Authority vs. Automation: From Corpora to LLMs for Reliable Legal Terminology *

Mirjana Borucinsky[1,*,†] and Martina Bajčić[2,†]

[1] University of Rijeka, Faculty of Maritime Studies, Studentska 2, 51000 Rijeka, Croatia
[2] University of Rijeka, Faculty of Law, Hahlić 6, 51000 Rijeka, Croatia

Abstract

This paper examines the reliability of three approaches to extracting and translating multi-word legal terms (MWT) from English into Croatian within the domain of EU consumer law. We compiled specialized corpora of EU consumer-law texts referring to the digital single market and extracted terminology using a corpus-based methodology. We then translated the obtained MWTs using corpus tools, neural machine translation (NMT), generative AI (LLMs) and compared all outputs with authoritative reference terms from Interactive Terminology for Europe (IATE) and human judgment. When benchmarked against IATE, corpus-based translations achieved the highest accuracy and frequency-weighted accuracy, outperforming both the LLM and NMT outputs. However, when evaluated against expert raters, ChatGPT5.1 achieved the highest term adherence, exceeding both corpus translations and NMT, particularly for high-frequency and legally salient terms. These findings reveal that term accuracy is sensitive to the chosen authority. Based on this case study, corpus methods best replicate institutional terminology, whereas LLMs more closely reflect expert preferences. We argue that neither source alone offers a complete solution. Instead, effective legal translation workflows require the complementary use of LLMs for speed and adaptability, corpora for terminological precision and evidential transparency, and expert human judgment for authority.

Keywords

terminology, multi-word terms, legal translation, corpus, NMT, LLM

1. Introduction

In recent years, generative artificial intelligence models such as ChatGPT have been used extensively for translation and terminology management across many domains, including the legal domain. The intersection between law and language has always presented significant challenges due to the normative, institutional, and intertextual nature of legal discourse [1]. As large language models (LLMs) begin to play a role in terminology management and translation, questions arise concerning their reliability and authority in producing legally credible equivalents across languages.

This paper presents a case study that compares corpus-based, neural machine translation (NMT), and GenAI-generated translation of MWTs in the domain of EU consumer law within the digital single market. It examines the advantages and limitations of these approaches in supporting legal translators and terminologists and discusses the broader implications for trust in AI-assisted legal translation.

2. Theoretical background

Legal language is characterized by normativity, precision, and intertextuality, and legal terms embody institutional authority and conceptual specificity. Terminological consistency is crucial for ensuring legal certainty, particularly within the European Union, where harmonized terminology

* 5th International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2026, June 25-26, 2026, Zadar, Croatia.
[1,*] Corresponding author.
[†] These authors contributed equally.
mirjana.borucinsky@pfri.uniri.hr (M. Borucinsky); martina.bajcic@pravri.uniri.hr (M. Bajčić)
0000-0002-1132-9720 (M. Borucinsky); 0000-0003-0654-6779 (M. Bajčić)

ensures the uniform application of law [2]. At the same time, digitization has introduced new challenges in the translation of legal content.

Inconsistent terminology can threaten the cognitive authority of law. Machine translation (MT) systems have made progress in domain adaptation, yet previous studies indicate persistent issues in translating multi-word expressions (MWEs) and maintaining terminological coherence [3, 4, 5]. LLMs offer promising support in assisting translators and terminologists, yet they require rigorous validation to ensure legal adequacy and conceptual alignment.

3. Methodology

The aim of this study is to compare three approaches to identifying Croatian equivalents for selected EU legal terms in English: corpus-based retrieval, neural machine translation (NMT), and LLM-generated output.

To obtain a list of MWTs three specialized corpora have been compiled from EU consumer-law sources, including regulations, directives, and related preparatory documents. The three corpora were compiled through different retrieval procedures in order to test whether corpus-building strategy influences term extraction and the identification of Croatian equivalents. The corpora comprise EU legal texts relevant to consumer protection in the digital single market, including regulations, directives, and preparatory materials current at the time of corpus compilation. Corpus compilation was carried out between March and June 2025 using Sketch Engine [6], and the corpora are available at the Pravri Institutional repository. The corpora have been compiled using 1. web crawling resulting in a 109595 token corpus, 2. seed words obtained through prompting ChatGPT 4o, which resulted in a 5468386 token coprus, and 3. uploading authentic legal texts, the output of which was a 158756 token corpus. A detailed description of how corpora can be compiled and terminology extracted can be found in [7]. Terminology has been extracted from all three corpora in Sketch Engine using linguistic criteria such as frequency, keyness, and term grammar. In Sketch Engine, term grammar refers to a set of CQL rules used to identify the lexical patterns that qualify as terms. The obtained MWTs from the corpora have been manually filtered to remove noise (e.g., general vocabulary). Filtering was performed by the authors, who hold a PhD degree in linguistics. This was the starting point for the identified concepts. In the next step, the manually validated source-term list was translated into Croatian using corpus tools (Sketch Engine), Neural Machine Translation (Google Translate, GT) and LLMs (ChatGPT 5.1). The English MWTs were submitted to Google Translate and ChatGPT 5.1 as isolated term units rather than as full sentences or paragraphs in order to ensure comparability across all tested approaches. As corpus tools do not generate translations like NMT/LLMs, in the compiled corpora we used (parallel) concordance to find the Croatian equivalents. We then compared how each tool aligns with IATE as the authoritative terminology reference, or the so-called gold standard, i.e., we measured the accuracy and frequency-weighted accuracy for each term according to the following formula:

$$\text{Frequency-weighted accuracy} = \sum \text{freq}(\text{correct terms}) / \sum \text{freq}(\text{all terms with IATE entry}).$$

However, alignment with IATE was used here as a benchmark for institutional terminological standardization, not as proof that non-IATE variants are necessarily incorrect, since legal terminology may show variation across texts and contexts, such as denominative variation [8, 9, 10].

We also compared the accuracy of each tool against rater agreement. For that purpose, two independent raters with legal expertise evaluated the Croatian candidates. They were asked to select the term they considered the most legally acceptable Croatian equivalent for the English source term. The raters assessed the candidates independently and were not informed which output had been produced by which tool. The evaluation focused on legal acceptability and terminological appropriateness rather than stylistic preference.

4. Results and discussion

Once we compiled the three specialized corpora and extracted the MWTs from each, we manually filtered the resulting lists to eliminate noise. Then we looked at the frequencies of MWTs, and compiled a list of the most frequent MWTs present in all three corpora. After that, we translated the MWTs using ChatGPT5.1 Plus, Google Translate (GT) and our own corpora, from which we obtained the initial list of MWTs in English and their frequencies. The results are shown in Table 1 (see Appendix), and the "gold standard" term (if present) is given in the column titled IATE, whereas the term on which both raters agreed is given in the last column.

In the next step, we calculated accuracy and frequency-weighted accuracy of the corpus, LLM, and NMT method against IATE. The results are shown in Table 2.

Table 2
Accuracy and frequency-weighted accuracy of corpus, LLM, and NMT translations of the selected MWTs when compared to IATE

Method	Exact-match accuracy (%)	Frequency-weighted accuracy (%)
Corpus	83.8	89.8
ChatGPT5.1 Plus	70.3	71.3
Google Translate (GT)	54.1	53.8

As can be seen from Table 2, when benchmarked against the IATE database as an institutional reference point, the corpus-based method achieves the highest congruence. Out of 37 English MWTs that are evaluated (the original list in Table 1 shows 43 MWTs, but only those for which there is a match in IATE have been evaluated), corpus-based translations reproduce the IATE equivalent in 84% of cases, compared with 70% for ChatGPT5.1 and 54% for GT. When corpus frequency is taken into account, the corpus translations outperform all machine translation systems by a large margin. Corpus translations achieve a frequency-weighted accuracy of 90%, meaning they correctly render nearly all high-impact, high-frequency terms. ChatGPT5.1 Plus performs slightly higher with 71%, while Google Translate has the lowest score, with 54%. Frequency-weighted accuracy was used to provide a more realistic and usage-based assessment of translation quality. In legal texts, certain terms occur more frequently and carry greater communicative and terminological weight than others. Errors in these high-frequency items are more detrimental to comprehension and professional translation practice than errors in rare terms. Weighting accuracy by corpus frequency therefore ensures that the evaluation reflects the actual distribution of terminology in authentic texts, rather than treating all items as equally important. These findings corroborate earlier studies [11] suggesting that corpus-based methods remain essential for ensuring reliability and replicability in terminology work. This outcome is not surprising, given that the corpora were compiled from EU legal and policy materials closely related to the institutional environment from which IATE also emerges. The higher congruence may therefore partly reflect genre and source proximity rather than an absolute superiority of corpus methods in all legal translation settings. Furthermore, the time used to compile the corpora and extract and translate terminology from corpora is significantly longer than the time it takes LLMs and NMT to carry out these tasks.

In the next step we asked two independent raters (lawyers) to assess which of the Croatian terms they found to be more acceptable (legally credible). Only the terms from Table 1 that had different outputs from the corpus, ChatGPT5.1 Plus and GT were given (26 terms in total). The inter-rater agreement was very high, with 81% of raw agreement. We calculated accuracy only for the terms that the raters agreed upon (19 in total). The results reveal several unexpected patterns regarding terminological stability and the performance of translation tools. First, although rater agreement was high overall (80.8%), disagreement systematically clustered around conceptually

complex terms, such as *lack of conformity* and *dark pattern*. These are precisely the areas of EU digital consumer law where Croatian usage remains unsettled, with competing variants co-existing. In contrast, raters essentially showed no disagreement for less frequent terms. The comparison between tools and rater-selected translations produced another striking result: ChatGPT5.1 Plus aligned more closely with the raters (85%) than either the corpus (70%) or GT (20%). However, when we take frequency into account, then frequency-weighted accuracy is 69% for ChatGPT5.1, 64% for the corpus, and 14% for GT. Frequency seems to play an important role in terminology and translation. Across tools and raters, low-frequency MWTs were more stable and easier to translate. High-frequency MWTs, on the other hand, were substantially more error-prone for humans and systems alike. Taken together, the findings suggest that terminological instability is strongest for highly salient, high-frequency legal concepts undergoing rapid regulatory evolution. Hence, translators and terminologists alike are required to mediate between automation and authority. Indeed, a combination of LLMs and corpora is required. LLMs are to be used to increase speed in workflows, capture more recent and emerging terms, and corpora to ensure precision, consistency, and credibility. Such integration would allow legal translators and terminologists to validate and refine AI outputs through evidence-based terminology management, thereby enhancing trust and reliability. Even where an authoritative termbase such as IATE is available, corpora remain useful because they provide evidence of actual usage, variation, and contextual distribution in authentic legal texts.

5. Limitations and future work

In this study we compared LLMs to corpus tools in the legal translation task. While corpus methods are more time-consuming, they offer verifiable data that can be used to inform and refine GenAI-output and hence lay the groundwork for deploying Terminology-Augmented Generation (TAG) [12] in legal translation. Implementing TAG and different terminology formats in the legal translation remains to be examined in future work. Possible limitations of this study concern the fact that a relatively small dataset was analyzed and the context was not accounted for. For this reason, any conclusions about the relative advantage of LLMs for recent or emerging legal terminology should be treated as preliminary. In future research it will also be examined if providing the models with more context (paragraph-based translation) affects the accuracy scores [13]. The study contributes to ongoing discussions on the credibility and authority of AI-generated terminology in legal translation. Overall, the results point to a future in which LLMs and other terminology resources mutually reinforce one another, but where evaluation frameworks must be explicit about whether they reward prescriptive terminological standards or usage-based expert judgments.

Acknowledgements

This paper has been written within the scope of the project „LawTermAI: Legal Terminology Management in the Digital Age" (uniri-iz-25-63), funded by European Union – NextGenerationEU via the Croatian National Recovery and Resilience Plan 2021-2026, in conjunction with the University of Rijeka, Faculty of Law Programme Financing. The views and opinions expressed are solely those of the authors and do not necessarily reflect the official position of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Declaration on Generative AI

During the preparation of this work, the authors used Chat-GPT-5 in order to check grammar and spelling. The authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. Busse, Recht als Text: Linguistische Untersuchungen zur Arbeit mit Sprache in einer gesellschaftlichen Institution, volume 131 of Reihe Germanistische Linguistik, Max Niemeyer Verlag, Berlin, New York, Tübingen, 1992. doi:10.1515/9783111377063.

[2] M. Bajčić, Terminological variation in EU legal texts: Dynamicity vs. consistency of legal language, Fluminensia 36 (2024) 375–409. doi:10.31820/f.36.2.3.

[3] E. Wiesmann, Machine translation in the field of law: A study of the translation of Italian legal texts into German, Comparative Legilinguistics 37 (2019) 117–153. doi:10.14746/cl.2019.37.4.

[4] J. Killman, Machine translation and legal terminology: Data-driven approaches to contextual accuracy, in: Ł. Biel, H. J. Kockaert (Eds.), Handbook of Terminology. Volume 3: Legal Terminology, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2023, pp. 485–510. doi:10.1075/hot.3.mac2.

[5] P. D. Callister, Generative AI and finding the law, Law Library Journal 117 (2025). doi:10.2139/ssrn.4608268.

[6] A. Kilgarriff, V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, V. Suchomel, The Sketch Engine: Ten years on, Lexicography ASIALEX 1 (2014) 7–36. doi:10.1007/s40607-014-0009-9.

[7] M. Borucinsky, Primjena metoda korpusne lingvistike u jezikoslovnim istraživanjima. Pomorski fakultet Sveučilište u Rijeci, Rijeka, 2023.

[8] M. Bajčić, Terminological variation and conceptual divergence in EU law, in: Ł. Biel, H. J. Kockaert (Eds.), Handbook of Terminology. Volume 3: Legal Terminology, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2023, pp. 271–288. doi:10.1075/hot.3.ter11.

[9] Ł. Biel, Variation of legal terms in monolingual and multilingual contexts: Types, distribution, attitudes and causes, in: Ł. Biel, H. J. Kockaert (Eds.) Handbook of Terminology. Volume 3: Legal Terminology, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2023, pp. 90–123. doi:10.1075/hot.3.var1.

[10] J. Freixa, S. Fernández-Silva, Terminological variation and the unsaturability of concepts, in: P. Drouin, A. Francœur, J. Humbley, A. Picton (Eds.), Multiple Perspectives on Terminological Variation, volume 18 of Terminology and Lexicography Research and Practice, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2017, pp. 155–180. doi:10.1075/tlrp.18.07fre.

[11] M. Cabezas-García, P. León-Araúz, Machine versus corpus-based translation of multiword terms, Digital Scholarship in the Humanities 38 (2023) i6–i16. doi:10.1093/llc/fqad026.

[12] G. M. Di Nunzio, Terminology-augmented generation (TAG): Foundations, use cases, and evaluation paths, Journal of Digital Terminology and Lexicography 1 (2025) 97-104. doi:10.25430/pupj.jdtl.1752566034.

[13] A. Lackner, A. Vega-Wilson, C. Lang, An evaluation of terminology-augmented generation (TAG) and various terminology formats for the translation use case, Journal of Digital Terminology and Lexicography 1 (2025) 31–47. doi:10.25430/pupj.jdtl.1763379938.

A. Appendix

Table 1
MWTs and their Croatian equivalents

Multi-term - English	Relative frequency in corpora	Corpus translation	Google Translate	ChatGPT 5.1 Translation (28/11/2025)	IATE	Rater
dark pattern	4164.59	tamni obrazac	tamni uzorak	obmanjujući obrazac / mračni	tamni obrazac	/

				obrazac		
digital service	3439.16	digitalna usluga	digitalna usluga	digitalna usluga	digitalna usluga	/
member state	3347.98	država članica	članica države	država članica	država članica	država članica
business user	3169.57	poslovni korisnik	poslovni korisnik	poslovni korisnik	poslovni korisnik	/
consumer protection	2666.97	zaštita potrošača	zaštita potrošača	zaštita potrošača	zaštita potrošača	/
platform service	2467.7	usluga platforme	platforma usluga	usluga platforme	/	platformska usluga
core platform service	2396.11	osnovna usluga platforme	glavna platforma usluga	osnovna usluga platforme	osnovna usluga platforme	temeljna usluga platforme
contract law	1821.04	ugovorno pravo	ugovorno pravo	ugovorno pravo	ugovorno pravo	/
intermediation service	1747.2	usluga posredovanja	posrednička usluga	posrednička usluga	usluga posredovanja	usluga posredovanja
online intermediation service	1712.37	usluga internetskog posredovanja	online posrednička usluga	usluga internetskog posredovanja	usluga internetskog posredovanja	usluga internetskog posredovanja
digital single market	1699.7	jedinstveno digitalno tržište	digitalno jedinstveno tržište	jedinstveno digitalno tržište	digitalno jedinstveno tržište	jedinstveno digitalno tržište
internal market	1583.23	unutarnje tržište	interno tržište	unutarnje tržište	unutarnje tržište	unutarnje tržište
end user	1549.12	krajnji korisnik	krajnji korisnik	krajnji korisnik	krajnji korisnik	/
personal datum	1472.61	osobni podaci	osobni podaci	osobni podatak	osobni podaci	/
lack of conformity	1219.87	neusklađenost	nedostatak usklađenosti	nedostatak sukladnosti / neusklađenost	/	/
online platform	1178.95	online platforma	internetska platforma	internetska platforma	internetska platforma	internetska platforma
single market	1058.48	jedinstveno tržište	jedinstveno tržište	jedinstveno tržište	jedinstveno tržište	/
national law	1058.47	nacionalno pravo	nacionalno pravo	nacionalno pravo	nacionalno pravo	/
online sale	1005.78	online prodaja	internetska	internetska	online	internetska

			prodaja	prodaja	prodaja	prodaja
supply of digital content	770.13	isporuka digitalnog sadržaja	opskrba digitalnim sadržajem	isporuka digitalnog sadržaja	isporuka digitalnog sadržaja	isporuka digitalnog sadržaja
third party	750.68	treća strana	treća strana	treća strana	treća strana	/
data protection	731.31	zaštita podataka	zaštita podataka	zaštita podataka	zaštita podataka	/
sales contract	690.73	ugovor o prodaji	prodajni ugovor	ugovor o prodaji	/	/
union law	607.81	pravo unije	union zakon	pravo Unije	pravo Unije	pravo Unije
service provider	547.71	pružatelj usluga	davatelj usluga	pružatelj usluga	pružetalj usluga/davatelj usluga	pružatelj usluga
consumer law	539.25	pravo o zaštiti potrošača	potrošačko pravo	potrošačko pravo	potrošačko ugovorno pravo	potrošačko pravo
delivery service	453.52	usluga dostave	usluga dostave	dostavna usluga	usluga dostave	/
parcel delivery service	434.69	usluga dostave paketa	usluga dostave paketa	usluga dostave paketa	usluga dostave paketa	/
consumer right	394.14	pravo potrošača	potrošačko pravo	pravo potrošača	prava potrošača	pravo potrošača
digital markets act	390.3	akt o digitalnim tržištima	akt o digitalnim tržištima	Akt o digitalnim tržištima	akt o digitalnim tržištima	/
sale of goods	372.87	prodaja robe	prodaja robe	prodaja robe	kupoprodaja robe	/
digital sector	366.45	digitalni sektor	digitalni sektor	digitalni sektor	digitalni sektor	/
commercial practice	363.99	poslovna praksa	komercijalna praksa	poslovna praksa	poslovna praksa	/
digital single market strategy	346.73	strategija jedinstvenog digitalnog tržišta	strategija za digitalno jedinstveno tržište	strategija jedinstvenog digitalnog tržišta	strategija jedinstvenog digitalnog tržišta	strategija jedinstvenog digitalnog tržišta
digital economy	319.35	digitalno gospodarstvo	digitalna ekonomija	digitalno gospodarstvo	digitalna ekonomija. digitalno gospodarstvo	digitalno gospodarstvo
right of withdrawal	278.35	pravo na odustajanje	pravo na odustajanje	pravo na povlačenje iz ugovora	pravo odustajanja	/

consumer contract	274.55	potrošački ugovor	potrošački ugovor	potrošački ugovor	/	/
illegal content	257.17	nezakonit sadržaj	nezakonit sadržaj	nezakonit sadržaj	nezakonit sadržaj	nezakonit sadržaj
free movement	235.31	slobodno kretanje	slobodno kretanje	slobodno kretanje	slobodno kretanje	/
principle of subsidiary	228.11	načelo supsidijarnosti	princip politike supsidijarnosti	načelo supsidijarnosti	/	načelo supsidijarnosti
unfair practice	194.24	nepoštena praksa	nepravedna praksa	nepoštena praksa	nepoštena praksa	nepoštena praksa
country of origin principle	182.49	/	princip načela zemlje podrijetla	načelo zemlje podrijetla	/	načelo zemlje podrijetla