NATURAL VS ARTIFICIAL INTELLIGENCE AND NEURAL MACHINE TRANSLATION IN SPECIALISED TRANSLATION: A COMPARATIVE STUDY

Irina Stoyanova-Georgieva

Konstantin Preslavsky University of Shumen, Bulgaria

Abstract: *The current study aims to identify the differences between human and hybrid translation, analyse the impact of Neural Machine Translation and Artificial Intelligence on the final product, and evaluate the performance of two groups of students translating specialised texts. The first group relies on advanced technology, utilising both Neural Machine Translation and Artificial Intelligence, while the second group depends solely on natural intelligence. The results indicate that technology does not necessarily ensure quality in specialised translations. The quality assurance process shows that high-quality translation is only achieved by experienced translators who are fluent in the target language and possess a strong understanding of the subject matter. Such individuals are less prone to making inadequate translation decisions under NMT influence and are more likely to implement necessary modifications during post-editing.*

Keywords: *Machine Translation (MT), Neural Machine Translation (NMT), Post-Editing, Human Translation (HT), Hybrid Translation*

About the author: *Irina Stoyanova-Georgieva, PhD is a lecturer in the English Studies Department of Konstantin Preslavsky University of Shumen, Bulgaria. She has done a translation traineeship at the European Parliament and has a PhD thesis on the use of intensifiers in letters to the editor in British and Bulgarian newspapers and magazines. Her main interests are in the field of translation studies and translation technologies.*

e-mail: *i.stoyanova-georgieva@shu.bg* **ORCID:** *https://orcid.org/0000-0003-4065-4917*

Article history: Received: 4 May 2025; Reviewed: 13 October 2025; Revised: 1 December 2025; Accepted: 2 December 2025; Published: 19 December 2025

Citation (APA): Stoyanova-Georgieva, I. (2025). Natural vs artificial intelligence and neural machine translation in specialised translation: A comparative study. *Studies in Linguistics, Culture, and FLT, 13*(3), 135-148. https://doi.org/10.46687/XZWW6164

Introduction

Neural Machine Translation (NMT) and Artificial Intelligence (AI) are now dominating the translation industry. Today, translating any technical or administrative document is not only assisted by computer-aided translation tools but also often performed using Neural Machine Translation (NMT) and subsequently post-edited. Over the past two years, the process has become increasingly integrated with AI. Like any other technological breakthrough, society is still learning how to implement and harness it, not only in translation services but also in daily life. It is often demonised and blamed for potential future drawbacks. One such concern that has recently drawn much attention is that progress in AI will impact the employment prospects of many professionals, especially translators, leading to significant job losses. A survey conducted by the SOA POLICY TEAM found that "[O]ver a third of illustrators (37%) and over 4 in 10 translators (43%) say the income from their work has decreased in value because of generative AI" (SOA Policy Team, 2024). These findings are further supported by the study of Frey and Llanos-Paredes (2025), who report that "the adoption of Google Translate contributed to a 0.71 percentage point reduction in translator employment growth, translating into an estimated loss of more than 28,000 jobs over the 2010-2023 period" (p. 20). According to them, the situation will likely worsen as NMT technologies continue to develop. However, the increasing demand for better, faster, and more affordable translation services, driven by the need for expanded commercial communication, which has led to the improvement in MT quality and recent advancements in AI, is unlikely to reverse. Undoubtedly, the future of automation and the narrowing performance gap between humans and machines (Hassan et al., 2018; Moneus & Sahari, 2024) will prompt questions and more research.

A remark made by *The Economist* in 2023 has raised further concerns. The statement that "AI could make it less necessary to learn foreign languages" (Johnson, 2023) is even more distressing, as it reinforces technology's role as the sole means of communication in the future. In the long run, such a policy and development will most likely lead not to unlimited freedom of speech, expression, and communication, but, on the contrary, to increased control and regulation.

Pym (2024) uses a colourful illustration to highlight these concerns, suggesting that soon after people stop believing in the future of human translation, students will cease enrolling in language and translation degrees, professors will lose their jobs, and consequently, research in translation studies will come to an end.

However, for translators engaged in specialised translation, NMT and AI are tools that are here to stay. Therefore, the main concern revolves around the use and implications of these tools. As a result, the field needs further investigation into the widely publicised increases in productivity and turnover in NMT and post-editing, as well as their impact on the quality of the translation output.

Quality, serving as a marker of professional and dependable translation, remains particularly vital across different domains. For example, translation in highly specialised fields, such as legal and medical documents, still relies on human expertise (Ramos, 2015; Lee, 2023).

It is this situation that prompted the current research, which aims to examine the similarities and differences between NMT and HT approaches, as well as their quality.

Previous research in the field

Given the significance of AI for translators' work and the ever-growing apprehensions about its role in the translation industry and for translators in the future, the study contributes to a growing body of NMT and AI-oriented research. However, advances in automation led to various definitions of translation quality (Gaspari et al., 2015). Drugan (2013) highlights that there is no uniform, objective way to measure quality, but there is an abundance of error typology models that provide quantitative indicators of quality (Lommel et al., 2014). The latter categorise the traditional evaluation of quality conducted by bilingual reviewers as subjective, and find metrics developed specifically for a given translation project as unreliable. They point out that though customisation is necessary, consistency and interchangeability are desirable. In response to this variation, Lommel et al. (2014) developed MQM. MQM stands for Multi-dimensional Quality Metrics, which is based on the LISA QA Model (Dillinger & Lommel, 2004). The study defines the metrics as a flexible system that can be used to evaluate the quality of any translation (human or machine-generated) and any source text, and even to identify issues in the source text.

In their study, Lommel et al. elaborate that MQM addresses the shortcomings of previous quality assessment models and evaluates the entire project life cycle using a hierarchical listing of 114 specific issues. The upgraded 2.0 version of MQM (available at https://themqm.org/about-us/) enables detailed error analysis through a hierarchical tree structure whose categories mirror many of the quality assurance indicators in platforms like *Phrase* and *Trados*. MQM 2.0 has eight categories: terminology, accuracy, linguistic conventions, style, locale conventions, audience appropriateness, design and markup, and custom, which list 127 specific issues. Among them, linguistic conventions, design, and markup appear to account for the most issues. As Lommel et al. (2014) observed, linguistic issues and formatting seem to be central to the model.

The model's accuracy and adaptability make it particularly suitable for the current study, enabling a well-grounded evaluation.

As it was mentioned, due to the significance of the topic, there are many studies in the field of human vs AI and/ or NMT and Large Language Models discussing the results of the use of AI for specialised translation or even fiction (He et al., 2024; Moneus & Sahari, 2024; Alkhofi, 2025; Awashreh & Aboeisheh, 2025; Martos et al., 2025; Doan, 2025; Nedelcheva, 2025). Moneus and Sahari's 2024 study is comparable to the current one and also evaluates the quality of AI translations, identifying differences between AI-generated and human translations. However, the primary concern of their research is identifying differences between human translation results and those of generative AI (ChatGPT, ChatSonic). Therefore, it does not compare human translation from scratch with human post-edited NMT and AI, a crucial consideration for the future of translation that the industry should consider, given its impact on quality. The current study tests whether post-edited NMT, further backed by AI (i.e., hybrid translation), has the same high quality as human translation and under what circumstances. Another dissimilarity between this study and the other recent studies is the choice of domain. The majority of recent research discusses legal text translation, which has long been shaped by requirements for human intervention, understanding, and legal expertise (Lee, 2023).

The results of Moneus and Sahari's (2024) study indicate that while AI translation is undeniably faster and more cost-effective, it cannot deliver a human-quality product as a stand-alone tool. The reason is that specific contexts prevent it from properly reflecting legal terminology, metaphors, idiomatic expressions, and cultural nuances. Therefore, it has limitations that human translation does not.

Research Questions

Based on this, the current paper compares the translations of two groups of students training to become translators. The first group translates specialised texts using *Phrase,* a cutting-edge translation and localisation platform that streamlines the process by harnessing NMT and AI. In the company's words, *Phrase* "automates the selection of the most effective translation services, incorporates brand terms, and scores content quality based on an array of context and circumstances" (Phrase, 2025). While the first group utilises technology, the second translates from scratch using only a word processor.

The study thus aims to identify differences between human and hybrid translation, analyse the effects of NMT and AI on the final product, assess the performance of both groups, and evaluate the need for post-editing in specialised translation.

Unfortunately, the nature of the study poses some limitations. As the participants in one of the groups complete their projects without the use of any translation technology tool, the duration of the translation and revision stages of their projects cannot be measured. This restricts the ability to compare human and hybrid translation in terms of productivity.

Methods

To achieve this, the study relies on a corpus of 33,140 words compiled from the translations of five short academic texts (publication guidelines for two journals, a preface and an *About Us* section for a journal, and a call for papers). These were translated by twenty third- or fourth-year students specialising in either English Studies or Applied Linguistics. All texts followed a translator–reviser workflow model. In it, depending on the translation type (human or hybrid) the first student served as either the translator or post-editor, while the second student had to revise. All projects were to be completed and delivered within seven days.

The domain, texts, their length, the number of participants in the experiment, and the due date were determined by the client, an internal body of the higher education institution. As a result, the translated texts provide a clear picture of the complexity and diversity of the language in specialised texts and offer a glimpse into the translation procedures used by the two groups, thereby revealing their approaches to the texts. According to House (1997), "different views of translation lead to different concepts of translational quality, and hence different ways of assessing quality" (p. 1). Thus, the criteria for translation quality evaluation differ in different theories. Here, however, quality assurance was conducted through reviews and proofreadings of the texts, executed independently by two university professors using MQM.

For the purposes of the study, the term quality assurance is defined as "systems and processes used to help create or maintain quality" (Saldanha & O'Brien, 2013, p. 95). It is formed through quality control and quality assessment (Mossop, 2020). Ramos's (2015) study presents an excellent example of a quality evaluation methodology. It offers criteria for evaluating texts, which he defines as suitable for human translation or for AI-generated translation. It addresses semantic accuracy and consistency, the adequacy of translation decisions regarding terminology, phraseology, genre conventions, the microtextual level, and the cohesion of the text, its syntax, punctuation, etc., all united under the general linguistic correctness. It also provides a grading scheme, which defines the results:

Excellent (A/5)	Maximum accuracy and consistency, adequate decisions according to the legal conditions and communicative situation; no linguistic errors
Acceptable (B/4)	Only some minor inaccuracy, inconsistency, inadequate decision or linguistic error not affecting main functions or microtextual priorities.
Borderline (C/3)	Inadequate decisions hinder main functions or microtextual priorities; significant linguistic error or several minor ones (e.g. punctuation problems)
Poor (D/2)	Major problems of accuracy, consistency, adequacy or linguistic correctness even if the text is readable
Unacceptable (E/1)	Inaccurate content, systematically inadequate decision-making and serious linguistic errors

(Ramos, 2015, p. 25).

However, the indicators used to assess the quality of the translation as excellent, acceptable, borderline, poor or unacceptable seem vague. Issues are not explained in detail, leading to decisions that seem subjective and based on overall impressions of the text.

To this end, the study adopts MQM, which enables a more precise assessment of specific translation issues, and then uses Ramos's grading scheme to define the results.

Following the aforementioned procedure, each translation is submitted for quality assurance in accordance with the MQM typology and receives a mark.

Lastly, the study also draws on another typology, Pym's translation solutions for many languages (2016). It seeks to describe the translation procedures used by participants in the two groups, to examine differences between all-human and hybrid translation, and to provide an explanation for the results. The decision to use a translation procedure analysis was prompted by Al-Qinai's statement that the "tendency to ignore the process of decision-making lies behind the lack of objectivity in translation assessment" (Al-Qinai, 2000, p. 497).

Cruise mode (normal use of language skills, reference resources, parallel texts, intuition – anything prior to bump mode – so no special solutions are needed)

Copying	Copying Words	Copying Sounds Copying Morphology Copying Script …
	Copying Structure	Copying Prosodic Features Copying Fixed Phrases Copying Text Structure …
Expression Change	Perspective Change	Changing Sentence Focus Changing Semantic Focus Changing Voice …
	Density Change	Generalization/Specification Explicitation/Implicitation Multiple Translation Resegmentation …
	Compensation	New Level of Expression New Place in Text (notes, paratexts) …
	Cultural Correspondence	Corresponding Idioms Corresponding Culture-Specific Items …
Content Change	Text Tailoring	Correction/Censorship/Updating Omission of Content Addition of Content …

Pym's typology (2016, p. 220)

Data Analysis

The first part of the analysis looks at the texts generated, post-edited, and further revised in *Phrase*. Since the study participants had no prior experience with such texts and no translation memories were available for the project, all segments were machine-translated and then post-edited. The auto-select tool in

Phrase, which chooses the best NMT engine based on each job's domain and language pair, supports nine NMT engines: *Amazon Translate, DeepL, Google Translate, Microsoft Translator, Phrase Next GenMT, Phrase NextMT, Rozetta Translate, Tencent,* and *Widn.AI*. However, *DeepL* was the most frequently used source for the majority of the segments.

The analysis shows that three of the projects can be classified as acceptable according to Ramos (2015). The individual jobs (using *Phrase* terminology) in these projects have only minor inaccuracies, inconsistencies, and inadequate decisions. MQM highlights errors that can be categorised as issues in terminology, accuracy, linguistic conventions, style, or audience appropriateness.

Wrong terms are rarely encountered. Some projects used *summary* instead of *abstract* for a scientific paper, and *society* instead of *association* for the Association of Writers. Others used *Technical Requirements* where *Publication Guidelines* should have been used, and *Basic Steps* instead of *Submissions Basics* or *Technical Requirements*. *Departmental editor/ department editor/ Editor-in-Chief* appeared instead of editor. The situation is similar to the *Literature Section,* which is generally referred to as *References*. Further example is *the Unified format of electronic publications*, which should have been translated as *Submission Guidelines*. The analysis indicates that these examples are clearly produced by copying the structure of the original term.

Issues with accuracy were also common in these projects. They were represented by the overtranslation of Bulgarian book titles. Ivelina Savova's book „Съвременни графити (лингвистичен аспект)" was turned into *Modern graffiti (linguistic aspect)* instead of being transliterated as *Savremenni grafiti (lingvistichen aspect)*. The same happened with several other books whose titles were also translated, where copying the sounds should have been used: *Savova, I., Dobreva, Sn. Bulgarian syntax. Learning aid for students.; Popova, V. 2017a. The early ontogeny of event modality.; Bosilkov, K. 1981. Interaction between the traditional and the new in the early stage of the formation of the New Bulgarian literary language.*

Such errors result from automatic MT-generated translation; however, due to the limited experience of the translator and reviewer alike, the issues were neither identified nor corrected.

Another type of accuracy issue involves the translation of entity names that should remain in Bulgarian. The analysis found that the name of *Шуменски университет „Епископ Константин Преславски"* was translated on the institution's letterhead and logo, which changed them.

The analysis identified some punctuation errors, as students followed the source text's punctuation rather than the target language conventions. However, these errors were not deemed crucial to understanding the texts.

The most common issue across all the projects was white spaces. Leading and trailing spaces (using *Phrase* terminology) appeared in projects of both acceptable and borderline quality. They were likely the result of the students' insufficient experience with *Phrase* and their habit of translating from scratch using word processors or pen and paper.

Finally, some sentences, even in projects with acceptable quality, had an awkward, unidiomatic style and language-dependent logic. The translation, although post-edited, still retained too much of the MT-generated literal translation of the source segment:

> Текстовете за отпечатване трябва да се предадат най-късно до 1 септември 2024г. на CD или по e-mail. – Texts that are to be printed must be submitted by September 1, 2024. on CD or by e-mail.

The latter are typically caused by copying words or structures from the source text, when cultural correspondence or a change in perspective would have been more suitable. All these mistakes were generated and inherited from the NMT and later approved by the post-editor, who lacked sufficient experience and was misled by the NMT and AI results.

Not surprisingly, none of the projects exhibited any significant formatting issues. Thanks to the advanced software and quality assurance tools within the translation environment, even instances of missing tags did not impact the formatting of the target text.

However, despite using the same resources, the other two projects were rated as borderline, according to Ramos's (2015) classification. The analysis shows that the projects in this group exhibit all the errors found in the previous group, but their frequency and number are significantly higher. In addition to the higher percentage of the aforementioned terminology, accuracy, linguistic conventions, and stylistic issues, these projects also have additional problems: omissions and a greater number of character formatting errors.

For example, *библиография* is translated into the target text as *bibliography* rather than as *references*. Multiple mistakes were found in the name of the institution, which is translated instead of being transliterated or transcribed: „Еп. Константин Преславски" – "Ep. "BISHOP KONSTANTIN PRESLAVSKI". Even more concerning is that the title of the University's patron was repeated twice in consecutive segments, first transliterated and then translated. Many segments exhibit an awkward, unidiomatic style due to the overly literal translation of the source text:

> списание – разликата спрямо правилата при книга е, че с курсив се изписва наименованието на списанието, броят и съответните страници на цитираната публикация. – journal - the difference from

> the rules for a book is that the name of the journal, the number and the corresponding pages of the cited publication are written in italics.

The array of minor and major linguistic errors, coupled with poor translation decisions, creates difficulties for the reader and hinders the main functions of the text.

While some low-quality results can be explained by a lack of extensive experience in a translation environment, the rest cannot be solely attributed to this factor. An interesting fact is that, despite all students using the same resources, some projects were deemed acceptable, while others were classified as borderline, according to Ramos's (2015) categorisation. A plausible explanation may be the competency of the translators, post-editors, and reviewers involved in the study. A more detailed analysis of the participants reveals that the teams responsible for delivering high-quality translations consisted of two types of students. While the MT post-editing was performed by less experienced students whose competence could be assessed as that of independent users according to the Common European Framework of Reference (CEFR), the revision was carried out by more experienced users. The latter are qualified as proficient users. The teams that delivered projects assessed as borderline included inexperienced translators and reviewers whose competence was equivalent to B2 according to the CEFR. Therefore, participants in the second group were more inclined to rely on NMT rather than their own experience and knowledge.

The second stage of the analysis examines texts translated by humans from scratch, using only a word processor.

The results of this stage, however, reveal a similar division. Namely, the group of inexperienced and less skilled translators and revisers relied on translation procedures such as copying words and copying structure. As a result, sections of the texts were almost identical to those generated by the NMT in the first phase of the project. This suggests that the translators in this group relied on the same translation procedures as NMT. Consequently, they made similar mistakes to those detected in the projects of inexperienced post-editors.

As a result, the most commonly used translation procedure among the students is copying structure, i.e., literal or word-for-word translation (Vinay & Darbelnet, 1995), even though perspective change, cultural correspondence, or density change would have been more appropriate translation procedures. For example:

> Целта на списание „ЛитерМедия“ в самото начало е да популяризира литературоведски и медийни научни изследвания.

> The aim of ''LiterMedia'' magazine at the very beginning was the popularization of scientific researches related to literacy and media.

and

…в огледалото на Запада и междуславянското общуване.

…through the lens of the West and inter-Slavic communication.

…който си постави амбицията да бъде калейдоскопичното лице на живота във ФХН, като онагледява и озвучава този живот.

…which aspired to be a kaleidoscopic reflexion of life within the Faculty of Humanities, by illustrating it visually and auditorily.

Similar to the previous set of projects, these also involve issues related to terminology and linguistic conventions. In them, alternative words were used instead of the approved terms: *volume/magazine* instead of *journal*, *thematic cores* instead of *topics, Slavdom* instead of *Slavism*. Punctuation was omitted or incorrect, and there were spelling issues (see the previous examples): *dedline* (deadline) and *turkologists* (Turkologists), *reflexion* (reflection), *lenght* (length), etc.

In contrast, projects whose translators and/or reviewers were proficient in the language were rated as having acceptable quality. The proportion of terminological, accuracy, linguistic conventions, and style issues found in these projects was similar to that in the texts translated with *Phrase* using AI and NMT. What sets the acceptable-quality projects produced in the two phases of the study apart is the markedly higher number of character formatting issues in those translated from scratch.

The final indicator examined in the study is editing time. Editing time is frequently cited as a key factor in the use of NMT in translation over recent years (Ramos, 2015; Choudhury & McConnell, 2014; Mossop, 2007; Moneus & Sahari, 2024). However, reliable information about this indicator can only be obtained for projects generated through *Phrase* TMS, where the system records editing time in seconds. The analysis of this data shows that revisers spent between 16 and 133 minutes working on all four files in the project, which totalled 13365 characters/ 2145 words/ 8.51 pages.

Results

The results of the first phase of the study show that, although all participants used the same resources and the same NMT engines provided by *Phrase*, the quality of the translations varies from acceptable to borderline. These findings can only be explained by the experience of the translators acting as post-editors or reviewers of the project. As a result, it highlights the importance of humans in ensuring translation quality. A human translator or reviser understands the meaning of the text, its purpose, cultural nuances, and specific writing style. Even more importantly, humans consider the target readership, which allows for accurate communication not only of the message but also of its tone. Occasional inaccuracies in NMT would not mislead a qualified translator, post-editor, or reviser, who would make the necessary adjustments to produce a high-quality translation.

The second stage of the project offers insights into the parallels between the NMT approach and that of linguists. It indicates that experienced linguists utilise a range of translation solutions. Meanwhile, translators lacking training and language proficiency predominantly depend on copying mechanisms, such as copying words and structure from the source document. The aforementioned approach aligns with NMT's method in these cases. This can be seen as a demonstration not only of the progress of NMT technology but also of its time and cost efficiency.

The final metrics analysed in the study relate to the time spent on the project. Editing time, however, can only be obtained for NMT and AI-assisted projects. Information about the total time spent on both the translation and revision stages of human translation would have provided a clearer understanding of the results and revealed any connection between time and quality. Nevertheless, despite the lack of data for all projects, the analysis indicates that students with limited experience in the field can produce an acceptable-quality translation of five texts, totalling eight and a half pages, in less than two and a half hours.

Conclusion

The current study once again demonstrates that the gap between AI, NMT engines, and human translation is narrowing, and that the technology can readily replace less experienced translators. With advances in AI and its increased use to improve NMT, the gap is undoubtedly closing.

However, the results of the current study show that cutting-edge technology does not guarantee quality in specialised translations. Quality assurance proves that translation quality is maintained only by experienced translators with excellent command of the foreign language and a strong understanding of the subject matter. Such individuals are less susceptible to poor translation decisions made by NMT and are more likely to make necessary changes in documents. They rely more on perspective and density adjustments as well as on cultural correspondence rather than simply copying words and structure from the source document. However, combining NMT, AI, and linguists' post-editing with extensive experience streamlines translation quality, enabling higher-quality translations in less time.

The results of the study show that language training should not become obsolete, but future translators must be trained to work with NMT and AI to deliver high-quality post-editing. Although specialised translation across various fields still requires human translators to ensure quality, translators should be encouraged to develop skills that machines cannot replicate. Such skills and services demand a solid understanding of source and target cultures, traditions, and specificities, as well as creativity and domain-specific knowledge that remain beyond the reach of technology.

In conclusion, despite advances in NMT and AI, the disparity between human and machine results remains significant, and the human factor is indispensable for delivering quality translations. Though NMT and AI are here to stay, language education and translator training are not to be underestimated.

References

Alkhofi, A. (2025). Man vs. machine: Can AI outperform ESL student translations? *Frontiers in Artificial Intelligence*, *8*. https://doi.org/10.3389/frai.2025.1624754

Al-Qinai, J. (2000). Translation quality assessment. Strategies, parameters and procedures. *Meta: Translators' Journal, 45*(3), 497–519. https://doi.org/10.7202/001878ar

Awashreh, R., & Aboeisheh, A. (2025). The collaborative future of translation between human-ai partnerships. *Advances in Computational Intelligence and Robotics*, 205–236. https://doi.org/10.4018/979-8-3373-0060-3.ch008

Brunette, L. (2000). Towards a terminology for translation quality assessment. *The Translator*, *6*(2), 169–182. https://doi.org/10.1080/13556509.2000.10799064

Choudhury, R., & McConnell, B. (2013). Translation technology landscape report, De Rijp, The Netherlands: Translation Automation Users Society (TAUS).

Common European Framework of Reference for Languages. (2001). The CEFR Levels. Retrieved 10/10/2025, from https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Dillinger, M. & Lommel, A., (2004). *LISA Best Practice Guide: Implementing Machine Translation*. Geneva: Localization Industry Standards Association.

Doan, T. T. (2025). Assessment of human translation vs. AI translation in a literary work. *VNU Journal of Foreign Studies*, *41*(3), 126–141. https://doi.org/10.63023/2525-2445/jfs.ulis.5513

Drugan, J. (2013). *Quality in professional translation: assessment and improvement*. Bloomsbury, London

Frey, C. B., & Llanos-Paredes, P. (2025). *Lost in Translation: Artificial Intelligence and the Demand for Foreign Language Skills*. Oxford Martin School Working Paper.

Gaspari, F., Almaghout, H., & Doherty, S. (2015). A survey of machine translation competences: Insights for Translation Technology Educators and practitioners. *Perspectives*, *23*(3), 333– 358. https://doi.org/10.1080/0907676x.2014.979842

Gotti, M., & Sarcevic, S. (2006). *Insights into Specialized Translation*. Lausanne, Switzerland: Peter Lang Verlag. https://doi.org/10.3726/978-3-0351-0411-0

Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., Huang, X., Junczys-Dowmunt, M., Lewis, W., Li, M., Liu, S., Liu, T., Luo, R., Menezes, A., Qin, T., Seide, F., Tan, X., Tian, F., Wu, L., . . .& Zhou, M. (2018). Achieving human parity in automatic Chinese to English news translation. *arXiv preprint arXiv:1803.05567.*

He, Z., Liang, T., Jiao, W., Zhang, Z., Yang, Y., Wang, R., Tu, Z., Shi, S., & Wang, X. (2024). Exploring human-like translation strategy with large language models. *Transactions of the Association for Computational Linguistics*, *12*, 229–246. https://doi.org/10.1162/tacl_a_00642

House, J. (1997). *Translation quality assessment: A model revisited*. Tübingen: Gunter Narr Verlag.

Johnson. (2023, August 17). AI could make it less necessary to learn foreign languages. *The Economist.* https://www.economist.com/culture/2023/08/17/ai-could-make-it-less-necessary-to-learn-foreign-languages

Lauscher, S. (2000). Translation quality assessment: Where can theory and practice meet? *The Translator 6*(2), 149–168. https://doi.org/10.1080/13556509.2000.10799063

Lee, T. B. (2023, July 5). *How human translators are coping with competition from powerful AI. Understanding AI.* https://www.understandingai.org/p/how-human-translators-are-coping

Localization Industry Standards Association. (2004). Lisa - Best Practice Guide. https://ot2009.files.wordpress.com/2009/05/5-lisa-best-practice-guide.pdf

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multi-dimensional Quality Metrics (MQM): A Framework for declaring and describing translation quality metrics. *Tradumàtica Tecnologies de La Traducció*, (12), 455–463. https://doi.org/10.5565/rev/tradumatica.77

Martos, M., Fields, B., Finlayson, S. G., Hartell, N., Kim, T., Larimer, E., Lau, J. J., Lin, Y.-H., Salaguinto, T., Tran, N., & Lion, K. C. (2025). Accuracy of Artificial Intelligence vs professionally translated discharge instructions. *JAMA Network Open*, *8*(9). https://doi.org/10.1001/jamanetworkopen.2025.32312

Moneus, A. M., & Sahari, Y. (2024). Artificial Intelligence and human translation: A contrastive study based on legal texts. *Heliyon*, *10*(6). https://doi.org/10.1016/j.heliyon.2024.e28106

Mossop, B. (2007). Empirical studies of revision: What we know and need to know. *The Journal of Specialised Translation*, (8), 5–20. https://doi.org/10.26034/cm.jostrans.2007.695

Mossop, B. (2020). *Revising and editing for translators* (4th ed.). London: Routledge. https://doi.org/10.4324/9781315158990

MQM (multi-dimensional quality metrics). MQM (Multi-dimensional Quality Metrics). (n.d.). https://themqm.org/the-mqm-full-typology/

Nedelcheva, S. (2025). *Natural vs. Artificial Intelligence in Translating Stylistic Flair and Culture Specifics of Fiction Texts: A Case Study of Georgi Gospodinov's Time Shelter*. Konstantin Preslavsky University Press. ISBN 978-619-201-891-7

Phrase. (2025, September 25). Phrase. https://phrase.com

Pym, A. (2016). *Translation solutions for many languages: Histories of a flawed dream* (1st ed.). Bloomsbury Academic.

Pym, A. (2024). On the end of translation studies as we know it. *XII International Scientific Conference Major Problems of Translation Studies and Translator/ Interpreter Training*. V. N. Karazin Kharkiv National University, Ukraine, 19-20 April 2024.

Ramos, P F. (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

Saldanha, G., & O'Brien, Sh. (2013). *Research methodologies in translation studies*. Manchester: St. Jerome Publishing.

SOA POLICY TEAM (2024, April 11). *SOA survey reveals a third of translators and quarter of illustrators losing work to ai*. The Society of Authors. https://societyofauthors.org/2024/04/11/soa-survey-reveals-a-third-of-translators-and-quarter-of-illustrators-losing-work-to-ai/

Vinay, J. P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: A methodology for translation* (J. C. Sager & M.-J. Hamel, Trans.). Amsterdam/Philadelphia: John Benjamins Publishing Company. (Original work published 1958). https://doi.org/10.1075/btl.11