TRANSLATING LAW: A COMPARISON OF HUMAN AND POST-EDITED TRANSLATIONS FROM GREEK TO ENGLISH

Vilelmini Sosoni, John O'Shea, Maria Stasimioti*

Abstract

Advances in neural machine translation (NMT) models have led to reported improvements in machine translation (MT) outputs, especially for resource-rich language pairs (Deng & Liu, 2018), mainly at the level of fluency (Castilho et al., 2017a, 2017b). NMT systems have been used particularly for the translation of technical and life science texts with short, repetitive, formulaic, and unambiguous sentence types. In contrast, legal translation studies scholars have depicted legal translation as not particularly compatible with MT, mainly because legal texts include features that pose significant challenges to MT (Killman, 2014; Prieto Ramos, 2015; Matthiesen, 2017). As such, the quality of the output varies according to the legal genre and language pair. Using MQM-DQF error typology, this study evaluates the quality of the post-edited and human translation (HT) products of two normative property law texts from Greek to English, a language pair considered to be under-resourced. The time taken by the two translators who participated in the study to complete these products was monitored, and information was collected on their attitudes towards MT and post-editing (PE). The findings indicate neither productivity gains in the case of PE, nor major differences in accuracy or fluency between the post-edited and HT texts, although the number of errors was slightly higher overall in the case of HT, with most occurring at the level of accuracy. Conversely, the post-edited versions contained more errors at the levels of style and verity. Finally, the translators' views on MT and PE were dependent on the MT output quality, while their trust level in the output may have affected the end-product quality.

Keywords: neural machine translation; post-editing; legal translation; property law; translation quality.

TRADUIR EL DRET: COMPARACIÓ DE TRADUCCIONS HUMANES I POSTEDITADES DEL GREC A L'ANGLÈS

Resum

Els avenços en els models de traducció automàtica neuronal (TAN) s'han traduït en una millora dels resultats de la traducció automàtica (TA), especialment, en les combinacions lingüístiques amb molts recursos (Deng i Liu, 2018) i sobretot pel que fa a la fluïdesa (Castilho et al., 2017a, 2017b). Els sistemes de TAN s'han fet servir sobretot per traduir textos tècnics i de ciències de la vida amb frases breus, repetitives, predictibles i sense ambigüitats. Per contra, l'estudi acadèmic de la traducció jurídica ha assenyalat que aquesta no és gaire compatible amb la TA, sobretot perquè els textos jurídics tenen característiques que plantegen problemes importants per a la TA (Killman, 2014; Prieto Ramos, 2015; Matthiesen, 2017). Així, la qualitat dels resultats varia en funció del gènere jurídic i la combinació lingüística. Amb una tipologia d'errors MQM-DQF, en aquest estudi s'avalua la qualitat dels productes de la postedició i de la traducció humana (TH) amb dos textos prescriptius de dret de la propietat del grec a l'anglès, una combinació lingüística que es considera que té pocs recursos. L'estudi va controlar el temps que les dues persones participants en l'estudi van necessitar per acabar aquests dos textos i també es va recollir informació sobre la seva postura en relació amb la TA i la postedició (PE). Els resultats indiquen que la productivitat no millora en el cas de la PE ni tampoc s'observen grans diferències pel que fa a la precisió o la fluïdesa entre els textos posteditats i els fets amb TH. En general, però, el nombre d'errors era lleugerament més elevat en les TH i la majoria d'aquests errors tenia a veure amb la precisió. En canvi, les versions posteditades contenien més errors d'estil i veracitat. Per acabar, les opinions dels traductors sobre la TA i la PE depenien de la qualitat dels resultats de la TA, tot i que el seu nivell de confiança en els resultats pot haver afectat la qualitat del producte final.

Paraules clau: traducció automàtica neuronal; postedició; traducció jurídica; llei de propietat; qualitat de la traducció.

* Vilelmini Sosoni, assistant professor of Economic, Legal and Political Translation at the Department of Foreign Languages, Translation and Interpreting, Ionian University, sosoni@ionio.gr, 0000-0002-9583-4651

John O'Shea, legal translator and independent researcher, info@jurtrans.com, 0000-0002-0998-2368

Maria Stasimioti, PhD candidate at the Department of Foreign Languages, Translation and Interpreting, Ionian University, stasimioti@ionio.gr, 0000-0001-9541-4676

Article received: 11.08.2021. Blind reviews: 12.09.2021 and 29.09.2021. Final version accepted: 26.10.2022.

Recommended citation: Sosoni, Vilelmini, O'Shea, John, & Stasimioti, Maria. (2022). Translating law: A comparison of human and post-edited translations from Greek to English. *Revista de Llengua i Dret, Journal of Language and Law*, *78*, 92-120. https://doi.org/10.2436/rld.i78.2022.3704

Contents

1 Introduction

Recent advances in neural machine translation (NMT) models have led to reported improvements in machine translation (MT) outputs, especially in resource-rich language pairs (Deng & Liu, 2018), mainly at the level of fluency (Castilho et al., 2017a, 2017b). NMT models have been increasingly used in translation project workflows (Mellinger, 2018; Vieira et al., 2019) and have been widely associated with productivity gains over human translation (HT) "from scratch" (Gaspari et al., 2014; Toral et al., 2018; Moorkens et al., 2018; Jia et al., 2019; Vieira, 2019; Stasimioti & Sosoni, 2020). In fact, increasing the integration of MT into workflows has resulted in what has been termed *augmented translation* (Lommel, 2018a). This includes adaptive MT, which incorporates human corrections on the fly. Now included in computer-assisted translation (CAT) tools, augmented translation is facilitating the incorporation of MT in workflows and challenging the established notion of a concrete source text (Pym, 2013). Moreover, augmented translation has given rise to innovations in metrics, assessment, and expectations of quality. Productivity has emerged as a core metric of quality, with the notion of a "good enough" translation gaining visibility (Angelone et al., 2020, p. 4).

Augmented translation and MT are not necessarily suitable for all genres and text types, however. For example, MT has shown better performance with short, repetitive, formulaic, and unambiguous source text (ST) sentences (Bawden, 2018; Moorkens et al., 2018), leading to the extensive use of MT in commercial translation of technical and life science texts with these types of sentences. MT has also been used in legal translation scenarios to various ends. For example, MT has been used to generate a draft translation to triage electronically available materials that might require human translation (e.g., foreign language documents in e-discovery settings) (Foster & Northrop, 2011, p. 45; Nelson & Simek, 2017). MT has also been used for patent applications (Vieira et al., 2021) and in matters related to immigration applications (UNHCR, 2014; Oakes, 2016, p. 893). However, given that legal texts are often regarded as demanding to translate (Wiesmann, 2019, p. 121) and include features that pose significant challenges to MT (Killman, 2014; Prieto Ramos, 2015; Matthiesen, 2017), it is unsurprising that the quality of MT output varies according to the legal genre, language pair, and legal system. Therefore, there appears to be justified scepticism about the extent to which MT is compatible with legal texts, and by extension legal cultures, in such a way that PE could become a worthwhile alternative to HT.

The aim of the present article is threefold: (a) using MQM-DQF error typologies, to evaluate, in the resource-poor Greek–English language pair, the quality of post-edited and human translations of two normative, legal texts that are specific to Greek property law, (b) to record the time taken by two translators to complete these PE and HT tasks, and (c) to analyse their attitudes to MT and PE on the one hand and HT on the other.

The article is structured as follows: Section 2 discusses the particular characteristics of legal translation and reports on related work as regards MT and legal texts; Section 3 explains the methodology employed in the study, describing the texts used and the MT system deployed, the translators and annotators involved, the actual translation process followed, post-editing and error annotation stages, and the error typology framework used; and Section 4 presents and discusses the findings. Finally, Section 5 summarises the study and refers to future work.

2 Related works

Cao (2007) indicates that the translation of law has played an integral part in interaction among nations throughout history. One of the first legal texts translated from one language to another is the Treaty of Kadesh (1271 BCE), also known as the Eternal Treaty or the Silver Treaty (Mattila, 2006, p. 7). Inscribed on a silver tablet, this treaty was translated from Akkadian to Egyptian to establish peace and brotherhood between Hatti and Egypt, two strong empires of the Near Eastern world. Since then, legal translation has expanded in scope and volume, playing a growing role in an increasingly interconnected and globalised world in which companies operate globally and people travel, work, and carry out cross-border transactions.

Legal texts have always been regarded among the most complex specialised texts, with a range of features that pose challenges to MT (Killman, 2014; Prieto Ramos, 2015). More specifically, legal texts are written in a language that operates as a functional variant of natural language referred to as *lingua legis* (Matulewska,

2007), with its own domain of use and particular linguistic norms (e.g., phraseology, vocabulary, hierarchy of terms, and specialist meanings). Legal language is characterised by several special morphosyntactic, semantic, and pragmatic features (Mattila, 2006, p. 3) that set it apart from other registers; it is also known for its "formulaicity, standardisation, petrification and rituals" (Biel & Engberg, 2013, p. 5) and is rich in terminology, passive and impersonal forms, hypotactic structures, and complexity of modifiers. This specialised language lacks emotivity and seeks precision, yet is simultaneously characterised by vagueness, generality, ambiguity, and declarative and pompous style (Sosoni et al., 2018). More importantly, legal terms refer to system-bound concepts which are incongruous and unique to each legal system and legal tradition (Prieto Ramos, 2021), leading to terminological asymmetry that is difficult for translators to overcome. Consequently, it is commonly agreed that true equivalence in legal translation is either "random" (Gémar, 2002, p. 174) or futile (Cao, 2007, p. 34), suggesting that the equivalence sought in legal texts is necessarily functional in nature. Functionalism, which Šarčević (1997) fervently supports, presupposes that a target text (TT) is relatively independent from its ST, "a rendering of information by translators who have made active decisions based on their insights into the source and target situations and cultures and the communicative task emerging from the relation between the two situations" (Engberg, 2021, p. 8).

This discussion underscores the key areas of legal translation challenges that give rise to substantial complexity from the perspective of both humans and natural language processing systems. As Prieto Ramos (2015, p. 20) observes, "the qualitative analysis of the different variables and layers of system-bound legal meaning constitutes a challenge for the production of acceptable drafts through machine translation […], even in more 'linguistically-predictable' contexts." According to Wiesmann (2019, p. 121), several features of legal texts are likely to cause problems for MT, such as system-bound terminology, frequently occurring abbreviations, formulaic and elliptical usage, as well as genre-specific deviations from normal language usage. Therefore, since legal texts have features that pose major challenges to MT, the question arises as to what extent MT can now translate legal texts – or, more specifically, normative legal texts – into another legal language in a way that justifies the use of post-editing. Normative texts, such as codes of law, contracts and constitutions, are prescriptive rather than descriptive texts, as in the case of regulatory instruments (Šarčević, 1997, p. 11) that define rights and duties, and possibly consequences if the respective norm is breached. It can thus be understood why the misuse of MT can have serious consequences, particularly in high-stakes settings.

Given the potential complexity involved in the translation of legal texts, it is unsurprising that to date only a small group of researchers have attempted to investigate the use of MT in legal settings. Researchers have focused on MT output quality, its use by non-translators, and on PE in comparison with HT. Regarding MT quality, Yates (2006) examined the accuracy of Babel Fish (an RBMT system) in translating excerpts from the civil codes of Mexico and Germany into English, identifying severe errors in the output that interfered with the sense of these texts, and concluding that Babel Fish was mostly inaccurate. In a study focusing on Google Translate, an SMT system at the time, Killman (2014) evaluated Spanish-English output accuracy for a sample of over 600 legal terms and phrases from a collection of judgement summaries produced by the Supreme Court of Spain. Despite being a general-purpose system, Google Translate produced accurate English translations in nearly 65% of the cases, which posed a variety of legal translation challenges (Killman, 2014). In a study comparing Bilingual Evaluation Understudy (BLEU) scores for in- and out-of-domain SMT and NMT systems in the German-English language pair in the domains of law, medicine, IT, the Koran, and subtitles, Koehn and Knowles (2017) found that in- and out-of-domain SMT performance was superior to NMT in the case of law (Koehn & Knowles, 2017, p. 30).

In the context of EU translation, Arnejšek and Unk (2020) carried out a study at the Slovene Language Department of the European Commission's Directorate General for Translation (DGT), analysing errors reported in NMT output produced by eTranslation, the MT service provided by the European Commission, and concluding that, in legislative acts, "the biggest problem with NMT seems to be terminological errors and inconsistency", while in "terminology-heavy Annexes, the terminological errors and inconsistency might make NMT useless, especially if there are tables with fragmented text and many abbreviations" (Arnejšek & Unk, 2020, p. 7). Şahin and Dungan (2014) conducted a study with students, investigating the quality of their renditions and their reactions when they translated technical, literary, media, and legal texts from English into Turkish with the assistance of either printed/online resources or output they could post-edit from Google Translate, which was SMT at the time. Quality was slightly better when the legal texts were translated rather

than post-edited in the resource-poor Turkish-English pair. In addition, the participants noted that they would not choose to post-edit the legal text, as they did not feel comfortable with the output.

3 Methodology

Our study evaluates translation quality with respect to the translation of two normative legal texts from the property law domain, focusing on the resource-poor Greek-English language pair. Specifically, we investigated full PE of NMT output and HT from scratch of these texts. In addition, we compared the time taken to complete these tasks and surveyed the professional translators who participated in the study to gain insights into their attitudes toward and experience with using MT and PE on the one hand and HT on the other, given that openness to MT use is a positive factor for PE performance (De Almeida, 2013; Mitchell, 2015).

To that end, an NMT system was first trained using in-domain data (for a description, see sub-section 3.2). Two texts (a 500-word extract from urban planning legislation and a 500-word extract from a property sale agreement; see sub-section 3.1) were then selected. Two translators (see sub-section 3.3) were asked to translate and post-edit the two texts, and two annotators (see sub-section 3.3) were asked to annotate the translations and the post-edited texts using MQM-DQF error typology (see sub-section 3.5). Finally, questionnaires and interviews (see sub-section 3.6) were used to gather information about the attitudes of the two translators regarding the MT output before they completed the PE and HT tasks.

3.1 Source texts

The source texts (STs) used in this study were both normative, specialised texts from the domain of property law, given the importance of this domain in the Greek translation market as a result of increased inter-EU migration (Sosoni & O'Shea, 2021). The marketing of Greece as an investment destination in recent years and reforms to the Greek legal system have made the country more attractive to businesses and have increased demand for English translations in the area of property law.

For the purposes of this study, two texts from different genres were chosen: a 500-word extract from planning legislation (Text A; see Appendix A) and a 500-word extract from a property sale agreement (Text B; see Appendix A). While the Greek texts are primarily prescriptive and normative, their English translations are not. Based on Cao's distinction (2007, pp. 10–12), the translations in the present study are informative with a descriptive function and may be used for general legal or judicial purposes.

It should be underscored that property law texts can be particularly challenging to translate. Although the idea or philosophy of "property" underlies property law in all European systems and countries, there are substantial differences in the way the concept of property is perceived and interpreted across systems and countries. As far as Greek and British property law systems are concerned, significant differences exist at the conceptual and organisational levels, resulting in particularly difficult terminological challenges (Valeontis & Krimpas, 2014, pp. 247–248; Vlachopoulos, 2014, pp. 77–94). The work of translators is further complicated by the relative dearth of bilingual resources and limited documentation in English on the topic.[1]

Both texts have comparable readability scores, being of similar difficulty and text complexity according to the Greek readability analyser, which is based on several indices, including the Flesch Reading Ease test, Flesch-Kincaid Grade Level, SMOG, and Flesch Fog Index.[2] In particular, they were both found to be challenging texts, suitable for university graduates. Further analysis using Voyant Tools[3] revealed that Text A, the excerpt from planning legislation, had a vocabulary density of 0.512 and an average of 49.6 words per sentence, while Text B, the excerpt from the property sale agreement, had a vocabulary density of 0.537 and an average of 81.3 words per sentence.[4] The density scores for the texts indicate fairly complex texts, in line with the findings of the Greek readability analyser.

1 Additional resources in English can be found in Giannakourou (2006), and Serraos, Gianniris, and Zifou (2005).

2 For more information on readability, see the index.

3 Visit the website for more about Voyant Tools.

4 Vocabulary density is the ratio of the number of words in a document to the number of unique words in that document. The closer

3.2 The NMT engine and MT outputs

The MT system was trained by a Cyprus-based company that provides specialised services and software development for individuals and companies in the translation industry, Lexorama Ltd.[5] The system was trained using Transformer architecture (Vaswani et al., 2017) on 10.5 million sentences in total, of which 5.3 million sentences were specific to the legal/legislative domain and 5.2 million sentences were obtained from public sources, such as Opus,[6] containing a mix of domains, including EU legislation, medical, and news. Tokenisation, that is, parsing text into smaller units such as the division of a sentence into words, was performed with a SentencePiece model (Kudo & Richardson, 2018) relying on a vocabulary of 32,000 tokens. Certain custom domain adaptation techniques were applied during training.

BLEU scores were used to evaluate the MT output quality. BLEU is a metric for comparing a candidate translation to one or more reference translations by counting the number of matches for *n*-grams (Papineni et al., 2002). An *n*-gram is a sequence of *n* words: a 2-gram (which is referred to as a bigram) is a two-word sequence such as "please turn", or "your key", and a 3-gram (or trigram) is a three-word sequence such as "please turn your", or "turn your key". BLEU is the standard metric used in the MT community because it provides a quick, rough estimation of MT output quality. A perfect match results in a score of 1.0 – or 100, if we use percentage values – whereas a perfect mismatch results in a score of 0.0 (Papineni et al., 2002). However, very few translations attain a score of 100 unless they are identical to the reference translation. For this reason, even a human translator will not necessarily achieve a perfect score. The higher the BLEU score, the more similar the two texts are. Generally speaking, a score below 0.15 means that the engine is not performing optimally and PE is not recommended as it would require a lot of effort to finalise the translation and reach publishable quality, while a score above 0.50 is a very good score and means that significantly less PE is required to achieve publishable translation quality (Lavie, 2011). The reference translations used to calculate the BLEU score were produced by a professional translator with over 20 years of experience working with legal texts (see Appendix B).

The analysis showed that both MT outputs obtained not only comparable, but almost identical BLEU scores, that is, 32.59 for Text A and 32.71 for Text B. This means that the quality of the MT outputs was comparable and that the texts were good enough to be post-edited, albeit with a considerable amount of PE required.

3.3 Participants

Given the limited pool of translators working in the Greek-English (EL–EN) language combination, we issued a call for participation in the study, outlining very specific criteria (i.e., professional legal translators working in the EL–EN language pair with at least 5 years of experience), and shared it with members of the Panhellenic Association of Translators. Four professional translators replied to our call, two of whom were chosen to translate and post-edit the texts in the study on the basis of their immediate availability and comparable experience (i.e., each had 10 or more years of experience in legal translation); the remaining two carried out the error annotation task as described in Section 3.4.

Translator A (male, age group 35–44) and Translator B (female, age group 55–64) were both freelance translators with at least a decade of professional experience in legal translation (10 and 14 years, respectively), translating mainly notarial deeds, contracts, court judgments, legal opinions drawn up by jurists, defence statements, legal instrument service reports, property deeds, leases, articles of association, partnership and employment agreements, and family law petitions.

The other two translators were asked to perform the annotation task, given their familiarity with error annotation and greater professional experience. Their legal translation experience ranged from 15 to 25 years, having translated mainly contracts, property agreements, judgments and pleadings, articles of association, and legislation. Finally, both had extensive experience working with the MQM-DQF error typology framework

to one, the greater variety of words (denser), while a higher ratio indicates simpler text with words reused.

5 Visit the website.

6 For more information, see the Opus website.

(see Figure 1), as they provide MT evaluation services and training in MT and PE for continuing professional development programmes of translator associations in Greece.

Neither the translators nor the annotators received remuneration for their work and their participation was voluntary. To avoid bias, they were not informed of the aims of the study; they were simply told that the research was focused on the quality of the end products and that it was important to closely follow the guidelines provided. Before the study, both translators and annotators signed a consent form in accordance with the requirements of the Research Ethics and Deontology Committee of the Ionian University.

3.4 The study

Before carrying out the tasks, the two translators were asked to fill in a questionnaire.[7] The questionnaire consisted of 24 questions on demographics, professional translation experience in legal or other areas, their use of CAT tools and MT when translating legal texts, as well as their opinions of MT and PE. Fourteen of the 24 questions were close-ended, and 10 were open-ended.

Translator A translated Text A and post-edited Text B, while Translator B post-edited Text A and translated Text B. Since the translators were not familiar with the practice of PE, a brief training session was provided before completing the tasks, along with the specific guidelines they were to follow closely. These guidelines were based on the comparative overview of full PE guidelines provided by Hu and Cadwell (2016), as proposed by the Translation Automation Users Society (TAUS) (2016) and Flanagan and Christensen (2014). More specifically, the participants were asked to retain as much raw MT output as possible, ensure that the message transferred was accurate, correct any omissions and/or additions, mistranslations, morphological errors, misspellings and typos, wrong terminology, inconsistencies, and errors in punctuation and style. Finally, they were asked not to introduce preferential changes.

Following the training, the translators received by email a Microsoft (MS) Word file with the ST they had to translate (File A), and another MS Word file with the ST and MT output they had to post-edit (File B). They were asked to type their translation under the ST in File A and also correct the MT output using the track changes feature in Word in File B. They could use resources such as paper and online dictionaries and databases like IATE, but they were told not to use any CAT tools, translation memories or MT systems. This was done to ensure that the translators would translate without any reliance on pre-existing solutions, such as translation memories built up over the years by themselves or provided to them by clients. The fact that they were asked to follow a workflow that did not match their normal workflow may limit the conclusions that can be drawn from the comparison. The translators were also asked to use the time-tracking software Clockify,[8] chosen because it is easy-to-use freeware that allows individual and group projects to be created and monitored by the project owner, in this case one of the researchers.

After completing the HT and PE tasks, a structured mini-interview was conducted to determine their experience with the different tasks and take note of any differences or similarities. The study concluded with the annotation of the HT and post-edited texts by the two annotators, who followed an adapted version of the MQM-DQF error typology framework discussed below. Both annotators carried out the error analysis independently, but were then asked to convene and review instances where they disagreed. Their final joint decision was the one used in the study.

3.5 Error typology framework

The harmonised MQM-DQF error typology is a *de facto* standard for translation quality assessment that is widely used in both industry and academia. It provides a common framework for the description and categorisation of translation errors from a functional perspective (Lommel, 2018b), and examines whether a translation meets particular specifications and identifies specific error types in the translated text. Figure 1 shows how the framework establishes a hierarchy of error types consisting of up to two levels. At the top,

7 See the questionnaire for more information.

8 Visit the Clockify website.

there are seven primary error types, also known as *branches* or *dimensions*, while each of the primary error types includes several sub-types. The error types are defined as follows:

Accuracy: Refers to errors related to the semantic relationship between the source and target texts. It includes omissions, additions, mistranslations, untranslated content, and improper exact TM matches. It is extremely important that terminological errors be included in the sub-type of mistranslation (of technical relationship).

Fluency: Refers to errors related to the form or content (rather than the meaning) of a target text, such as errors in punctuation, spelling, grammar, grammatical register (e.g., use of informal pronouns or verb forms when their formal counterparts are required), errors in links/cross-references and in character encoding, as well as inconsistencies.

Terminology: Interestingly, terminology does not include terminological errors, which belong to the category of accuracy. On the contrary, terminology includes errors related to the inconsistent use of a term with a termbase (a term is used inconsistently with a specified termbase), as well as inconsistent use of terminology within a text.

Style: Refers to errors related to the overall feel of a text or adherence to style guides. It includes cases of awkward style, unidiomatic style, as well as deviations from company or third-party style.

Locale convention: Refers to errors related to adherence to locale-specific guidelines (e.g., for numbers and/or digits, addresses, dates, telephone format, currencies, measurements, shortcut keys, and other types of local formatting).

Design: Refers to errors related to non-textual (design) aspects of the content, such as links, markup errors, formatting, length, and text truncation or expansion.

Verity: Refers to errors dealing with the relationship of the content to the world in which it exists, that is, culture-specific references.

In the present study, as can be seen in Figure 1, the typology was slightly modified, to (a) accommodate the specificities of the texts under analysis, that is, the fact that no CAT tools were used and that such normative legal texts are rich in terminology, and (b) take into account previous studies which revealed that certain categories confused the annotators, while others were missing from the typology (Klubička et al., 2018, p. 9). In particular, the terminology category was deleted, as no termbase was provided and inconsistencies in terminology fell under the sub-type of "inconsistencies within the text" in the accuracy error type. In addition, the collocations subcategory was added to the style category as it is clearly missing from the typology and is one of the major sources of translation errors. According to Hoey (2005, p. 5), "collocation is a psychological association between words […] and is evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution," and is considered to be a subcategory of formulaic language (Wray, 2002). As aptly put by Newmark (1988, p. 213), "[i]f grammar is the bones of a text, collocations are the nerves, more subtle and multiple and specific in denoting meaning." Finally, the category of design was removed from the typology, because there was no special formatting in our texts.

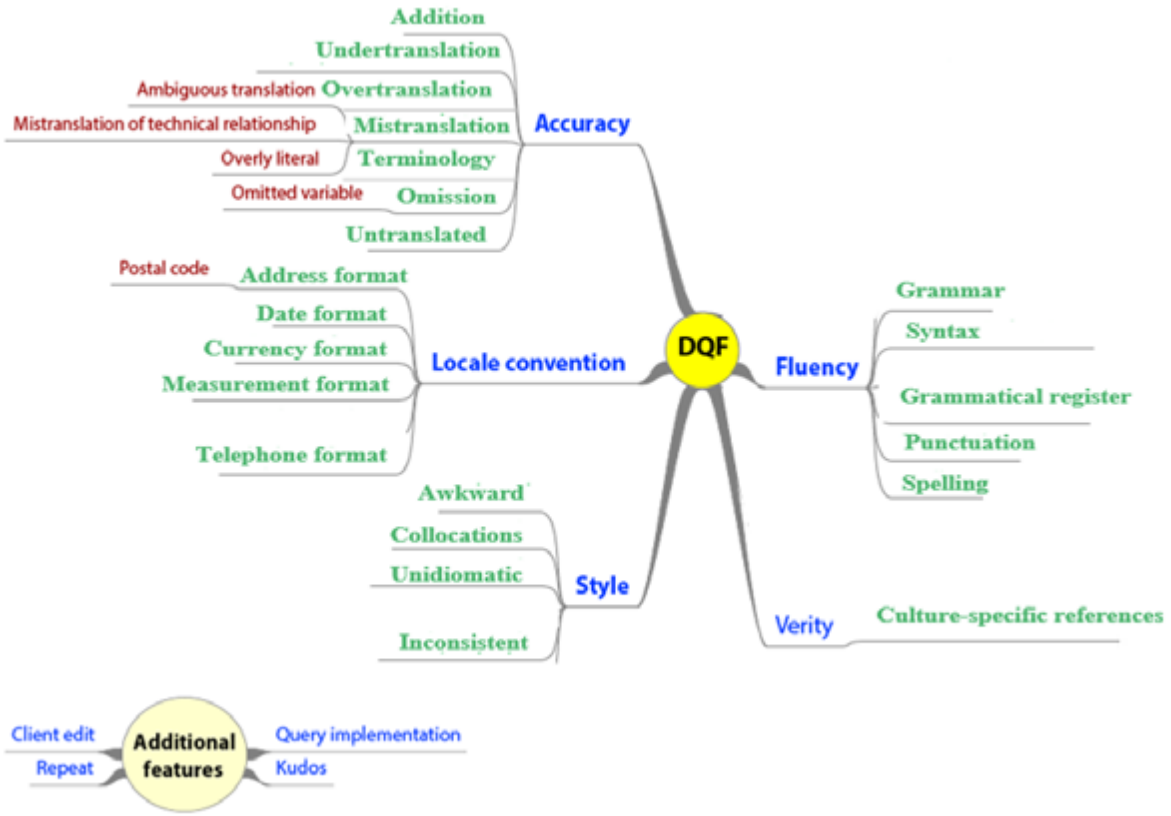

Figure 1. The adapted version of the harmonised MQM-DQF error typology used in the study

4 Findings

4.1 Task time

With respect to the time taken to complete the task, no productivity gains were observed in the case of PE. In fact, the participants needed 38 minutes on average to translate the texts from scratch, and 41.6 minutes on average to post-edit the MT output. More specifically, Translator A needed 23 minutes for translation from scratch and 23.2 minutes for PE, while Translator B needed 53 minutes and 60 minutes, respectively.

4.2 Error analysis and discussion

The qualitative analysis, however, indicated differences, given that the number of errors was slightly higher overall in the case of HT (on average, 21 HT errors vs. 20 in the case of PE). The number of errors was higher at the level of accuracy (on average, 17 HT errors vs. 11 in the case of PE). Conversely, PE entailed more errors at the level of style and verity (see Table 1 and Appendix C for a detailed presentation of all the errors identified).

As for PE, in Text A, we identified two instances in which Translator A introduced new errors and five cases where he spotted the error in the MT output but did not provide an appropriate rendition. Nevertheless, Translator A identified all the errors in the MT output. In the case of Text B, Translator B failed to correct three errors in the MT output. In three other instances, she spotted the error in the MT output but did not provide the correct rendition, and in one instance she incorrectly modified accurate and fluent MT output, thus introducing an error where one had not previously existed. Finally, in the two segments left untranslated by the MT engine, Translator B's HTs included seven errors.

Table 1. Number of errors per error category per modality

Number of errors per error category per modality						
	TEXT A		**TEXT B**		**AVERAGE**	
	HT (Translator B)	**PE (Translator A)**	**HT (Translator A)**	**PE (Translator B)**	**HT**	**PE**
accuracy	20	7	13	15	17	11
fluency	3	4	1	1	2	3
style	2	4	2	5	2	5
locale convention	0	0	0	0	0	0
verity	0	1	0	0	0	1

A closer look at the sub-categories in Table 2 reveals that the main accuracy errors were terminological, mistranslations and omissions, in the case of both HT and PE. Unexpectedly, HT, on average, gave rise to more mistranslations (6 vs. 2) and omissions (3 vs. 1). However, the number of terminological errors was slightly higher on average in the case of PE (9 vs. 8). One example of a terminological error involved the Greek term "αγροτεμάχια" in Text A, which was translated erroneously as "fields" in the HT and equally erroneously as "agricultural parcels" in the post-edited version. Interestingly, this is a case where the MT output, "land parcels," was correct, but was ignored by the translator. Another example of a terminological error involves the multiword expression "τίτλος κτήσης" in Text B, which was translated erroneously as "acquisition deed" in the HT and equally erroneously as "property title" in the post-edited version. This is a case where the erroneous MT output, "deed of conveyance," was recognised by the translator, who however failed to provide a correct rendition such as "title deed." Another example from Text B is a mistranslation involving the Greek expression "νόμιμη ή μη υπόγεια στάθμη," which means a legal or illegal underground property level. The expression was mistranslated as "legal or non-underground level" in the HT and as "lawful or non-underground level" in the post-edited version. Interestingly, the translators' choices in both cases were similar and arose from a lack of understanding of the ST, due to its syntax. The "non" element was taken to refer to the underground level rather than its lawfulness in the MT output, "at a lawful or non-basic level." The output includes an additional error, namely that "underground" was rendered as "basic."

Table 2. Number of accuracy errors per error subcategory per modality

Number of accuracy errors per error subcategory per modality						
	TEXT A		**TEXT B**		**AVERAGE**	
	HT (Translator B)	**PE (Translator A)**	**HT (Translator A)**	**PE (Translator B)**	**HT**	**PE**
addition	0	0	0	0	0	0
omission	3	1	2	0	3	1
mis-translation	5	1	7	2	6	2
over-translation	12	5	3	13	8	9
under-translation	0	0	1	0	1	0
untranslated	0	0	0	0	0	0

At the level of fluency, as can be seen in Table 3, the main errors involved grammar and syntax, given the lengthy and complex sentence patterns of the original Greek legal texts. Interestingly, the HT and the post-edited products had the same average number of errors in the areas of syntax, grammar and grammatical register. One example of a fluency error at the level of syntax in the case of both products involved the rendering of a very long sentence with many subordinate clauses. The Greek sentence in Text A read:

> Οι διατάξεις της παρ. 1 του άρθρου 1 του ν.δ. 1024/1971 εφαρμόζονται και επί γηπέδων, που κείνται εκτός σχεδίου πόλεως και εκτός ορίων οικισμών και ανήκουν σε έναν ή πλείονες ιδιοκτήτες, επί

> των οποίων έχουν ανεγερθεί μέχρι τις 28.7.2011 οικοδομήματα νομίμως ανεγερθέντα ή αυθαίρετα, υπαγόμενα στις διατάξεις του παρόντος, με την επιφύλαξη των οριζομένων στις διατάξεις του άρθρου 89 του παρόντος.

The HT rendition of this sentence did not contain the word order changes needed to make it fluent in English. This disfluent rendition read as:

> The provisions of Article 1(1) of Legislative Decree 1024/1971 also apply to fields located outside the urban plan and outside settlement boundaries which belong to one or more owners, and on which permanent or irregular structures were built before 28.7.2011 which are subject to the provisions hereof, without prejudice to the stipulations of Article 89 hereof.

A similar rendering was observed in the post-edited product:

> The provisions of Article 1(1) of Legislative Decree 1024/1971 shall also apply to plots which lie outside town plans and outside the boundaries of hamlets and which belong to one or more owners, *and* on which plots buildings have been erected, whether lawfully or illegally, by 28 July and which are subject to the provisions hereof, without prejudice to the provisions of Article 89 hereof.

Notably, the syntax was equally unnatural in the MT output:

> The provisions of Article 1(1) of Legislative Decree 1024/1971 shall not apply to plots which are outside the town plan and outside the boundaries of settlements and belong to one or more owners on which buildings have been lawfully erected or arbitrarily erected by 28.7.2011, which fall within the provisions hereof, without prejudice to the provisions of Article 89 hereof.

A more natural rendition would be:

> Without prejudice to the provisions of Article 89 hereof, the provisions of Article 1(1) of Legislative Decree 1024/1971 shall also apply to plots located outside of a town plan and outside the boundaries of settlements, which belong to one or more owners, and on which prior to 28.7.2011 buildings falling within the provisions hereof - whether erected lawfully or without planning permission - had been erected (see Appendix B, Text A).

Table 3. Number of fluency errors per error subcategory per modality

Number of fluency errors per error subcategory per modality						
	TEXT A		**TEXT B**		**AVERAGE**	
	HT (Translator B)	**PE (Translator A)**	**HT (Translator A)**	**PE (Translator B)**	**HT**	**PE**
grammar	1	1	1	1	1	1
syntax	1	2	0	0	1	1
grammatical register	1	1	0	0	1	1
punctuation	0	0	0	0	0	0
spelling	0	0	0	0	0	0

At the level of style, as Table 4 shows, the main errors both in both HT and PE modalities involved awkward renderings not consistent with the stylistic conventions of the English legal genre and the expectations of the TT readership. Here, PE involved more errors on average than HT (5 vs. 2). In Text B (see Appendix B) the Greek expression "Σε περίπτωση μεταμέλειας των πωλητριών […]" was rendered as "In the event the vendors change their mind […]" in the post-edited translation version, which is not equivalent at the idiomatic level. The HT, on the contrary, was stylistically correct: "In the event of remorse on the part of the vendors […]."

Table 4. Number of stylistic errors per error subcategory per modality

Number of stylistic errors per error subcategory per modality						
	TEXT A		**TEXT B**		**AVERAGE**	
	HT (Translator B)	**PE (Translator A)**	**HT (Translator A)**	**PE (Translator B)**	**HT**	**PE**
collocation	1	0	1	1	1	1
awkward	1	3	1	2	1	3
unidiomatic	0	0	0	1	0	1
inconsistent	0	1	0	1	0	1

Finally, at the level of verity, the absence of errors in the case of HT is remarkable, while, in the case of PE, neither participant was able to properly render culture-specific items. In particular, in Text B, in the case of the term "υποθηκικό βάρος," the translator retained the MT output "mortgage lien" which, though semantically accurate, is tied to the common law system and the reality of the countries that use it, such as the UK and the US. A less specific variant such as "mortgage" would suffice in this case. In Text A, the translator disregarded the correct MT rendering of the culture-specific item "οικισμός" ("settlement"), and translated it instead as "hamlet," thus introducing an error into the PE version.

4.3 Questionnaire and interviews: translators' perceptions and translators' actual experience

In the pre-task questionnaire, the translators were asked to give their opinion of MT. Both indicated that MT can be useful if used responsibly and transparently, and the MT output is effectively edited. However, they specifically expressed a preference not to post-edit, as they found it slowed them down and was frustrating. Both translators also noted their top concerns as the cost involved in training an engine and the myriad issues surrounding confidentiality and ethics, such as ownership of training data. The translators indicated their work does not involve PE tasks, observing that PE is suitable for MT that is used for boilerplate documents, such as user manuals. However, the translators also mentioned that, should PE ever encroach on the freelance world of legal translation such that their daily work involved more PE than HT, they would find this new work configuration frustrating. Both translators pointed to the fact that PE would always be essential to ensure the quality of any final product intended for dissemination, and underscored the fact that proper training of translators in PE is a necessity if MT is to be effectively used in the translation sector.

In the post-task interview, Translator A indicated that he would have preferred to translate the two texts from scratch, that is, to rely on his own ability to translate. He approached the MT output with scepticism and fear and felt inclined to double-check terms before validating them in the output. He was frustrated because a segment was left untranslated in the MT output and pointed out that this untranslated segment made him worry about the impact that such segments could have on pricing models for PE that may be used in the industry. He would also have preferred to work in a CAT tool environment where he would have felt more confident in his translation choices, especially with issues stemming from non-equivalence at the conceptual and terminological levels.

Translator B found the MT output of very high quality and would have preferred to post-edit both texts rather than translate from scratch in the other task. She identified ST comprehension as her biggest challenge and observed that this was due to the long and complex syntactic patterns of the STs and, more specifically, the distance between the subject and the verb. Of particular interest in this respect in the case of Translator B is the fact that her trust of the MT output may have interfered with her ability to correct MT errors. She appeared to rely too heavily on the MT output at times, failing to correct three errors in the post-edited version, unlike Translator A, who identified all the errors in the MT output. Therefore, Translator B's disposition towards MT may have affected her judgement.

5 Conclusion

This is a small-scale analysis involving small text samples, and only two translators and legal genres, namely, property planning legislation and sale agreements. Yet the analysis indicates an apparent lack of productivity gains in the case of PE. It remains to be explored whether there are technical or cognitive gains to be had when engaging in PE or HT that could justify choosing one translation mode over the other.

In addition, there appear to be no major differences in accuracy or fluency between the post-edited and the human-translated texts, although the qualitative analysis indicates that such differences are in fact present, since in the HT the number of errors is slightly higher overall, with the majority of errors found at the level of accuracy, whereas, conversely, the post-edited versions contained more errors at the level of style and verity.

Interestingly, Translator B seemed to rely more heavily on the MT output, failing to identify the erroneous MT output and leaving three errors in the post-edited version, in contrast with Translator A, who identified all the errors in the MT output. Translators' attitudes towards MT may therefore affect their judgement and, by extension, the quality of their work. Another observation involves the nature of errors that posed challenges. Translator A identified non-equivalence at the conceptual and terminological levels as the biggest challenge, while Translator B identified the complex syntactic pattern of the STs.

In the future, we intend to expand upon this research with further text samples from different legal domains and additional translators in order to arrive at more generalisable conclusions. We also wish to compare in-domain MT systems with generic ones, such as eTranslate and DeepL, and integrate them in a CAT tool environment in order to (a) establish whether in-domain systems are superior to generic systems and lead to better quality post-edited texts, and (b) replicate the working conditions of translators who work with CAT tools and often use generic MT systems rather than tailor-made ones.

References

Angelone, Erik, Ehrensberger-Dow, Maureen, & Massey, Gary. (2020). Introduction. In Erik Angelone, Maureen Ehrensberger-Dow, & Gary Massey (Eds.), *The Bloomsbury companion to language industry studies* (pp. 1–14). Bloomsbury. https://doi.org/10.5040/9781350024960.0005

Arnejšek, Mateja, & Unk, Alenka. (2020). Multidimensional assessment of the eTranslation output for English–Slovene. In André Martins, Helena Moniz, Sara Fumega, Bruno Martins, Fernando Batista, Luisa Coheur, Carla Parra, Isabel Trancoso, Marco Turchi, Arianna Bisazza, Joss Moorkens, Ana Guerberof, Mary Nurminen, Lena Marg, & Mikel L. Forcada (Eds.), *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation* (pp. 383–392). European Association for Machine Translation.

Bawden, Rachel. (2018). *Going beyond the sentence: Contextual machine translation of dialogue. Computation and language* [Doctoral dissertation]. Université Paris Saclay (COmUE).

Biel, Łucja, & Engberg, Jan. (2013). Research models and methods in legal translation. *Linguistica Antverpiensia*, *12*, 1–11. https://doi.org/10.52034/lanstts.v0i12.316

Cao, Deborah. (2007). *Translating law*. Multilingual Matters. https://doi.org/10.21832/9781853599552

Castilho, Sheila, Moorkens, Joss, Gaspari, Federico, Calixto, Iacer, Tinsley, John, & Way, Andy. (2017a). Is neural machine translation the new state of the art? *Prague Bulletin of Mathematical Linguistics*, *108*, 109–120. https://doi.org/10.1515/pralin-2017-0013

Castilho, Sheila, Moorkens, Joss, Gaspari, Federico, Sennrich, Rico, Sosoni, Vilelmini, Georgakopoulou, Yota, Lohar, Pintu, Way, Andy, Miceli Barone, Antonio, & Gialama, Maria. (2017b). A comparative quality evaluation of PBSMT and NMT using professional translators. In Sadao Kurohashi & Pascale Fung (Eds.), *Proceedings of Machine Translation Summit XVI, Vol. 1: Research Track* (pp. 116–131). MTSummit.

De Almeida, Giselle. (2013). *Translating the post-editor: An investigation of post-editing changes and correlations with professional experience across two Romance languages* [Doctoral dissertation]. Dublin City University.

Deng, Li, & Liu, Yang. (2018). A joint introduction to natural language processing and to deep learning. In Li Deng & Yang Liu (Eds.), *Deep learning in natural language processing* (pp. 1–22). Springer. https://doi.org/10.1007/978-981-10-5209-5

Engberg, Jan. (2021). Legal translation as communication of knowledge: On the creation of bridges. *Parallèles*, *33*(1), 6–17.

Flanagan, Marian, & Christensen, Tina Paulsen. (2014). Testing post-editing guidelines: How translation trainees interpret them and how to tailor them for translator training purposes. *The Interpreter and Translator Trainer*, *8*(2), 257–275. https://doi.org/10.1080/1750399X.2014.936111

Foster, Trevor, & Northrop, Seth. (2011). A lawyer's guide to source code discovery. *The Federal Lawyer*, 42–46.

Gaspari, Federico, Toral, Antonio, Naskar, Sudip, Groves, Declan, & Way, Andy. (2014). Perception vs reality: Measuring machine translation post-editing productivity. In Sharon O'Brien, Michel Simard, & Lucia Specia (Eds.), *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas* (pp. 60–72). Association for Machine Translation in the Americas.

Gémar, Jean-Claude. (2002). Le plus et le moins-disant culturel du texte juridique : Langue, culture et équevalence. *Meta*, *47*(2), 163–176. https://doi.org/10.7202/008006ar

Giannakourou, Georgia. (2006). Planning regulation, property protection, and regulatory takings in the Greek planning law. *Washington University Global Studies Law Review*, *5*(3), 535–557.

Hoey, Michael. (2005). *Lexical priming: A new theory of words and language*. Routledge.

Hu, Ke, & Cadwell, Patrick. (2016). A comparative study of post-editing guidelines. *Baltic Journal of Modern Computing*, *4*(2), 346–353.

Jia, Yanfang, Michael Carl, & Xiangling Wang. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *JoSTrans: The Journal of Specialised Translation*, *31*, 60–86.

Killman, Jeffrey. (2014). Vocabulary accuracy of statistical machine translation in the legal context. In Sharon O'Brien, Michel Simard, & Lucia Specia (Eds.), *Third Workshop on Post-Editing Technology and Practice (WPTP–3)* (pp. 85–98). AMTA.

Klubička, Filip, Toral, Antonio, & Sánchez-Cartagena, Víctor M. (2018). Quantitative fine-grained human evaluation of machine translation systems: a case study on English to Croatian. *Machine Translation*, 32, 195–215.

Koehn, Philipp, & Knowles, Rebecca. (2017). Six challenges for neural machine translation. In Thang Luong, Alexandra Birch, Graham Neubig, & Andrew Finch (Eds.), *Proceedings of the First Workshop on Neural Machine Translation* (pp. 28–39). Association for Computational Linguistics.

Kudo, Taku, & Richardson, John. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In Eduardo Blanco & Wei Lu (Eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 66–71). Association for Computational Linguistics.

Lavie, Alon. (2011, September 19). *Evaluating the output of machine translation systems* [Conference tutorial]. Machine Translation Summit XIII, Xiamen, China.

Lommel, Arle. (2018a). Augmented translation: A new approach to combining human and machine capabilities. In Janice Campbell, Alex Yanishevsky, Jennifer Doyon, & Doug Jones (Eds.), *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 2: User Track)* (pp. 5–12). Association for Machine Translation in the Americas.

Lommel, Arle. (2018b). Metrics for translation quality assessment: A case for standardising error typologies. In Joss Moorkens, Sheila Castilho, Federico Gaspari, & Steven Doherty (Eds.), *Translation quality assessment* (pp. 109–127). Springer. https://doi.org/10.1007/978-3-319-91241-7_6

Matthiesen, Aaron. (2017). *Maschinelle Übersetzung im Wandel. Die Auswirkungen von künstlicher Intelligenz auf maschinelle Übersetzungssysteme. Mit einer vergleichenden Untersuchung von Google Translate und Microsoft Translator*. epubli.

Mattila, Heikki. (2006). *Comparative legal linguistics*. Ashgate.

Matulewska, Aleksandra. (2007). *Lingua legis in translation*. Peter Lang.

Mellinger, Christopher D. (2018). Re-thinking translation quality: Revision in the digital age. *Target*, *30*(2), 310–331. https://doi.org/10.1075/target.16104.mel

Mitchell, Linda G. (2015). *Community post-editing of machine-translated user-generated content* [Doctoral dissertation]. Dublin City University.

Moorkens, Joss, Toral, Antonio, Castilho, Sheila, & Way, Andy. (2018). Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces*, *7*, 240–262. https://doi.org/10.1075/ts.18014.moo

Nelson, Sharon, & Simek, John. (2017). *Running with the machines: Artificial intelligence advances bring benefits, threats to practice of law* [White paper]. Sensei Enterprises.

Newmark, Peter. (1988). *A textbook of translation*. Prentice Hall.

Oakes, Jacob. (2016). U.S. immigration policy: Enforcement & deportation. Trump fair hearings—Systematic violations of international non-refoulement obligations regarding refugees. *North Carolina Journal of International Law*, *XLI*(4), 833–918.

Papineni, Kishore, Roukos, Salim, Ward, Todd, & Zhu, Wei-Jing. (2002). BLEU: A method for automatic evaluation of machine translation. In Pierre Isabelle, Eugene Charniak, & Dekang Lin (Eds.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311–318). Association for Computational Linguistics. https://doi.org/10.3115/1073083.1073135

Prieto Ramos, Fernando. (2015). Quality assurance in legal translation: Evaluating process, competence and product in the pursuit of adequacy. *International Journal for the Semiotics of Law – Revue internationale de Sémiotique juridique*, *28*, 11–30. https://doi.org/10.1007/s11196-014-9390-9

Prieto Ramos, Fernando. (2021). Translating legal terminology and phraseology: Between inter-systemic incongruity and multilingual harmonization. *Perspectives*, *29*(2), 175–183. https://doi.org/10.1080/0907676X.2021.1849940

Pym, Anthony. (2013). Translation skill-sets in a machine translation age. *Meta*, *58*(3), 487–503. https://doi.org/10.7202/1025047ar

Şahin, Mehmet, & Dungan, Nilgün. (2014). Translation testing and evaluation: A study on methods and needs. *Translation & Interpreting*, *6*(2), 67–90.

Šarčević, Susan. (1997). *New approach to legal translation*. Kluwer Law International.

Serraos, Kostantinos, Gianniris, Elias, & Zifou, Maria. (2005). The Greek spatial and urban planning system in the European context. In Gabriella Padovano & Cesare Blasi (Eds.), *Complessitá e Sostenibilitá, Prospettive per i territorio europei: strategie di pianificazione in dieci Paesi, Rivista bimestrale di pianificazione e progrettazione* (pp. 1–24). Poli.design.

Sosoni, Vilelmini, & O'Shea, John. (2021). Translating property law terms: An investigation of Greek notarial deeds and their English translations. *Perspectives*, *29*(2), 184–198. https://doi.org/10.1080/0907676X.2020.1797840

Sosoni, Vilelmini, Kermanidis, Katia-Lida, & Livas, Sotirios. (2018). Observing Eurolects: The case of Greek. In Laura Mori (Ed.), *Observing Eurolects: Dynamics of language variation in EU law* (pp. 170–198). John Benjamins. https://doi.org/10.1075/scl.86.08sos

Stasimioti, Maria, & Sosoni, Vilelmini. (2020). Translation vs post-editing of NMT output: Measuring effort in the English-Greek language pair. In Janice Campbell, Dmitriy Genzel, Ben Huyck, & Patricia O'Neill-Brown (Eds.), *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas, 1st Workshop on Post-Editing in Modern-Day Translation* (pp. 109–124). Association for Machine Translation in the Americas.

TAUS. (2016). *MT post-editing guidelines.*

Toral, Antonio, Wieling, Martijn, & Way, Andy. (2018). Post-editing effort of a novel with statistical and neural machine translation. *Frontiers in Digital Humanities: Digital Literary Studies*. https://doi.org/10.3389/fdigh.2018.00009

UNHCR. (2014). *Findings and recommendations relating to the 2012–2013 missions to monitor the protection screening of Mexican unaccompanied children along the U.S.-Mexico border*. Regional Office Washington, D.C. for the United States and the Caribbean.

Valeontis, Kostas, & Krimpas Panagiotis. (2014). Νομική γλώσσα, νομική ορολογία: θεωρία και πράξη [Legal language, legal terminology: Theory and practice]. Nomiki Bibliothiki.

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jalob, Jones, Llion, Gomez, Aidan. N., Kaiser, Lukasz, & Polosukhin, Illia. (2017). Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, & Roman Garnett (Eds.), *31st Conference on Neural Information Processing Systems (NIPS 2017)*. NIPS. https://doi.org/10.48550/arXiv.1706.03762

Vieira, Lucas Nunes. (2019). Post-editing of machine translation. In Minako O'Hagan (Ed.), *The Routledge handbook of translation and technology* (pp. 319–335). Routledge.

Vieira, Lucas Nunes, Alonso, Elisa, & Bywood, Lindsay. (2019). Introduction: Post-editing in practice – Process, product and networks. *JoSTrans: The Journal of Specialised Translation*, *31*, 2–13.

Vieira, Lucas Nunes, O'Hagan, Minako, & O'Sullivan, Carol. (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. https://doi.org/10.1080/1369118X.2020.1776370

Vlachopoulos, Stefanos. (2014). Πολυγλωσσία και δίκαιο [Multilingualism in the law]. Nomiki Vivliothiki.

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

Wray, Allison. (2002). Formulaic language in computer-supported communication: Theory meets reality. *Language Awareness*, *11*(2), 114–131. https://doi.org/10.1080/09658410208667050

Yates, Sarah. (2006). Scaling the tower of Babel Fish: An analysis of the machine translation of legal information. *Law Library Journal*, *98*(3), 481–500.