_{i=1}^n B_i^2}}$$

μοιάζουν τα κείμενα σημασιολογικά είναι η ομοιότητα του συνημιτόνου (cosine similarity). Η αξιοπιστία αυτής της μεθόδου έγκειται στο γεγονός ότι το μέτρο (μήκος) των προς σύγκριση διανυσμάτων (δηλαδή των κειμένων, στην προκειμένη περίπτωση) δεν επηρεάζει τον βαθμό ομοιότητάς τους (Karsdorp κ.ά., 2021, σ. 100-101). Αυτό αποτελεί την ειδοποιό διαφορά της σε σχέση με την ευκλείδεια απόσταση η οποία, με απλοϊκούς όρους, συνδέει τα δύο σημεία (κείμενα) στον χώρο με μια ευθεία γραμμή. Η ομοιότητα του συνημιτόνου, $\text{συν}(\theta)$, για δύο διανύσματα διαστάσεων n , A και B , υπολογίζεται ως το εσωτερικό γινόμενο των διανυσμάτων προς το γινόμενο του μέτρου τους:



Γράφημα 4: Θερμοχάρτες της ομοιότητας συνημιτόνου του Παραδείγματος 1 στις δύο βασικές εκδόσεις και στις εκδόσεις αναφοράς (αριστερά) και μεταξύ των γερμανικών μεταφράσεων του ελλαΚ-1940 (δεξιά).

Για να χρησιμοποιήσουμε αυτή τη μέθοδο, μετατρέψαμε, χρησιμοποιώντας το fastText, τις λέξεις του κάθε κειμένου σε μια σειρά από διανύσματα τα οποία αντικατοπτρίζουν τη σημασία τους, δίνοντας μία τιμή στο κλειστό διάστημα $[-1, 1]$ για κάθε σημασιολογική διάσταση (300 στο σύνολο). Στη συνέχεια, υπολογίσαμε τον μέσο όρο όλων των διανυσμάτων αυτών για να προκύψει το διάνυσμα κάθε περιόδου και, τέλος, τα συγκρίναμε μεταξύ τους ώστε να προκύψει το ποσοστό ομοιότητάς τους. Ας δούμε το παράδειγμα και στις δύο περιπτώσεις:

Διαπιστώνουμε πως το πρωτότυπο κείμενο (ελλαΚ-1940) και η μετάφρασή του στα Γερμανικά από τον Γώγο (ελλαΚ-1951) εμφανίζουν σημασιολογική ομοιότητα 94%, το οποίο είναι αναμενόμενο αφού πρόκειται για πρωτότυπο και μετάφραση. Στην άλλη πλευρά, δεξιά, παρατηρούμε πως, αν συγκρίνουμε μόνο τις γερμανικές μεταφράσεις, το ποσοστό ομοιότητας μεταξύ τους κυμαίνεται μεταξύ 79-90%, με μόνη εξαίρεση τις δύο μεταφράσεις του GPT-4 (με ομοιολεειτουργική και τεκμηριωτική προσέγγιση και ποσοστό 93%), που όμως δεν επιδέχονται σύγκρισης μεταξύ τους καθώς πρόκειται για τον ίδιο «μεταφραστή». Σημειώνουμε πως εδώ δεν επικεντρωνόμαστε στην ποιότητα των μεταφράσεων, αλλά αποκλειστικά στη σημασιολογική τους διάσταση. Σε άλλα παραδείγματα, ωστόσο, παρατηρούνται ελαφρώς υψηλότερα ποσοστά, περίπου 85-95%, αναλόγως με την πολυπλοκότητά τους.

4 Συμπέρασμα

Η μελέτη μας θίγει πολλά ζητήματα, θέτει αρκετά τα ερωτήματα και προσπαθεί να τα απαντήσει δεδομένης της πολυπλοκότητας αλλά και της τυχαιότητας των αποτελεσμάτων, ειδικά από τα συστήματα αυτόματης μετάφρασης και τεχνητής νοημοσύνης. Κρίθηκε αναγκαίο να μη δοθεί βάση στην ποιότητα των μεταφρασμάτων αυτών, καθώς κάτι τέτοιο θα καθιστούσε την έρευνα σαφώς μεγαλύτερη σε έκταση. Η μελέτη μας υποδεικνύει ότι γλωσσικές διαφορές στα νομικά κείμενα μπορούν να ανιχνευτούν και να μετρηθούν, ενώ επίσης υπάρχει σαφής διάκριση της ανθρώπινης από την αυτόματη νομική μετάφραση, και πάλι με γνώμονα τις διαφορές σε επίπεδο γραμματικής, σύνταξης και μορφολογίας και ειδικότερα όσον αφορά σε δομές που απαντώνται κατά κόρον σε τέτοιου είδους κείμενα. Η γενική εικόνα μάς δείχνει πως τέτοιες μελέτες προσφέρουν πολλές αφορμές για περαιτέρω έρευνα που δεν θα περιορίζεται αποκλειστικά στη νομική γλώσσα.

Βιβλιογραφία

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. Στο: *Proceedings of the 3rd International Conference on Learning Representations*; 2015 May 7–9; San Diego, USA. [10.48550/arXiv.1409.0473](https://arxiv.org/abs/1409.0473).

- Biskup, M. (2013). Zu Besonderheiten der Rechtssprache Deutschlands. *Beiträge zur allgemeinen und vergleichenden Sprachwissenschaft, Sprache in Wissenschaft und Unterricht* (σ. 17-27). https://beitraege-contributions.pl/articles/02/02_biskup.pdf (Διαθέσιμο στις 8 Ιουλίου 2023).
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. [10.48550/arXiv.1810.04805](https://arxiv.org/abs/1810.04805).
- Georgiades, A. (2000). Der Einfluß des deutschen BGB auf das griechische Zivilrecht. *Archiv Für Die Civilistische Praxis*, 200 (3/4), 493–502. <http://www.jstor.org/stable/40995728> (Διαθέσιμο στις 8 Ιουλίου 2023).
- Gogos, D. (1951). *Das Zivilgesetzbuch von Griechenland (1940) mit dem Einführungsgesetz*. Walter De Gruyter & Co.
- Honnibal, M. & Montani, I., (2017). *spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing*.
- Karsdorp, F., Kestemont, M., & Riddell, A. (2021). *Humanities Data Analysis: Case Studies with Python*. Princeton University Press.
- Nivre, J., Agić, Ž., Ahrenberg, L., Antonsen, L., Aranzabe, M. J., Asahara, M., Ateyah, L., Attia, M., Atutxa, A., Badmaeva, E., Ballesteros, M., Banerjee, E., Bank, S., Bauer, J., Bengoetxea, K., Bhat, R. A., Bick, E., Bosco, C., Bouma, ... Zhu, H. (2017). *Universal Dependencies 2.0 – CoNLL 2017 Shared Task Development and Test Data*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <http://hdl.handle.net/11234/1-2184> (Διαθέσιμο στις 8 Ιουλίου 2023).
- Nord, C. (1997). *Translating as a Purposeful Activity. Functionalist Approaches Explained*. St Jerome Publishing.
- Nord, C. (2014). *Η Μετάφραση ως στοχευμένη δραστηριότητα. Εισαγωγή στις λειτουργικές προσεγγίσεις* (Σίμος Π. Γραμμενίδης & Δέσποινα Δ. Λάμπρου, Μετ.). Δίαυλος.
- OpenAI. (2023). *GPT-4 Technical Report*. [10.48550/arXiv.2303.08774](https://arxiv.org/abs/2303.08774).
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J. & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. [10.48550/arXiv.2203.02155](https://arxiv.org/abs/2203.02155).
- Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. OpenAI. <https://openai.com/research/language-unsupervised> (Διαθέσιμο 8 Ιουλίου 2023).
- Sutskever, I., Vinyals, O., Le, QV. (2014). Sequence to sequence learning with neural networks. Στο: *Proceedings of the 27th International Conference on Neural Information Processing Systems*; 2014 Dec 8–13; Montreal, QC, Canada. [10.48550/arXiv.1409.3215](https://arxiv.org/abs/1409.3215).
- Weinrich, H. (1993). *Textgrammatik der deutschen Sprache*. Duden.
- Zeman, D., Nivre, J., Abrams, M., Ackermann, E., Aepli, N., Aghaei, H., Agić, Ž., Ahmadi, A., Ahrenberg, L., Ajede, C. K., Akkurt, S. F., Aleksandravičiūtė, G., Alfina, I., Algom, A., Alzetta, C., Andersen, E., Antonsen, L., Aplonova, K., Aquino, A., ... Ziane, R. (2022). *Universal Dependencies 2.11*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <http://hdl.handle.net/11234/1-4923> (Διαθέσιμο στις 8 Ιουλίου 2023).

Βαλεοντής, Κ. Ε. & Κριμπάς, Π. Ι. (2014). *Νομική γλώσσα, νομική ορολογία: Θεωρία και πράξη*. Νομική Βιβλιοθήκη.

Βλαχόπουλος, Σ. (1999). *Η μετάφραση νομικών κειμένων επί τη βάσει της θεωρίας του σκοπού*. [Αδημοσίευτη Διδακτορική Διατριβή, Ιόνιο Πανεπιστήμιο. Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας]. <https://www.didaktorika.gr/eadd/handle/10442/12000> (Διαθέσιμο στις 8 Ιουλίου 2023).

Βλαχόπουλος, Σ. (2014). *Πολυγλωσσία στο δίκαιο: Μετάφραση νομικών κειμένων και δικαστηριακή διερμηνεία*. Νομική Βιβλιοθήκη.

Γιουτίκας, Δ. (2016). *Μετάφραση νομικών κειμένων από τη γαλλική στην ελληνική γλώσσα: η περίπτωση των δικαστικών αποφάσεων σε υποθέσεις διαζυγίων*. [Αδημοσίευτη Διδακτορική διατριβή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τμήμα Γαλλικής Γλώσσας και Φιλολογίας]. <https://www.didaktorika.gr/eadd/handle/10442/39058> (Διαθέσιμο στις 8 Ιουλίου 2023).

Κοζομπόλης, Σ. (2022). *Όψεις τις δεοντικής τροπικότητας στη νομική γλώσσα και μετάφραση του Δικαίου του Ανταγωνισμού της ΕΕ: κριτική επεξεργασία και ανάλυση σώματος κειμένου στα αγγλικά και στα ελληνικά*. [Αδημοσίευτη Διδακτορική Διατριβή, Ιόνιο Πανεπιστήμιο, Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας]. <https://www.didaktorika.gr/eadd/handle/10442/51436> (Διαθέσιμο στις 8 Ιουλίου 2023).

Λάμπρου, Δ. Δ. (2020). *Η διδακτική της διαγλωσσικής μετάφρασης στον ελληνόφωνο χώρο: ιστορική και σύγχρονη μελέτη των προγραμμάτων σπουδών, το προφίλ των διδασκόντων και οι διδακτικές προσεγγίσεις στην ελληνόφωνη βιβλιογραφία*. [Αδημοσίευτη Διδακτορική διατριβή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τμήμα Γερμανικής Γλώσσας και Φιλολογίας]. <https://www.didaktorika.gr/eadd/handle/10442/47589> (Διαθέσιμο στις 8 Ιουλίου 2023).

Παναρέτου, Ε. (2009). *Νομικός Λόγος: Γλώσσα και δομή των νόμων*. Παπαζήση.

Σκυτώτη, Σ. (2022). *Συγκριτικό Δίκαιο & Γλώσσα*. Αντ. Ν. Σάκκουλα Ε.Ε.

Σταυράκης, Α. Ν. (2010). *Νεοελληνική νομική γλώσσα και ορολογία*. Π. Ν. Σάκκουλας.

Χρήστος Γιαννούτσος

Υποψήφιος Διδάκτωρ

Ιόνιο Πανεπιστήμιο

Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας (ΤΞΓΜΔ)

Ηλ-ταχ.: giachr1997@gmail.com

Δημήτρης Μπουμπάρης

Μηχανικός Επεξεργασίας Φυσικής Γλώσσας

Απόφοιτος Ιονίου Πανεπιστημίου (ΤΞΓΜΔ) και Πανεπιστημίου της Αμβέρσας

Ηλ-ταχ.: dimboump@pm.me