transLogos 2026 Vol 9 Issue 1
Vraila, Stavroula Paraskevi, pp. 23–46
https://dx.doi.org/10.29228/transLogos.86

transLogos
A Translation Studies Journal

Stylistic Flattening and Terminological Displacement in AI-Mediated Legal Translation: Evidence from German–Greek and English–Greek Language Pairs

Stavroula Paraskevi VRAILA*

Large language models (LLMs) are increasingly being integrated into translation workflows. However, their systemic impact on the quality of legal translation in non-hegemonic language pairs remains insufficiently explored. This article examines two interrelated phenomena, namely stylistic flattening and terminological displacement via English mediation, as observed in AI-mediated legal translation into Greek. Drawing on a purpose-built corpus of controlled outputs from two neural machine translation engines (DeepL and Google Translate) and two generative LLMs (OpenAI GPT-5.3 and Anthropic Claude Opus 4.5), tested on German–Greek and English–Greek legal terms, the study investigates how these systems, whose training data are heavily weighted towards Anglophone legal corpora, reduce register variation and impose Anglo-American categories onto continental legal traditions, particularly for concepts lacking direct equivalence across legal systems. The analysis reveals that stylistic leveling operates at the level of register and broader discourse conventions, while terminological displacement functions at the conceptual level, substituting source-system categories with common law analogues. Together, these mechanisms constitute what the article terms algorithmic legal hegemony, a structurally embedded asymmetry through which Anglophone legal reasoning colonizes target systems via ostensibly neutral technological mediation. The article concludes by discussing implications for translator training, post-editing literacy, and the critical evaluation of AI-generated outputs in specialized contexts.
Keywords: GenAI; specialized translation; legal translation; English mediation; translator training

1. Introduction

The integration of large language models (LLMs) into professional translation workflows has accelerated rapidly since 2022, reshaping not only expectations of productivity but also the epistemic conditions under which translation is produced and evaluated. In legal translation, a field defined by terminology closely tied to specific legal systems and jurisdictions, as well as by stylistic conventions rooted in linguistic and cultural norms, the use

*Assistant professor at Ionian University, Corfu.
E-mail: stavivra@gmail.com; ORCID ID: https://orcid.org/0009-0000-5404-3952.
(Received 20 April 2026; accepted 18 June 2026)

of artificial intelligence (AI) raises concerns that extend well beyond surface-level accuracy. When an LLM trained primarily on English-language data translates a German legal term into Greek, it does not simply select an equivalent expression. Rather, it filters the source concept through the conceptual framework of Anglo-American legal categories, producing outputs that may be fluent but are conceptually and institutionally shifted.

This article examines two interrelated phenomena that consistently emerge in LLM-mediated legal translation into Greek: stylistic flattening and terminological displacement through English mediation. Stylistic flattening refers to the systematic reduction of register variation, the neutralization of formal legal conventions, and the erosion of genre-specific linguistic features in LLM outputs. Terminological displacement denotes the replacement of concepts from continental law, rooted in the European legal tradition shared by both German and Greek legal systems, with approximations drawn from common law, reflecting the Anglophone bias embedded in model training.

The present study is motivated by the decision of the Greek government to introduce AI into the translation of legal texts. It also responds to a significant gap in existing literature. While research on the quality of machine translation in legal contexts has grown substantially (Prieto Ramos 2021; Killman 2023; Rossi and Chevrot 2019), including work on the risks of automated translation in high-stakes institutional settings such as courts (Vieira, O'Hagan, and O'Sullivan 2021) and recent empirical evaluations of generative models in the legal domain (Briva-Iglesias, Cavalheiro Camargo, and Dogru 2024), the majority of empirical studies focus on language pairs involving English, particularly English–French, English–Spanish, and English–Chinese. Non-hegemonic language pairs, that is, those that do not include English as either source or target, remain markedly underexplored, even though it is precisely in such pairs that meaningful shifts are most likely to occur as a result of English-centric training data.

Greek, as a language with a fully developed legal terminology grounded in the continental European legal tradition and, more specifically, in Roman law, but with relatively limited digital representation in LLM training corpora, offers a particularly suitable testing ground for examining this dynamic. The choice of the German–Greek (DE→EL) language pair is motivated by the structural affinity between these two legal systems: both belong to the Roman-Germanic tradition, and the Greek Civil Code was modeled substantially on the German

Bürgerliches Gesetzbuch (Civil Code). This means that English mediation in DE→EL translation is theoretically unexpected, making it a particularly revealing test case.

The study draws on a purpose-built corpus of 50 system-bound German legal terms and their English common law equivalents, translated into Greek by four systems (DeepL, Google Translate, OpenAI GPT-5.3, and Anthropic Claude Opus 4.5) in both the German–Greek (DE→EL) and English–Greek (EN→EL) directions. The dual-direction design is methodologically central: DE→EL provides the primary test of English mediation, while EN→EL serves as a control so that the contrast between the two directions distinguishes English as an actual source language from English as an invisible pivot.

Three research questions guide the analysis. First, to what extent do system outputs for non-hegemonic language pairs (DE→EL) exhibit English mediation, and does this differ across different system architectures (NMT vs. LLM)? Second, how do stylistic flattening and terminological displacement interact across legal domains, and are certain categories of system-bound terminology more vulnerable to displacement? Third, what are the implications of these findings for translator training and the critical evaluation of LLM outputs in specialized translation?

2. Theoretical Framework

2.1 Stylistic Flattening as a System-Level Phenomenon

The systematic reduction of register variation in machine-generated legal translation, which this article terms ‘stylistic flattening,’ is conceptualized here not as an aggregate of individual errors but as a structural property of how contemporary translation systems are built and aligned. The phenomenon is therefore theorized at the level of system design rather than that of the translating subject: register loss follows from the statistical character of the models and their optimization objectives and is systemic rather than incidental.

The framework proposed here draws one explanatory factor from each of three research traditions: from corpus linguistics, the distributional composition of training data; from the study of machine-learning alignment, the objectives that shape generative models; and from comparative legal linguistics, the register conventions specific to legal discourse. Each of these factors is rendered analytically tractable in the study’s coding scheme, and from each the framework derives a specific, testable prediction about the form that register loss should take.

The first factor is 'corpus underrepresentation.' Neural machine translation (NMT) engines and generative LLMs alike estimate the probability distributions of their training data, in which the formal register of legal texts, characterized by nominalization, fixed collocations, and formulaic Latinate phrasing, is statistically rare relative to the general-purpose text that dominates the web-scale corpora on which such systems are trained (Bommasani et al. 2021; Dodge et al. 2021). Research on machine translation has documented the consequence directly: domain mismatch and limited training data are among the principal sources of degradation in neural systems (Koehn and Knowles 2017), and because domain-specific corpora are typically scarce, general-purpose models perform poorly on specialized text without adaptation (Chu and Wang 2018).

A model under-exposed to high-register legal discourse has, correspondingly few examples from which to reproduce its conventions. The mechanism is a frequency effect: where the signal for register-appropriate formulation is weak, the model's prior pulls outputs toward higher-frequency, lower-register ordinary language. The prediction is specific. The expected deformation is a downward leveling toward a neutral middle rather than colloquialization or hyper-formalization, both of which would require a positive signal that data sparsity withholds; neutralization should therefore dominate the register shifts, a distinction the study's coding typology is designed to capture.

The second factor is 'alignment.' Generative LLMs are not trained merely to reproduce parallel text but are fine-tuned to satisfy human or model-generated preferences, and the objective chosen at this stage independently shapes register. Reinforcement learning from human feedback (RLHF) optimizes for fluency, helpfulness, and perceived naturalness because its reward model is trained on human judgments that favor readable text (Ouyang et al. 2022). Such reward models have been shown to favor surface features such as response length well beyond their contribution to substantive quality (Singhal et al. 2023).

In translation, this introduces a smoothing pressure that competes with terminological and register fidelity, rewarding the fluent paraphrase over the formally exact rendering. Constitutional AI, by contrast, supervises behavior against explicitly stated principles (Bai et al. 2022) that can encode faithfulness and precision as values that attenuate that pressure. Alignment is thus a second lever on flattening, independent of corpus composition: models with comparable pretraining may flatten differently depending on their alignment objective. The

prediction is intra-paradigm variation. If flattening were a uniform property of generative architectures, all instruction-tuned models would flatten to the same degree. A systematic gap between systems aligned under different regimes would instead show flattening to be, in part, an alignment artefact and therefore contingent rather than technologically inevitable.

The third factor is the 'register specificity of legal discourse' itself. The formal markers of legal texts are functional rather than ornamental: they signal the authority, performativity, and systemic position of the text, and in performative language they carry illocutionary force (Šarčević 1997; Mattila 2013). A rendering that preserves propositional content while stripping these markers may remain intelligible yet fail the communicative norms of the target legal system, a failure of what Łucja Biel terms the "textual fit" (2014, 104) of translated law and what Brian Mossop discusses as a failure to revise the text's "tailoring" (2014, 149) with regard to register, except that here no human reviser has intervened. Flattening must, therefore, be treated as a distinct error category rather than a sub-case of terminological displacement: a translation can be conceptually correct yet register-inappropriate, since register operates on a separate axis from conceptual equivalence. The prediction follows that register erosion is the more pervasive deformation, since it can occur even where the terminological choice is accurate. So construed, stylistic flattening is not a borrowed literary category but an operational variable, defined as the loss of legal register, explained by data composition and alignment, and measured against target-system discourse conventions.

2.2 Legal Translation and System-Bound Terminology

Susan Šarčević's foundational work on legal translation established that legal terms are "system-bound" (1997, 9): their meaning is constituted by the legal system in which they operate, and they cannot be transferred across systems without remainder. This insight is central to the present study. The German concept of *Vorsatz* (dolus directus and dolus eventualis) does not map neatly onto the English *intent* or *mens rea*, which carry common law connotations of purpose and knowledge derived from an entirely different doctrinal architecture. When an LLM translates *Vorsatz* into Greek as *πρόθεση* (*prothesi*: intention) rather than *δόλος* (*dolos*: dolus), it is not simply selecting a synonym; it is substituting one legal conceptual framework for another.

This phenomenon is compounded by what Enrique Alcaraz Varó (2009) terms 'anisomorphism' between legal systems: the structural asymmetry whereby legal concepts in one system have no counterpart in another. The German three-tier classification of criminal offenses (*Verbrechen*, *Vergehen*, *Ordnungswidrigkeit*) corresponds to the Greek tripartite distinction (*κακούργημα* [*kakourgima*], *πλημμέλημα* [*plimmelima*], *πταίσμα* [*ptaisma*]) but has no equivalent in the common law's binary *felony/misdemeanor distinction*. When LLMs produce translations for this classification, they consistently impose the English binary framework, collapsing the tripartite structure into generic terms (*έγκλημα* [*egklima*], *αδίκημα* [*adikima*]) that erase the doctrinal specificity of both the German source and the Greek target.

Heikki E. S. Mattila's (2013) comparative legal linguistics provides further analytical tools, particularly the distinction between autonomous and heteronomous legal terminology. Autonomous terms are those whose meaning is defined exclusively within the legal system, while heteronomous terms derive their meaning from ordinary language and are merely adopted into legal usage. LLM errors cluster disproportionately around autonomous terms, precisely because these are the terms for which the training data's English bias produces the most distorted outputs.

2.3 Algorithmic Legal Hegemony

The concept of 'algorithmic legal hegemony' as proposed here draws on Robert Phillipson's (1992) linguistic imperialism thesis and Norman Fairclough's (1989, 2003) critical discourse analysis, which examines how power relations are reproduced through discursive practices that appear neutral but serve ideological functions.

Algorithmic legal hegemony operates through a specific mechanism: LLMs trained predominantly on English-language corpora internalize the conceptual architecture of the common law as a default framework for legal reasoning. When these systems produce translations into languages embedded in different legal traditions, they do not simply transfer lexical items; they impose the ontological categories of the dominant legal system onto the target, producing outputs that are fluent, plausible, and systematically misleading. This hegemony is algorithmic because it operates through the statistical regularities of training data rather than through conscious ideological intent; it is legal because it specifically affects the

conceptual integrity of legal systems; and it is hegemonic in the Gramscian sense because it presents itself as technically optimal rather than as a particular cultural or legal perspective.

Crucially, algorithmic legal hegemony is not an inevitable consequence of LLM technology but a consequence of specific design choices, namely training-data composition, fine-tuning objectives, and alignment strategies. Because different architectures exhibit substantially different rates of English mediation, the hegemonic tendency is architecturally contingent rather than technologically necessary, which both guards the analysis against technological determinism and implies that mitigation is possible.

3. Methodology

3.1 Corpus Design

The empirical corpus comprises 50 German legal terms selected according to three criteria: (a) systemic incongruence between German, Greek, and English legal systems; (b) terminological autonomy in the sense of Mattila (2013), ensuring that terms are system-bound rather than borrowed from ordinary language; and (c) distributional coverage across four legal domains: criminal law (15 terms), civil law (15 terms), procedural law (10 terms), and public/constitutional law (10 terms). This thematic distribution reflects both the structure of the Greek legal system and the areas where system-bound terminology is most concentrated.

For each term, the corresponding English common law equivalent or near-equivalent was identified on the basis of comparative legal analysis (e.g., *Vorsatz* / *intent*, *Bewährung* / *probation*, and *Rechtsgeschäft* / *legal transaction*). These English equivalents served as source terms for the EN→EL translation direction, enabling a controlled comparison between direct DE→EL translation and English-mediated translation of the same legal concepts.

The reference translation in Greek for each term was established through triangulation of four sources: (a) standard bilingual legal dictionaries (Zacharopoulos et al. 2013), (b) the IATE terminology database,[1] (c) official Greek legislative texts containing the equivalent concept, and (d) authoritative Greek legal scholarship. The reference translations represent the terminologically correct rendering as recognized within the Greek legal system, rather than just a linguistically acceptable paraphrase. Each term was also classified according to its system

[1] *IATE (Interactive Terminology for Europe)*, accessed March 13, 2026, https://iate.europa.eu/.

incongruence type (pseudo-equivalent, partial correspondence, anisomorphism, common law gap, polysemy, broader EN term, Latin-origin term, or full correspondence) through comparative legal analysis.

The selection of 50 terms represents a deliberate methodological choice. The corpus is large enough to permit meaningful quantitative analysis across legal domains and system types, yet small enough to allow detailed qualitative analysis of individual translation decisions. Each term was selected because it poses a specific theoretical challenge to LLM-mediated translation: either it has no common law equivalent, its English translation is a pseudo-equivalent that distorts the source concept, or it belongs to a semantic field where the common law draws different conceptual boundaries.

3.2 Model Selection and Translation Protocol

Four models were selected to represent two distinct architectural paradigms. The NMT category includes DeepL and Google Translate, both dedicated translation engines optimized for fluency and trained on large parallel corpora. The generative LLM category includes GPT-5.3 and Claude Opus 4.5, both general-purpose language models capable of translation, among other tasks, and trained with instruction tuning and alignment procedures (RLHF and Constitutional AI, respectively).

This typological distinction is methodologically significant. NMT systems, trained on parallel text through encoder-decoder architectures, are directly dependent on their training distribution; where data for a pair are limited, as for DE→EL, they are more likely to route translation through an English pivot. Generative LLMs can instead reason about translation choices and ambiguity. Including both types lets the study distinguish structural English mediation, inherent to NMT data, from architecturally contingent mediation that reasoning may mitigate.

Each term was translated in two directions: DE→EL and EN→EL, yielding 400 individual outputs. The translation prompt was kept minimal to test default LLM behavior. The DE→EL direction constitutes the primary test of English mediation: if an LLM translating directly from German to Greek produces outputs that reflect common law categories rather than the shared continental legal framework, this provides direct evidence of English mediation in

the translation process. The EN→EL direction serves as a control, revealing whether the same LLMs perform differently when English is the actual source language.

The study is cross-sectional: all four systems were sampled at a single point in time, with outputs collected between February and March 2026. Because these systems, NMT engines especially, are updated continuously, the specific renderings reported here may not reproduce identically later; the archived corpus is the empirical record, and the claims below concern its aggregate patterns rather than any individual output.

3.3 Coding Framework

Each system output was independently coded along five dimensions: (a) English mediation (Y/N): whether the output reflects common law categorization rather than continental European legal reasoning; (b) stylistic flattening (Y/N): whether the output exhibits loss of register, formal legal discourse conventions, or terminological precision; (c) register shift type (Neutralization, Colloquialization, Hyper-formalization, or None); (d) terminological displacement (Y/N): whether a source-legal-system concept has been substituted with a common law analogue; and (e) displacement type (Conceptual substitution, Calque from EN, Hybrid, or None).

The coding was performed by the author and independently validated by two practicing legal translators specializing in DE→EL and EN→EL language pairs. Discrepancies were resolved through discussion until consensus was reached. A detailed coding guide with definitions, examples, and decision criteria was developed to ensure consistency across the 400 outputs. Summary statistics were computed automatically using formula-driven spreadsheets with per-system granularity, enabling cross-tabulation by system type, translation direction, legal domain, and displacement type.

3.4 Methodological Considerations: Architecture, Alignment, and the Limits of Comparability

A methodological caveat is in order regarding the comparability of NMT and generative LLM outputs. Instruction-tuned models such as GPT-5.3 and Claude Opus 4.5 are subject to alignment procedures (RLHF, Constitutional AI) that imbue them with metacognitive capabilities absent from NMT systems. When a generative LLM translates a legal term, it does not merely produce the statistically most probable target; it activates mechanisms of accuracy

assessment, disambiguation, and contextual reasoning that are architecturally embedded in its alignment. The superior performance of Claude Opus 4.5 observed in this study cannot, therefore, be unambiguously attributed to domain-specific training data, general reasoning capabilities, or the alignment objective, as the exact composition of proprietary LLM training corpora remains undisclosed (Bender et al. 2021).

However, two indirect indicators suggest a mixed explanation. First, Claude Opus 4.5's slightly higher mediation in EN→EL (12%) than DE→EL (10%) indicates that English source terms activate different processing pathways even in otherwise accurate systems. Second, Claude's systematic failures on system-bound terms lacking common law equivalents suggest that its accuracy has limits precisely where domain-specific knowledge, rather than reasoning, is required. Together these indicators are consistent with the hypothesis that Claude's performance reflects a combination of general reasoning capabilities and some degree of exposure to legal texts, but not specialized training in Greek law.

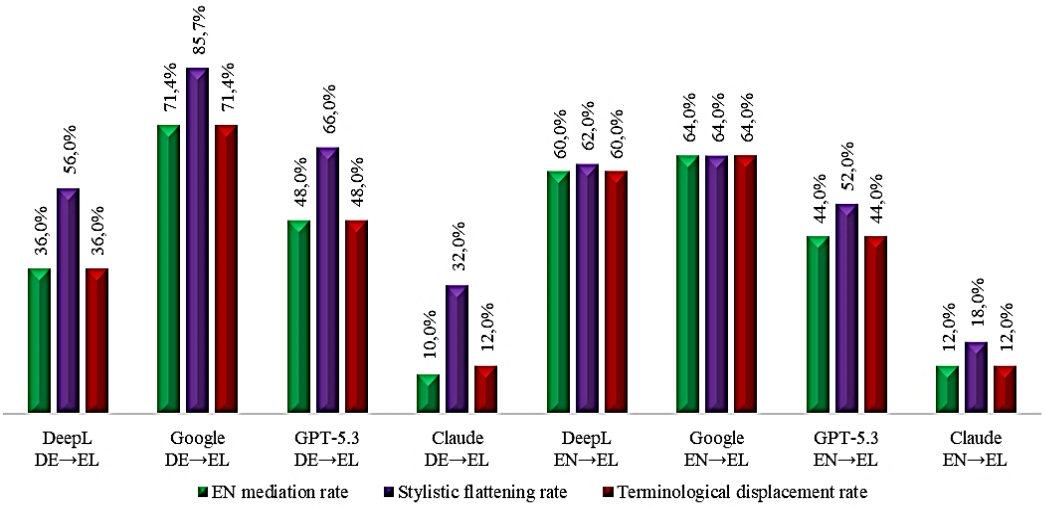
This caveat does not invalidate the comparative findings but qualifies their interpretation: the contrast is between two fundamentally different approaches, pattern matching and reasoning-augmented pattern matching, and the large differential in mediation rates between them is itself an important finding.

4. Findings

4.1 Overall Patterns

The results reveal a clear gradient across system types (fig. 1). Google Translate exhibits the highest rates of English mediation (71.4% DE→EL, 64.0% EN→EL), followed by GPT-5.3 (48.0% / 44.0%), DeepL (36.0% / 60.0%), and Claude Opus 4.5 (10.0% / 12.0%). This gradient is consistent across all three coding dimensions, confirming that English mediation, stylistic flattening, and terminological displacement are correlated phenomena rather than independent variables.

Figure 1. Aggregate coding results across systems and translation directions



Two structural observations emerge from the aggregate data. First, the NMT systems (DeepL, Google Translate) exhibit substantially higher rates of English mediation than the generative LLMs (GPT-5.3, Claude Opus 4.5), suggesting that the architectural difference between pattern-matching and reasoning-capable systems has measurable consequences for translation quality in specialized domains. Second, stylistic flattening rates consistently exceed terminological displacement, with one equal-rate exception, indicating that register erosion is a more pervasive phenomenon than conceptual substitution. LLMs may produce the correct term but strip it of its register-appropriate context.

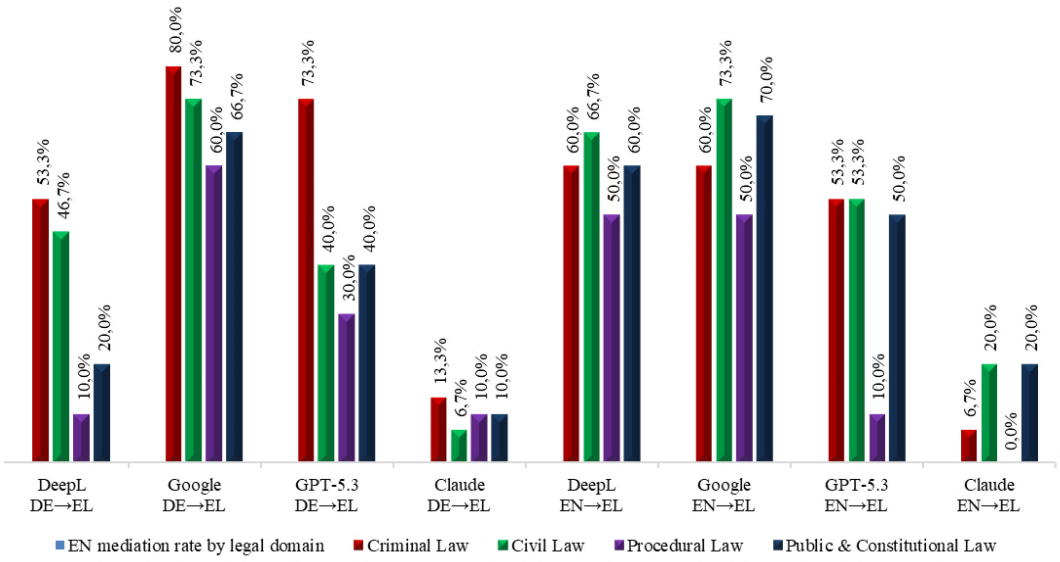
A notable asymmetry emerges in DeepL's performance: it exhibits markedly lower English mediation in the DE→EL direction (36.0%) than in EN→EL (60.0%). This pattern suggests that DeepL maintains a functional DE→EL translation pathway that is partially insulated from English mediation but defaults to English-mediated strategies when English is the actual source language. Google Translate, by contrast, shows consistently high mediation rates in both directions (71.4% and 64.0%), suggesting a more thoroughgoing dependence on English as a pivot language. This difference may reflect DeepL's investment in direct language pair models for European languages, whereas Google Translate's architecture relies more heavily on English-pivoted multilingual models.

transLogos
A Translation Studies Journal

4.2 Distribution by Legal Domain

The domain-level analysis reveals significant variation (fig. 2). Criminal law exhibits the highest overall rates of English mediation, reflecting the profound structural differences between the continental tripartite classification (*κακούργημα* [*kakourgima*], *πλημμέλημα* [*plimmelima*], *πταίσμα* [*ptaisma*]) and the common law binary distinction (*felony*, *misdemeanor*). Terms such as *Vorsatz*, *Eventualvorsatz*, and *Bewährung*, whose correct Greek equivalents are *δόλος* (*dolos*), *ενδεχόμενος δόλος* (*endechomenos dolos*) and *αναστολή ποινής* (*anastoli poinis*), respectively, are consistently displaced by their common law functional analogues (*intent*, *recklessness*, *probation*). The criminal law domain also features the most extreme cases of conceptual substitution, where the LLM output goes beyond imprecision to refer to a fundamentally different legal institution.

Figure 2. English mediation rate by legal domain and system



Procedural law shows the lowest rates of English mediation, likely because many procedural concepts have well-established Greek equivalents that appear frequently in LLM training data. The term *Berufung* (appeal) is correctly rendered by all LLMs in both directions, as is *Beweislast* (burden of proof). These are cases where the conceptual overlap between legal systems is substantial and the Greek term is sufficiently represented in digital corpora to override any English-mediation tendency. However, even within procedural law, system-bound

terms like *Revision* (a specific appellate mechanism absent from common law) and *Streitverkündung* (a procedural institution specific to continental systems) are displaced by calques.

Civil law presents a particularly instructive pattern. Core property-law concepts (*νομή* [*nomi*], *κυριότητα* [*kyriotita*]) are systematically displaced by property law calques from English. The term *νομή* (*nomi*: *possessio* in the Romanistic tradition) is rendered as *κατοχή* (*katochi*: detention) by all systems except Claude Opus 4.5, collapsing the conceptual distinction between factual control and legal possession that is fundamental to continental property law. Similarly, *κυριότητα* (*kyriotita*: *dominium* in the Romanistic tradition) is consistently rendered as *ιδιοκτησία* (*idioktisia*: property), a term that in Greek covers a broader semantic field, including both ownership and property as an economic category. These are not marginal errors; they affect the foundational concepts of an entire legal domain.

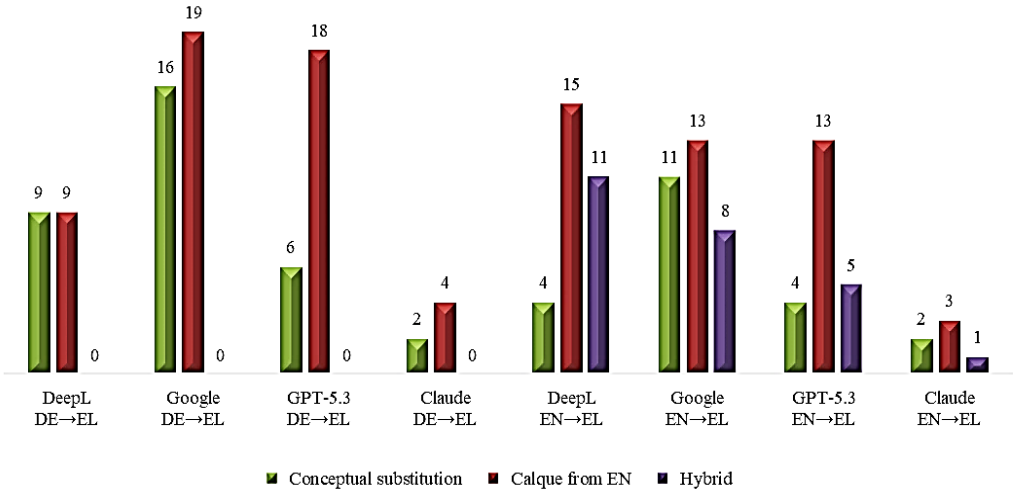
4.3 Displacement Types and Register Shifts

Three displacement types were identified. Conceptual substitution involves the wholesale replacement of a civil law concept with a common law analogue: *Vorsatz* → *πρόθεση* (*prothesi*: intent) rather than *δόλος* (*dolos*: dolus); *Bewährung* → *δοκιμαστική περίοδος* (*dokimastiki periodos*: probation) rather than *αναστολή ποινής* (*anastoli poinis*: suspension of sentence). Conceptual substitution is the most serious displacement type because it replaces the entire conceptual framework, not merely the lexical form. Calque from English produces neologistic constructions that mirror English syntax: *Leistungsstörung* → *αθέτηση σύμβασης* (*athetisi symvasis*: breach of contract), which represents only one subtype of the broader German concept, or *Vollstreckungsbescheid* → *διαταγή εκτέλεσης* (*diatagi ektelesis*: enforcement order), which conflates a specific German procedural instrument with a broader English category. Hybrid displacement combines elements of both, typically producing outputs that pair the correct term with an English-mediated alternative: *πρόθεση/δόλος* (*prothesi*: intent; *dolos*: dolus), *ιδιοκτησία/περιουσία* (*idioktisia*: property; *periousia*: estate).

A striking finding is that hybrid displacement appears exclusively in the EN→EL direction (fig. 3): when English is the explicit source language, all four systems occasionally pair a direct calque with the correct Greek term rather than committing to a single rendering. Hybrid displacement is therefore a property of the translation direction rather than of model

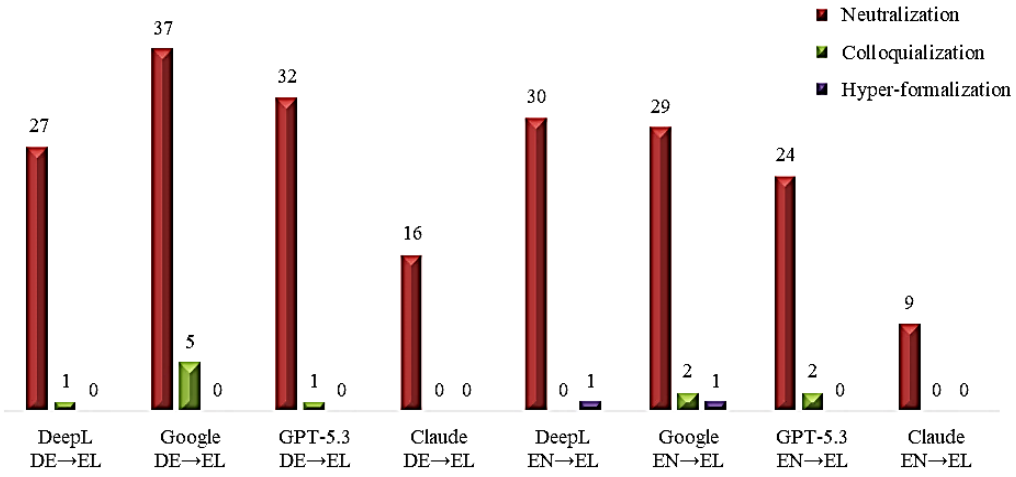
architecture. It is, moreover, more frequent in the NMT systems (DeepL and Google Translate) than in the generative LLMs, with Claude Opus 4.5 producing the fewest such outputs of any system. The co-occurrence of an Anglophone calque and its Greek equivalent thus cannot be read as evidence of superior metacognitive awareness in the generative models; rather, the presence of an English source term appears to make both candidate renderings salient simultaneously, an effect most pronounced in the NMT engines.

Figure 3. Distribution of displacement types across systems



Register shift analysis reveals that neutralization is by far the dominant shift type (fig. 4). Colloquialization (the lowering of register below that of the source) is rare and concentrated in Google Translate outputs, where terms like *Notstand* (*κατάσταση έκτακτης ανάγκης* [*katastasi ektaktis anagkis*]: state of emergency) are rendered simply as *ανάγκη* (*anagki*: need), a marked register collapse that transforms a technical legal concept into an everyday word. Hyper-formalization (the imposition of a register higher than that of the source) is virtually absent. The systems thus do not so much distort register as erase it, producing a 'middle register' that is neither formally appropriate for legal texts nor colloquially natural.

Figure 4. Distribution of register shift across systems



4.4 NMT vs. Generative LLM: A Structural Comparison

The most significant finding of the study is the marked performance differential between NMT systems and generative LLMs. Averaging across both directions, the NMT systems (DeepL and Google Translate) exhibit a mean English mediation rate of 57.9%, while the generative LLMs (GPT-5.3 and Claude Opus 4.5) average 28.5% (fig. 1). Within the generative category, Claude's performance (11.0% mean mediation rate) is substantially superior to GPT-5.3's (46.0%), suggesting that alignment strategies and training approaches within the generative paradigm also produce significant variation.

This differential is most pronounced for autonomous legal terminology with no common law equivalent. *Rechtsgeschäft* illustrates it directly: Google Translate returns *δικαιοδοσία* (*dikaiodosia*: jurisdiction) rather than *δικαιοπραξία* (*dikaiopraxia*: legal act), replacing a private-law act with a concept of procedural authority, whereas Claude Opus 4.5 produces the correct term. Claude is likewise the only system to produce *νομή* (*nomi*: possessio) rather than *κατοχή* (*katochi*: detention) for *Besitz*, *κυριότητα* (*kyriotita*: ownership) rather than *ιδιοκτησία* (*idioktisia*: property) for *Eigentum*, and *έγκληση* (*egklisi*: complaint) rather than *ποινική καταγγελία* (*poiniki katangelia*: criminal report) for *Strafantrag*, all terms that require knowledge of the Greek legal system rather than pattern matching from English.

Conversely, system-bound terms that lack equivalents in any tradition, such as the German *Grundschuld*, rendered in Greek as *έγγεια οφειλή* (*engeia ofeili*: land debt), and *Strafbefehl*, rendered as *ποινική απόφαση* (*poiniki apofasi*: penal judgment), defeat all systems,

including Claude Opus 4.5, confirming that even advanced reasoning capabilities cannot substitute for domain-specific knowledge. This finding delineates the boundary of what generative LLMs can achieve in legal translation: they can reason about equivalences between known legal concepts, but they cannot infer the correct rendering of concepts that exist only in a specific legal tradition and are absent from their training data.

The generative category is itself internally split: GPT-5.3 performs closer to the NMT systems than to Claude Opus 4.5, indicating that generative architecture alone does not prevent English mediation. The differentiator appears to be alignment: Constitutional AI may encode stronger preferences for accuracy than RLHF, which can favor fluency over terminological exactness, a hypothesis consistent with the patterns though not verifiable without the training protocols.

4.5 Qualitative Analysis of Key Error Patterns

Beyond the quantitative patterns, several qualitative observations merit attention. The first is what may be termed the cascade effect: when a system displaces a foundational concept, the displacement propagates to related terms. The German tripartite classification of offenses and its Greek counterpart (Section 2.2) illustrate this: when LLMs render all three categories through generic terms such as *έγκλημα* (*egklima*: crime) or *αδίκημα* (*adikima*: offence), the severity distinctions that carry procedural and sentencing consequences in both systems become invisible. The distortion can take particularly misleading forms: DeepL, for instance, renders *Vergehen* as *πταίσμα* (*ptaisma*) and *Ordnungswidrigkeit* as *παράβαση* (*paravasi*: violation), swapping two categories while eliminating the third altogether.

A second pattern is the garbling effect observed in Google Translate outputs. In several cases, Google Translate produces outputs that are displaced to the point of semantic incoherence: *Pfandrecht* (pledge) rendered as *εμπράγμα* (*empragma*), a truncated non-word, and *Gemeindeordnung* as *δημοτική παραγγελία* (*dimotiki parangelia*: municipal order). These outputs suggest that Google Translate's DE→EL pathway is so attenuated that it falls back on word-level pattern matching rather than phrasal translation, producing nonsensical combinations that no human translator would produce. This behavior has a concrete, non-speculative basis. The output quality of an NMT system in a given direction is strongly affected by the volume of parallel data available for that specific pair; DE→EL is a low-resource

direction with comparatively little direct parallel legal text, leaving its dedicated model more sparsely trained than DE→EN or EN→EL.

When phrase-level alignments are weak, the system falls back on the most reliable signals that remain, subword and lexical correspondences, rather than on phrasal or terminological units. This fallback is especially damaging in the DE→EL direction because German legal terminology is dominated by long nominal compounds (e.g., *Amts-* + *-haftung*), which a data-sparse model is prone to mis-segment and recombine literally, as in *Amtshaftung* → *δημόσια κράτηση* (*dimosia kratisi*: public detention). The result is therefore not random noise but a predictable degradation mode of compound-rich, low-resource NMT, one that is correspondingly rarer in the EN→EL direction, whose source terms are not compounded in the same way. The frequency of such garbled outputs (approximately 8.1% of Google Translate's total) raises questions about the reliability of NMT systems for low-resource language pairs in specialized domains.

A third finding concerns terms with established international equivalences. *Rechtsstaat* / *κράτος δικαίου* (*kratos dikaiou*: rule of law), *Hypothek* / *υποθήκη* (*ypothiki*: mortgage), and *Berufung* / *έφεση* (*efesi*: appeal) are rendered correctly by all systems in both directions: because they are standardized through EU instruments and comparative scholarship, they appear frequently enough in multilingual corpora to override any English-mediation tendency.

5. Discussion

5.1 Stylistic Flattening and English Mediation as Structural Effects

The findings confirm that stylistic flattening in LLM-mediated legal translation is not an occasional error but a structural feature of the technology. With rates ranging from 18% (Claude Opus 4.5 EN→EL) to 85.7% (Google Translate DE→EL), it affects the majority of outputs across all systems and legal domains, exactly as the framework of Section 2.1 predicts, and its magnitude tracks the alignment regime in a way that confirms the effect is contingent rather than inherent. The practical consequence is that outputs, even when terminologically correct, may fail to meet the register expectations of the target legal system, reducing formal designations to generic equivalents or stripping a specific procedural instrument of its institutional specificity, and thereby signaling a lack of register competence that undermines the document's authority.

English mediation is the second effect with structural roots. That these systems exhibit it even in the DE→EL direction, a pair that does not involve English, is direct evidence of a pivot language problem: rather than translating directly, these systems route the translation through an internal English representation, which not only reduces precision but also imposes the conceptual architecture of a different legal system onto the target. The problem is not new, having been documented in statistical machine translation (Utiyama and Isahara 2007) and in NMT systems using English as a bridge (Lakew, Cettolo, and Federico 2018); the present study shows that it persists in state-of-the-art LLMs and is especially damaging where conceptual precision is paramount. It manifests differently by category: in NMT, it appears architectural, with DeepL's lower mediation suggesting that direct language-pair models partially insulate translation from the pivot, whereas in generative systems it is subtler, arising from the English dominance of training data rather than explicit routing.

5.2 Implications for Training and Post-Editing Practice

The findings bear directly on translator training. First, they underscore the need for critical LLM literacy as a core competence, increasingly reflected in competence models for legal and institutional translation that now incorporate machine translation and post-editing (Prieto Ramos 2024): translators must recognize the specific errors LLMs produce, namely terminological displacement, register flattening, and English-mediated conceptual substitution, the more so because the fluency of these outputs makes a smooth but inaccurate rendering more readily accepted than an awkward but correct one. Second, the differential across architectures makes tool selection a professional competence: generative LLMs may suit legal translation better than dedicated NMT systems given informed post-editing, though Claude's advantage may reflect alignment strategy rather than domain-specific training and so prove unstable across versions. Third, the errors identified here are invisible to monolingual quality assessment; recognizing a formally similar but conceptually distinct substitution requires comparative legal knowledge, which argues for integrating comparative law into translation curricula for pairs spanning different legal traditions.

These competences converge on post-editing practice. Guidelines developed for general-purpose NMT focus on grammar, fluency, and terminological consistency, but legal post-editing must additionally address conceptual displacement and register appropriateness,

two dimensions that require domain expertise rather than general linguistic competence, an added cognitive load consistent with process research showing that post-editing effort varies with text type and that specialized genres impose distinctive demands (Cui, Liu, and Cheng 2023). The study, therefore, suggests a three-tier framework: terminological verification against authoritative sources; conceptual validation of systemic position within the target legal tradition; and register assessment against target-language discourse conventions. Because base systems differ so markedly, post-editing a low-mediation generative model is a fundamentally different task from post-editing a high-mediation NMT system.

5.3 The Limits of Algorithmic Legal Hegemony

Algorithmic legal hegemony must itself be qualified by the evidence. The marked variation in mediation rates shows the tendency is neither uniform nor inevitable: strongest in NMT and weakest in instruction-tuned models, it is significantly mitigated by architectural and alignment choices. This contingency converges with emerging empirical work on generative models in legal translation. Comparing a proprietary and an open-source LLM against an NMT baseline across four English-source legal pairs, Vicent Briva-Iglesias, João Lucas Cavalheiro Camargo, and Gokhan Dogru (2024) found that although automatic metrics favored NMT, human evaluators judged the proprietary model comparable or superior on adequacy and fluency. This divergence both corroborates the present expert-coded finding of generative superiority and explains it: the fluency that inflates automatic scores is what masks the conceptual displacement documented here. That the proprietary model also outperformed the open-source one mirrors the Claude–GPT-5.3 gap above, reinforcing that alignment and training, not architecture as such, govern performance; the present study extends this to a non-hegemonic pair and to conceptual equivalence rather than fluency.

Such variation complicates any simple narrative of technological determinism. Algorithmic legal hegemony is real but contingent, capable of being reinforced or attenuated by design decisions, data curation, and domain-specific knowledge; the problem lies not with LLM technology as such but with the market-dominant configurations that prioritize fluency over conceptual accuracy and treat English as a universal pivot rather than one legal language among many.

The stakes are concrete. Machine translation already serves as a primary channel in institutional settings where non-linguist users cannot detect conceptual displacement, from administrative procedures to migration and the courtroom (Vieira, O'Hagan, and O'Sullivan 2021; Vieira 2024). Without determining legal outcomes itself, it embeds distortions such as a collapsed severity classification or criminal intent rendered as *πρόθεση* (*prothesi*: intent) rather than *δόλος* (*dolos*: dolus), in the very categories on which procedural and sentencing decisions turn. The path forward is not to reject LLMs but to build configurations that respect the conceptual integrity of target legal systems, a question of epistemic responsibility: where disadvantage to non-English legal traditions is predictable and measurable, failing to address it is not a technical limitation but an ethical choice.

6. Conclusions

This study has shown that AI-mediated legal translation into Greek exhibits systematic stylistic flattening and terminological displacement through English mediation: not random errors but structural features rooted in the English-centric composition of training data. Its central contribution is that these effects are architecturally contingent. Generative LLMs, Claude Opus 4.5 in particular, mediate through English far less than dedicated NMT systems, so the problem is amenable to architectural solutions rather than being inherent to the technology.

Three findings stand out. Quantitatively, English mediation and stylistic flattening affect enough outputs and vary enough across systems and directions to constitute a systemic problem requiring structural solutions. Qualitatively, displacement operates through three distinct mechanisms: conceptual substitution, calque from English, and hybrid displacement, each with different implications for post-editing; the confinement of hybrid displacement to the EN→EL direction, across all four systems, indicates that the pairing of a calque with its correct equivalent is driven by the presence of an English source rather than by any architecture-specific reasoning capacity. Theoretically, the concept of algorithmic legal hegemony captures a real but contingent phenomenon: the imposition of Anglo-American legal categories onto non-Anglophone systems through translation technology is neither uniform nor inevitable but a function of design choices that can be interrogated, contested, and modified.

The differential performance of these architectures therefore offers both a diagnostic tool and a normative benchmark for evaluating the impact of translation technology on legal systems. The existence of systems that mediate less demonstrates that solutions exist; what remains open is whether market incentives will favor their adoption or whether the industry will continue to optimize for English-centric metrics at the expense of non-hegemonic legal traditions.

Future research should extend this analysis to full-text legal documents, to additional non-hegemonic language pairs, and to longitudinal designs that track LLM performance across model versions. The integration of glossary-based customization as a mitigation strategy, the development of specialized legal-translation benchmarks for non-hegemonic pairs, and the propagation of LLM-mediated errors into professional practice and legal decision-making all warrant systematic investigation, with implications that reach well beyond translation studies into the sociology of law and legal epistemology.

References

Alcaraz Varó, Enrique. 2009. "Isomorphism and Anisomorphism in the Translation of Legal Texts." In *Translation Issues in Language and Law*, edited by Frances Olsen, Alexander Lorz, and Dieter Stein, 189–192. London: Palgrave Macmillan.

Bai, Yuntao, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, et al. 2022. "Constitutional AI: Harmlessness from AI Feedback." arXiv:2212.08073v1. doi:10.48550/arXiv.2212.08073.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. doi:10.1145/3442188.3445922.

Biel, Łucja. 2014. *Lost in the Eurofog: The Textual Fit of Translated Law*. Frankfurt am Main: Peter Lang.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. 2021. "On the Opportunities and Risks of Foundation Models." arXiv:2108.07258v1. doi:10.48550/arXiv.2108.07258.

Briva-Iglesias, Vicent, João Lucas Cavalheiro Camargo, and Gokhan Dogru. 2024. "Large Language Models 'Ad Referendum': How Good Are They at Machine Translation in the Legal Domain?" *MonTI: Monographs in Translation and Interpreting*, no. 16, 75–107. doi:10.6035/MonTI.2024.16.02.

Chu, Chenhui, and Rui Wang. 2018. "A Survey of Domain Adaptation for Neural Machine Translation." In *Proceedings of the 27th International Conference on Computational Linguistics*, 1304–1319. Santa Fe, New Mexico: Association for Computational Linguistics. https://aclanthology.org/C18-1111.pdf.

Cui, Ying, Xiao Liu, and Yuqin Cheng. 2023. "A Comparative Study on the Effort of Human Translation and Post-Editing in Relation to Text Types: An Eye-Tracking and Key-Logging Experiment." *SAGE Open* 13 (1). doi:10.1177/21582440231155849.

Dodge, Jesse, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. 2021. "Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus." In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 1286–1305. Association for Computational Linguistics. doi:10.18653/v1/2021.emnlp-main.98.

Fairclough, Norman. 1989. *Language and Power*. London: Longman.

———. 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.

Killman, Jeffrey. 2023. "Machine Translation and Legal Terminology." In *Handbook of Terminology*, edited by Łucja Biel and Hendrik J. Kockaert, 3:485–510. Amsterdam: John Benjamins.

Koehn, Philipp, and Rebecca Knowles. 2017. "Six Challenges for Neural Machine Translation." In *Proceedings of the First Workshop on Neural Machine Translation*, 28–39. Vancouver: Association for Computational Linguistics. doi:10.18653/v1/W17-3204.

Lakew, Surafel M., Mauro Cettolo, and Marcello Federico. 2018. "A Comparison of Transformer and Recurrent Neural Networks on Multilingual Neural Machine Translation." In *Proceedings of the 27th International Conference on Computational Linguistics*, 641–652, Santa Fe, New Mexico, USA. Association for Computational Linguistics. https://aclanthology.org/C18-1054.pdf.

Mattila, Heikki E. S. 2013. *Comparative Legal Linguistics: Language of Law, Latin and Modern Lingua Francas*. 2nd ed. Translated by Christopher Goddard. Farnham: Ashgate.

Mossop, Brian. 2014. *Revising and Editing for Translators*. 3rd ed. London: Routledge.

Ouyang, Long, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, et al. 2022. "Training Language Models to Follow Instructions with Human Feedback." In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, 27730–27744. doi:10.52202/068431-2011.

Phillipson, Robert. 1992. *Linguistic Imperialism*. Oxford: Oxford University Press.

Prieto Ramos, Fernando. 2021. "Translating Legal Terminology and Phraseology: Between Inter-systemic Incongruity and Multilingual Harmonization." In "Legal Terminology and Phraseology in Translation," edited by Fernando Prieto Ramos. Special Issue, *Perspectives* 29 (2): 175–183. doi:10.1080/0907676X.2021.1849940.

———. 2024. "Revisiting Translator Competence in the Age of Artificial Intelligence: The Case of Legal and Institutional Translation." In "Competence in Legal and Institutional Translation: Training Challenges and Innovations," edited by Fernando Prieto Ramos. Special Issue, *The Interpreter and Translator Trainer* 18 (2): 148–173. doi:10.1080/1750399X.2024.2344942.

Rossi, Caroline, and Jean-Pierre Chevrot. 2019. "Uses and Perceptions of Machine Translation at the European Commission." *The Journal of Specialised Translation*, no. 31, 177–200. doi:10.26034/cm.jostrans.2019.182.

Šarčević, Susan. 1997. *New Approach to Legal Translation*. The Hague: Kluwer Law International.

Singhal, Prasann, Tanya Goyal, Jiacheng Xu, and Greg Durrett. 2023. "A Long Way to Go: Investigating Length Correlations in RLHF." arXiv:2310.03716v1. doi:10.48550/arXiv.2310.03716.

Utiyama, Masao, and Hitoshi Isahara. 2007. "A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation." In *Proceedings of NAACL-HLT 2007*, 484–491. https://aclanthology.org/N07-1061/.

Vieira, Lucas Nunes. 2024. "Machine Translation and Migration." In *The Routledge Handbook of Translation and Migration*, edited by Brigid Maher, Loredana Polezzi, and Rita Wilson, 221–234. London: Routledge.

Vieira, Lucas Nunes, Minako O'Hagan, and Carol O'Sullivan. 2021. "Understanding the Societal Impacts of Machine Translation: A Critical Review of the Literature on Medical and Legal Use Cases." *Information, Communication & Society* 24 (11): 1515–1532. doi:10.1080/1369118X.2020.1776370.

Zacharopoulos, Vassilios, Christina Koutsogianni Hanke, Dionysia Bourgiezi, Antonia Paradelli, Christina Papatsori, Evdokia Priovolou, and Vassilis Triantafyllidis. 2013. *Deutsch–Griechisches / Griechisch-Deutsches Rechtswörterbuch* [German–Greek / Greek–German legal dictionary]. Edited by Chrisoula Tsepisi and Konstantinos Vathiotis. Athens: Nomiki Bibliothiki.