Embracing New Translation Technologies at International Organizations:

Translators' Perceptions of Machine Translation and its Impact on Institutional Translation

*Fernando Prieto Ramos**

Abstract

This study presents the results of a large-scale survey of institutional translation professionals on the distinctive features of post-editing (as compared to revision in particular), the main issues found in neural machine translation (NMT) and the impact of its use on translation processes, products and competences. The quantitative and qualitative findings illustrate the diversification of tools and inputs in current translation workflows, the risks and efforts associated with NMT, and the reinforced need for solid translation competence and subject matter knowledge to ensure translation quality. The (mostly limited) variations found between institutions and translators' experience levels are also discussed. The evidence gathered supports a revised concept of translation that considers an increasingly hybrid and dynamic mix of translational actions in augmented translation environments. In turn, these results highlight the persistent relevance of translation expertise and specialized translator training, and contribute to nuancing our understanding of the implications of NMT for professional translation, beyond institutional settings.

---

* *Fernando Prieto Ramos*: Centre for Legal and Institutional Translation Studies (Transius), Faculty of Translation and Interpreting, University of Geneva, Fernando.Prieto@unige.ch. This article is based on research presented at the 2024 conference on "Legal Translation in the European Union: What are the Challenges in Light of AI?", organized by EUR FRAPP, ISIT and the University of Torino.

1. Introduction

As in other areas of professional translation, the recent advances in neural machine translation (NMT) systems have led to a tangible increase in the use of these technologies as part of the daily work of translation services of international organizations (IOs). Our recent examination of patterns of human-machine interaction in these settings (Prieto Ramos, 2024a) revealed a landscape of "augmented translation" (as coined by Lommel, 2018) in which translation memories (TMs) are the most popular tool and MT suggestions are integrated into translation assignments approximately two thirds of the time, while from-scratch translation is becoming limited to a minority of cases.

The patterns point to tool relevance, risks and efficiency considerations for each text type and purpose as the main factors to activate the MT add-on in each case, rather than based on attitudes or preferences connected to translators' backgrounds or experience levels, or as a result of other institutional specificities. For example, as in the case of legal translation, which is the main area of specialization and paradigm in institutional settings, high consistency with previously translated texts and discourse conventions (including, crucially, established terminology) is a key quality requirement in institutional translation, and this explains the need to verify the relevance and reliability of the sources of translations being reused. These sources are directly provided in TM matches, but not in MT suggestions, which justifies the prevalence of TMs as a priority tool for recycling previously rendered translations in institutional contexts (Prieto Ramos, 2024a).

This study delves into more specific aspects of technology use surveyed within the same project,[1] in particular, how institutional translation professionals perceive: (1) the distinctive features of editing MT suggestions (i.e. post-editing (PE)) as compared to revising translations produced or validated by human translators or retrieved through TMs; (2) the main issues found in MT output; and (3) the impact of MT use on working processes, translation product quality and translation competences required. The results in each case will be presented and discussed in the following sections, after a focused review of previous studies and a methodological outline. The main goal is to describe current perceptions of NMT to refine our understanding of how technology-assisted processes are embedded in institutional translation production, and thereby contribute empirically to the ongoing debate on how the concept of translation and the role of translators are reshaped by these changes (e.g. Daems & Macken, 2021; Do Carmo & Moorkens 2020; Olohan 2021).

---

[1] "Legal Translation in International Institutional Settings: Scope, Strategies and Quality Markers", led by the author and supported through a Consolidator Grant (transius.unige.ch/letrint/).

2. Surveying Perceptions of MT in IOs

Multilingual IOs are an ideal observation deck for examining the impact of technological changes on translation processes, competences and products, the three main aspects of translation quality analyzed in our project (see holistic framework in Prieto Ramos, 2015). These institutions are at the frontline of integrating and tailoring translation tools to their needs, which encompass a wide range of text types requiring high quality and specialized translation skills. They must ensure the reliability and continuity of institutional communication, while meeting cost optimization and efficiency requirements with the support of changing technologies (see e.g. Strandvik, 2020). The stability of text production, quality expectations and in-house professional profiles, including diverse age groups in the case of medium-sized or large translation services, makes institutional translation particularly suited for the analysis of technological impacts.

Most research on translators' perceptions of MT and PE in institutional settings thus far has focused on the largest EU institution, the European Commission (EC) and its Directorate-General for Translation (DGT). Cadwell et al. (2016), based on a focus group study of 70 translators of all 24 DGT language departments, and Rossi and Chevrot (2019), through a survey of 89 translators of 15 departments, reported the factors determining the perceived usability of MT@EC, the EU's statistical MT (SMT) system gradually replaced by the current NMT-based eTranslation since late 2017. Time savings or productivity gains stood out as the main reason to adopt MT while quality issues were the main negative factor. The perception of MT is very much conditioned by MT performance, which changed dramatically for some languages with the advent of NMT. Macken et al. (2020) compared the views of 12 DGT translators using SMT from English into French and eight using NMT from English into Finnish, and the latter provided more consistent benefits, although the difference in speed gain was limited overall (14% for NMT as opposed to 12% for SMT).

The reduction of typing efforts and having an initial version to start with and inspire the translation process were mentioned as positive aspects in these studies, and were also mentioned in the interviews conducted by Ágnes Lesznyák with a group of 38 Hungarian language translators to assess eTranslation after using it for 8 to 14 months, mainly when translating from English (Lesznyák, 2019). The perceptions were extremely disparate, with most of the critical views connected to the unreliability of MT output and recurring errors such as mistranslation, incorrect or inconsistent terminology and inappropriate style. These issues were also among the error types found in an analysis of eTranslation output from English into Slovene by Arnejšek and Unk (2020). They found that, despite the progress of NMT when compared to SMT, none of the ten sampled documents pre-translated with eTranslation were fit for purpose and required full PE. In the case of English-Polish translation, based on the analysis of 48 texts translated by eTranslation, Karolina Stefaniak found that productivity gains were modest, but risks

for quality were significant as "less than 20% of NMT segments did not contain any errors; and most of the mistakes in the remaining segments were mistakes in accuracy or terminology" (Stefaniak, 2020: 268).

In a more comprehensive study of the terminology of 34 documents translated by eTranslation from English into the other 23 EU official languages, Stefaniak (2023) reported a high degree of accuracy variability between languages, texts and terms, corroborating the unpredictability of eTranslation. PE interventions to fix NMT issues at DGT have also been the subject of analyses, e.g. in Vardaro et al. (2019), on the effects of NMT error types on eye movements among 30 translators from the German Department; and Vega Expósito (2024), on the usability of English-Spanish eTranslation suggestions among DGT translators. The eTranslation proposals used in translating 406 documents reached 30% of word volume as opposed to 50% in the case of TM matches.

Our study aims to fill a gap by examining perceptions of NMT and its impact on institutional translation across IOs, and not only DGT and EU institutions. It takes stock of trends emerging from the widespread adoption of NMT between the late 2010s and 2023, i.e. after the initial period of workflow adaptations in most organizations and the gradual NMT developments and users' familiarization with several MT systems. These include both institution-specific engines (e.g. eTranslation in the EU institutions and WIPO Translate at the World Intellectual Property Organization (WIPO)) and other commercially available NMT systems (e.g. DeepL or Google Translate). A comprehensive approach to stock-taking (i.e. comprising a myriad of institutions, text types and official languages, as well as multiple tools) has become urgently needed not only to gain a better understanding of ongoing technology-driven changes across the board, but also to nuance the "hype" about NMT with the actual expressed views of translation professionals.

To this end, we designed a survey drawing on preliminary analyses, including two rounds of structured interviews between 2017 and 2021,[2] additional exchanges with translation professionals working in or for IOs, and a literature review of the latest developments in NMT and PE. The survey was circulated with the support of the world's largest network of institutional language services, the International Annual Meeting on Language Arrangements, Documentation and Publications (IAMLADP), between late 2022 and February 2023. The survey collected anonymized quantitative and qualitative data on the respondents' profiles (including years of institutional translation experience), working methods and perceptions of the impact of NMT on institutional translation. 474 translation professionals from 24 IOs, including EU institutions and intergovernmental organizations (IGOs) with translation services of various sizes, participated in the survey, with the largest ones (the EC, the European Parliament (EP), the Council of the EU (COEU), the Court of Justice of the EU (CJEU) and the United Nations (UN))

---

[2] 33 structured interviews with 45 service heads and quality advisers between 2017 and 2019, on all aspects of quality assurance and the use of tools; and five additional interviews with translation managers between 2021 and 2022, with a focus on the latest workflow and technological developments.

contributing the highest numbers of respondents. A cohort of 31 translators working exclusively for IOs onsite or remotely, as temporary staff or external service providers, also participated in the survey as "freelancers", a heterogenous group whose increasing involvement in institutional translation could not be ignored.

Participants were assigned a unique alphanumeric identifier. They were asked to express their level of agreement with a number of statements and to provide comments in response to several open-ended questions. These replies were scrutinized drawing on keyword analysis to identify the most common views or arguments, which in turn added qualitative nuance to the patterns based on quantitative data. The components and results of the survey are discussed in detail in the next three sections.

3. The Integration of PE in Translation Processes

The first question examined in the study is how translation professionals perceive the integration of post-editing MT[3] suggestions in the translation process, as compared with editing TM matches or revising peers' translations. The distinctive features and overlaps of translation, revision and PE competences are the subject of ongoing debate. The increasing hybridization of translation processes in augmented translation environments, with diverse machine-generated inputs, are leading to the "partial confluence of translation, revision and post-editing tasks" (Konttinen et al., 2021: 188). Even if specific competence descriptors have been put forward for revision and PE (e.g. Nitzke et al., 2019; Rico & Torrejón, 2012; Robert et al., 2017; Scocchera, 2017), they inevitably overlap and rely on the common core competence required to verify and improve translation adequacy and target text quality, i.e. translation methodological (or strategic or decision-making) competence. As noted by Konttinen et al. (2021: 202), "the need to make corrections and improvements in translated texts, whether the text is a human translation, a machine translation or a fuzzy match from a translation memory", will remain the same.

Indeed, revision and PE share the need to detect and fix quality issues in a translation draft. The main difference between these activities, i.e. the human or machine origin of the draft, can be blurred; and editing or rewriting segments as part of both revising and PE is close to "classical" translation competence (see also Do Carmo & Moorkens, 2021; Robert et al., 2023; Teixeira, 2014). At the same time, editing and validating TM fuzzy matches entails revising or adapting text segments previously translated or validated by human professionals, with or without machine-generated inputs (see also Sánchez-Gijón et al., 2019). This previous human validation, together with the fact that the source

---

[3] The use of large language models (LLMs) for translation, albeit reported as rare or only experimental in some of the institutions considered, was also included in the concept of MT for the purposes of the survey.

of the re-used translation segment is typically retrieved by the TM, explains why editing TM segments can be associated with revision, even if ISO 18587:2017 on MT PE also considers editing and correcting TM content as PE. ISO 11669:2024 on translation project management also suggests that the distinction between these tasks is increasingly blurred, and that the uptake of specific technological tools will depend on their reliability and the quality required in each assignment.

This is also what emerged from our preliminary analyses. Translation managers and staff were unanimous about the flexible approach to using various resources as long as the final product conforms to the relevant institutional conventions and expectations. MT is one of such tools used (or not) in conjunction with other resources depending on the translator's expert judgment of the potential efficiency gains and risks involved in each case. PE was thus viewed as a routine action embedded in the CAT process, rather than a working method systematically adopted for entire pre-translated texts. MT inputs have enriched the existing patterns of automation based on TM use, and have added to the more diversified skillset of today's translators, who remain "in the loop". The fact that translations produced using MT are also revised (i.e. "other-revision" of full texts translated by human translators), especially when high quality must be assured, also illustrates the nature of PE as a translational action together with TM match editing.

The integration of PE, together with revision skills, into the core translation methodological or strategic competence was validated by almost all respondents[4] as part of a broader descriptor of institutional translator competence components, according to the following definition: "Strategic or methodological competence controls the application of the other competences and comprises: analysis of translation briefs and purposes, identifying and solving translation problems and, more broadly, implementing strategies or methods to achieve communicatively adequate translations, self-revise and justify translation decisions, and revise or post-edit texts to ensure the expected translation quality" (Prieto Ramos, 2024b: 255).

In order to gain more insights into the distinctive features of PE, respondents were also asked to what extent they considered that validating or revising human translations or TM matches and MT system suggestions were similar processes calling for essentially the same core translation methodological competence to ensure translation quality (i.e. reading the source text and the proposed translations, and introducing changes or rewriting segments to correct errors or improve the quality of the text, even if the quality of human translations and that of automated inputs may differ substantially). They could choose between three levels of similarity between PE and revision: "very similar", "partially similar" and "not similar at all"; and they were invited to comment on the differences.

[4] Only 2.11% of respondents stated that they would express some of the competence components differently.

Table 1: Level of similarity between PE of MT suggestions and revision of human translations and TM matches per institution and among freelancers

	Very similar	Partially similar	Not similar at all
European Commission (EC)	7.55%	88.68%	3.77%
European Parliament (EP)	15.09%	73.58%	11.32%
Council of the EU (COEU)	36.00%	49.33%	14.67%
Court of Justice of the EU (CJEU)	5.97%	92.54%	1.49%
Other EU institutions[5]	36.59%	54.88%	8.54%
Overall – EU institutions	**22.12%**	**69.70%**	**8.18%**
United Nations (UN)	15.38%	84.62%	-
World Trade Organization (WTO)	16.67%	83.33%	-
Other IGOs[6]	26.67%	68.89%	4.44%
Overall – IGOs	**20.59%**	**77.45%**	**1.96%**
Freelancers (several inst.)	**30.00%**	**66.67%**	**3.33%**
TOTAL	**22.29%**	**71.21%**	**6.49%**

---

[5] For the sake of simplicity, each breakdown of results per institution refers to the most statistically-significant groups of survey participants, which correspond to the largest translation services examined, while "other EU institutions" and "other IGOs" comprise institutions that contributed with fewer participants, also generally in line with the sizes of their respective language services. "Other EU institutions" include: the European Central Bank (ECB), the European Court of Auditors, the European Economic and Social Committee (EESC) and the European Committee of the Regions (CoR), the European Investment Bank (EIB) and the Translation Centre for the Bodies of the European Union (CdT).

[6] "Other IGOs" include: the African Development Bank, EPO, IAEA, ICC, ICJ, ILO, OECD, UNESCO, WHO, WMO, WIPO and the World Bank.

Figure 1: Level of similarity between PE of MT suggestions and revision of human translations and TM matches per institution and among freelancers
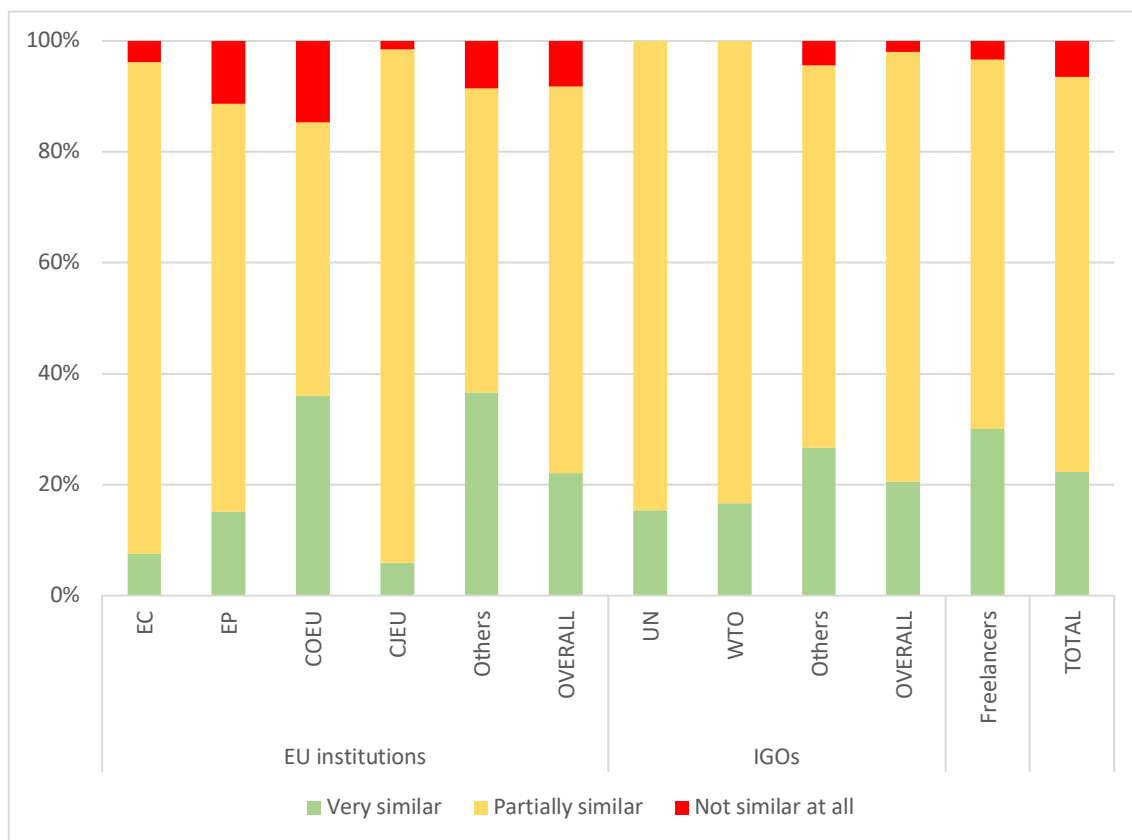
An average of only 6.49% of respondents (8.18% from EU institutions and 1.96% from IGOs) replied "not similar at all", while partial similarity was the most widespread view (71.21% on average *versus* 22.29% of perceptions as "very similar") (see Table 1 and Figure 1). No significant differences were observed per levels of experience (Table 2). The results of the COEU deviate the most, with the greatest proportions of both high similarity (36%) and no similarity answers (almost 14.67%) for a single organization, and the lowest rate of all the institutions for partial similarity (49.33%), even if this remained the most prevalent perception (see Table 1 and Figure 1). However, if we look into the comments made on the differences between PE and revision, the various levels of similarity assigned drew on comparable arguments. This suggests that such differences were assessed disparately but converged in a common translational competence core.

Table 2: Level of similarity between PE of MT suggestions and revision of human translations and TM matches per experience level

	0-5 years	**5-15 years**	**≥15 years**	**Overall**
Very similar	19.77%	25.16%	21.27%	**22.29%**
Partially similar	74.42%	69.68%	71.04%	**71.21%**
Not similar at all	5.81%	5.16%	7.69%	**6.49%**

The comments compiled on these differences, a total of 261, were analyzed using keyword (see Table 3) and discourse analysis to identify the main argumentation patterns. Practically all responses highlighted the differences between the errors or issues detected (see also next section) and/or the effort or level of intervention required. In fact, "error(s)" and "mistake(s)" were the most frequent keywords, followed by "time" (usually in connection with the attention demanded by PE), "term(s)/terminology", "correct/-ion(s)", and other keywords used to describe MT output issues ("(in)consistency/-ies", "context", "(un)reliable/-ility/rely", "logic", "(in)accuracy/-ies", "unpredictable") and the implications and effort required to fix these issues or mitigate risks ("cognitive effort", "risk(s)", "rewrite/-ing", "concentration").

Table 3: Most frequent keywords in comments about the differences between PE of MT suggestions and revision of human translations and TM matches

error(s)	59
mistake(s)	53
time	44
term(s)/terminology	43
correct/-ion(s)	34
(in)consistency/-ies	29
context	26
(un)reliable/-ility/rely	24
logic	16
(in)accuracy/-ies	13
(un)predictable	11
cognitive effort	9
risk(s)	8
rewrite/-ing	8
concentration	6

In line with the quantitative results, many of the comments underlined that the *competence required for all the tasks was essentially the same*, but that each task called for different approaches to applying the competence, in particular various levels of attention because of the divergent nature of error types in machine *versus* human translation, for example:

> *In my opinion, the competences are the same. What varies is the methodology. As one may expect different errors to occur in MT, you pay more attention to some specific aspects, such as style, register, prepositions, terminology inconsistencies, geographical variations, source of quotes, clients' preferences. Besides, I am usually more respectful with TM matches (provided they are fragments of human translations in origin) than with MT proposals.*

> *The competence itself remains the same, but the way of applying the competence and approaching a text has changed.*

Regarding the difference between human- and MT-originated segments, even if the translation competence is the same, it is crucial to understand their different genesis, as the type of issues that have to be corrected is very different in nature.

Post-editing MT suggestions requires a different kind of attention as far as the content is concerned as MT produces different mistakes than human translators. MT is also often less predictable in terms of the types of mistakes it produces. The skillset overall, however, is similar.

The competences are essentially the same, but the kind of changes and their relative frequency is somewhat different.

They are similar processes and call for the same competence, but MT can be much more labour intensive.

The *dissimilarity of error types* between machine and human translation was most commonly mentioned, as also suggested by the most frequent keywords, for example:

MT does not take into account terminology consistency, quotations from specific (legal or other) texts, specific in-house language.

MT results contain unpredictable mistakes. Mistakes in human translation are more or less predictable.

MT produces many more mistakes, especially hardly understandable mistakes, and is often not context targeted.

In human translations, errors typical of MT almost never appear. Moreover, in complex sentences, MT capabilities are very limited and the result is often not very usable. And MT still has problems with the use of appropriate terminology, which is less of a problem in the case of human translation.

The types of mistakes made by MT systems are often different from those made by competent human translators or from the discrepancies found in translation memories.

A significant proportion of comments specifically referred to the *risks* entailed by MT output "sounding fluent" but being inaccurate, and the concomitant *effort required* to ensure that such inaccuracies are not overlooked (on cognitive effort in PE, see e.g. Moorkens et al., 2015, and Herbig et al., 2019; on the priming effects of MT and quality gaps detected in post-edited texts, see e.g. Farrell, 2018, and Vardaro et al., 2019), for example:

Revising MT is far more time-demanding than translating a text from scratch. And when you revise MT results, you are prone to introduce mistakes in the sense that, at some point, you do not see mistakes/inconsistencies any longer.

More attention is needed with MT, which many times offers logical flow of text that does not match the source text.

Much more work involved in post-editing MT content. Errors are sometimes "hidden in plain sight" and you need to apply great attention to detail in order not to miss errors. Human error (from a good translator) is much more predictable.

Machine translation will often read very well at first sight, but may contain fundamental errors in meaning. Therefore, it may be harder to spot errors.

> *MT text can often seem to be well translated and grammatically correct when just reading the target text, but can be completely wrong when compared against the source text. This is rarely the case when checking human translations. This can lead to a sense of false confidence in MT.*

> *Humans and machines make different kinds of mistakes. In my experience the mistakes in MT are very tricky. You don't necessarily spot them immediately. So, if you are in a hurry, you take the risk of having serious translation mistakes.*

Finally, it is worth noting that a minority of comments referred to from-scratch translation and MT being more similar when human translation is of low quality or in the case of usable MT renderings of certain repetitive technical texts:

> The process is quite similar, sometimes human translation that has to be edited is worse or more unnatural than the machine translation. Yet, for the time being machine translation has its limits.

> It will always depend on the type of text. For some repetitive technical texts it is very similar or even better. For legal texts with lots of references and footnotes it is sometimes dangerous to rely too much on neural [machine] translation or even TMs as there are lots of quotes from other documents that have to be individually searched. A good translator should use the proper mix of CAT tools and personal research/intellect.

4. Main Issues in MT Output

As outlined in the previous section, perceptions of NMT characteristics and usability are critically conditioned by the quality of its output, and more specifically, the most typical errors made by NMT systems. This is the second major aspect examined in our study. Based on our preliminary literature review, interviews and exchanges, a taxonomy of seven *error categories* was established for survey participants to indicate the frequency of such errors in their practice. A Likert scale was used with values ranging from 0 for "never" to 3 points for "very often". Respondents were also invited to add other error types or further comments about NMT performance issues. The aim was to expand the scope of smaller-scale studies and (1) get a broad picture of the issues detected by professional users of NMT in institutional translation services, including a wide range of MT engines, languages and text types; and (2) identify any significant fluctuations between institutional settings.

Inconsistencies (average rating of 2.27) were flagged as the most frequent issue across institutions, with the highest number of "very often" replies (almost half of all the responses) (see Table 4 and Figure 2). Consistency issues are followed by a group of four error categories: idiomaticity or style issues, inappropriate terminology or phraseology, non-conformity to genre or institutional conventions and content inaccuracies, with averages of between 2.13 and 1.95 (i.e. issues detected often). A third group of two categories, incorrect references or data and linguistic errors, obtained average frequency

scores below 2 (1.66 and 1.65, respectively). They were the only categories for which "sometimes" was the most common frequency level assigned (43.40% and 46.31% of replies, respectively). "Other issues" registered an average close to never at 0.37, which means that (1) the survey population found our taxonomy comprehensive to a very high degree, and (2) other errors were marginal in their professional practice. Most respondents did not actually specify any additional error types beyond those included in the proposed categories.

Table 4: Frequency of issues detected when editing MT output or suggestions (overall scores)

	Never (0)	**Sometimes (1)**	**Often (2)**	**Very often (3)**	**Average**
Inconsistencies	1.79%	17.23%	32.89%	48.10%	**2.27**
Idiomaticity or style issues	0.89%	25.28%	34.23%	39.60%	**2.13**
Inappropriate terminology or phraseology	0.45%	26.62%	39.60%	33.33%	**2.06**
Non-conformity to genre or institutional conventions	2.68%	28.86%	32.98%	35.57%	**2.01**
Content inaccuracies	0.45%	31.77%	40.49%	27.29%	**1.95**
Incorrect references or data	5.82%	43.40%	29.53%	21.25%	**1.66**
Linguistic errors	2.68%	46.31%	34%	17%	**1.65**
Other	78.52%	9.84%	7.38%	4.25%	**0.37**

Figure 2: Frequency of issues detected when editing MT output or suggestions (overall scores)
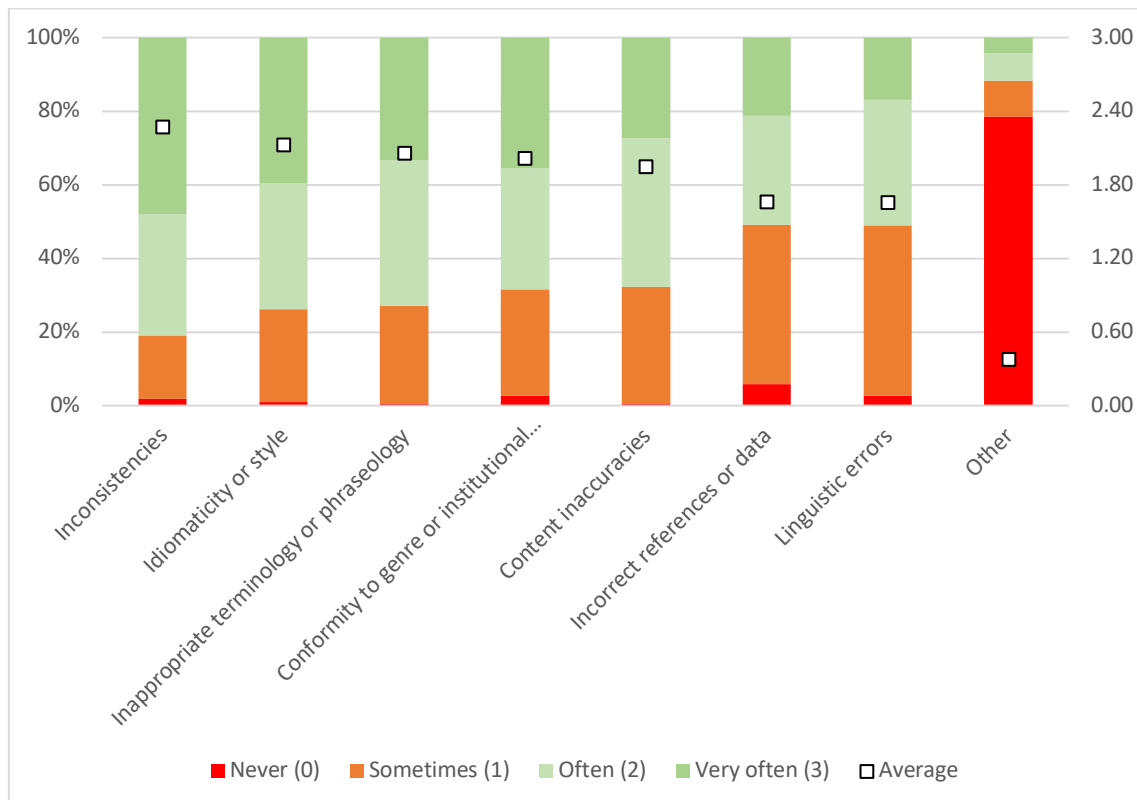
Overall, the results are in line with previous studies on NMT output (see Section 2), especially with regard to inconsistencies, terminology, mistranslations and style. Consistency issues in particular, including intratextual consistency, have been identified as a weakness of NMT, even in comparison with SMT (see e.g. Guzmán & Prieto Ramos, 2024; Toral & Sánchez-Cartagena, 2017; Arnejšek & Unk, 2020). The findings are also consistent with the main NMT deficiencies elicited in the open-ended descriptions of NMT summarized in the previous section, with data or grammar being less recurrent issues according to respondents, as opposed to consistency, terminology, house style or accuracy errors. Rather than additional error types, some of the qualitative insights compiled in the survey section on NMT issues referred to another interesting aspect reported in previous research (e.g. Koehn & Knowles, 2017; Lesznyák, 2019; Pouget-Abadie et al., 2014): the fact that, in the case of *long or complex sentences*, NMT quality tends to worsen, and therefore NMT suggestions become less usable:

> *I encounter all the above issues in greater degree if the segment translated is particularly long.*

> *MT requires checking the translation from a different point of view than that of a human translator: there is little chance of omissions or incorrect numbers but it is more likely not to "understand" what relates to what in longer phrases/sentences, not to use the correct terminology if a word has multiple meanings or choice of terminology depends on context.*

> *In complex sentences, MT capabilities are very limited and the result is often not very usable.*

> *In long sentences, MT may omit parts or repeat words without any reason, especially when it comes to nouns very close in meaning connected with "and" or "or".*

The results were highly similar among all the *institutions* considered (see Table 5), with inconsistencies systematically rated as the most frequent issue, followed by the second and the third groups of issues referenced above per frequency levels, and with only some modest frequency variations among the various IOs within each group of errors. While our findings are not indicative of the performance of any specific NMT system, the fact that EU translation services overwhelmingly use eTranslation provides a basis for comparison with the more diversified use of NMT tools among (and even within) the IGOs.

Despite the differences in tools, the results are strikingly similar, with all frequency scores for specific error categories (i.e. excluding the negligible category of "other") within a +/-0.09 divergence between EU and IGO averages, except for two categories, non-conformity to genre or institutional conventions and incorrect references or data. The more significant differences registered for these categories (0.25 and 0.14 points, respectively, less frequent in the EU institutions than in IGOs) can be interpreted as a sign of potential benefits of training custom-made engines with institution-specific data (eTranslation in this case).

Table 5: Frequency of issues detected when editing MT output or suggestions per institution

	Incon-sist.	Idio-maticity or style	Termi-nology	Genre or institu-tional conven-tions	Content inaccu-racies	Refer-ences or data	Linguis-tic er-rors	Other
EC	2.55	2.24	2.12	2.10	2.12	1.67	1.78	0.53
EP	2.20	2.14	1.86	1.70	1.74	1.42	1.68	0.36
COEU	2.20	2.07	2.05	1.80	1.82	1.54	1.51	0.32
CJEU	2.29	1.97	2.16	2.00	2.13	1.88	1.91	0.46
Other EU inst.	2.32	2.10	2.03	2.10	1.88	1.64	1.55	0.19
Overall – EU	**2.30**	**2.09**	**2.05**	**1.95**	**1.93**	**1.64**	**1.67**	**0.36**
UN	2.26	2.18	2.13	2.26	2.05	1.68	1.55	0.47
WTO	2.29	2.06	1.88	2.06	2.12	2.12	1.59	0.88
Other IGOs	2.67	2.33	2.50	2.58	2.33	1.92	1.67	0.58
Overall – IGOs	**2.24**	**2.11**	**2.08**	**2.20**	**2.01**	**1.78**	**1.58**	**0.51**
Average	**2.27**	**2.13**	**2.06**	**2.01**	**1.95**	**1.66**	**1.65**	**0.37**

Figure 3: Frequency of issues detected when editing MT output or suggestions per institution
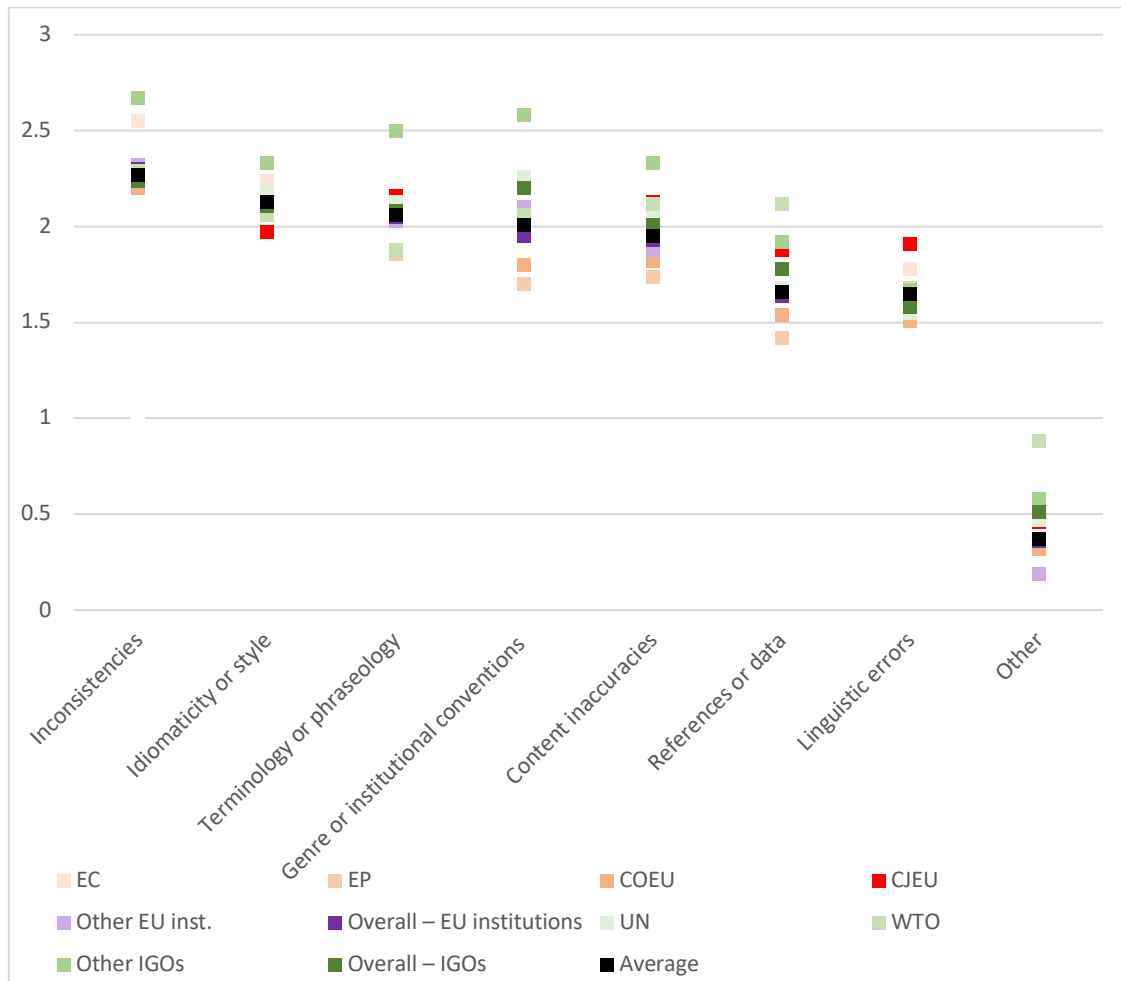
However, internal fluctuations within the EU institutions nuance this hypothesis and suggest an additional connection between error frequency and text types or the nature of the translation work at hand, which would warrant further statistical testing. For example, eTranslation suggestions seem to require more editing in the EC, which is responsible for translating legislative proposals as part of the EU's ordinary legislative procedure, than in the case of the EP and the COEU, which deal with amendments to such proposals. This critical factor was also highlighted by EU institutional informants. In fact, the scores of the EC and the CJEU were not only the highest of all the large EU institutions surveyed, but were also above the global averages in all the categories, except for the CJEU's scores for idiomaticity or style issues (1.97, the lowest of the series for this category) and genre or institutional conventions (2.00, almost equal to the average of 2.01). In contrast, lawyer-linguists translating for the CJEU also perceived linguistic errors as more recurrent than in any other institution (score of 1.91 *versus* the overall average of 1.65).

Nonetheless, apart from this score for linguistic errors and that of incorrect references or data (more frequently indicated at the WTO), the other IGOs category registered the highest error frequency scores of all the IOs, i.e. the most negative perceptions of NMT output with regard to the five most recurrent error categories listed above (see Table 5). This might derive from the fact that most of the smaller organizations tend to work with less tailored MT systems and account for a more limited proportion of text data used to train commercially available MT engines, as opposed to the larger organizations. Among these larger institutions, EC translators seem to encounter inconsistencies and idiomaticity issues more often than in any other large institution; and CJEU lawyer-linguists registered the highest frequency scores for terminological, content accuracy and linguistic issues; while non-conformity to genre or institutional conventions is most recurrent in the UN (with no internally-developed MT system). This was the only institution where this category was considered the most frequent issue together with inconsistencies.

5. The Impact of MT Use on Translation Processes, Competences and Products

Based on our preliminary analyses, six key statements were proposed to capture some of the most pervasive or controversial ideas about the impact of NMT use on institutional translation processes, product quality and competences required.

(a) Translation process
A1: MT use has contributed to improving my productivity.

A2: MT use generally enables me to save time on typing or routine searches and to focus on solving translation problems and improving translation quality.

A3: Overall, using MT has more advantages than disadvantages.

(b) Translation product

B1: MT use entails higher risks of overlooking inaccuracies and other errors than in the case of texts translated without MT.

B2: MT use is having a negative impact on translation or language quality, in particular, idiomaticity and style.

(c) Translation competence

C1: MT use has reinforced the need for solid translation competence and subject matter knowledge in order to spot issues and ensure translation quality.

Survey respondents were asked to express their level of agreement or disagreement with these statements on a Likert scale ranging from 0 ("I fully disagree") to 3 ("I fully agree"). Once again, they could also add comments. The statements on translation process (A1 to A3) were deliberately conceived to measure the positive aspects of NMT use, particularly with regard to time efficiency and productivity, and thus balance the focus on errors and editing effort examined in the previous sections. More specifically, A2 aims to test the idea that MT suggestions provide a first basis for inspiration and may save time for the more complex problem-solving.

In contrast, statements B1 and B2 refer to the risks of MT errors remaining in the target text and the negative impact on translation quality and the target language, especially in terms of idiomaticity and style, even when the final target text is error-free (on MT-associated risks of impoverishing language, see e.g. Farrell, 2018; Vanmassenhove et al., 2021; Jiang & Niu, 2022). Finally, the statement on translation competence (C1) was meant to encapsulate what emerged as a rather consensual view during the preliminary phase, and to test the connection with B1 (i.e. translation expertise is needed to spot MT issues and mitigate the risk of overlooking them). As will be discussed below, the impact of NMT on translation competence was further elaborated in another section of the survey that warrants more in-depth analysis due to its multifaceted nature.

The results indeed corroborate the highest level of agreement with the abovementioned correlation (see Table 6 and Figure 4), as the reinforced need for translation expertise in the new MT-supported environments (statement C1) got the top average score of 2.40 (87.36% of "fully agree" or "mostly agree" replies), together with the perceived risks of overlooking errors (statement B1, with a score of 2.36 and 86.21% of approval) for which solid translation competence is required. These two statements gathered more than 50% of full agreement responses, unlike the remaining key statements (A1-A3, B2). The positive statements about productivity (A1) and overall advantages of using MT (A3)

ranked next, with averages just above the "mostly agree" rating (2.10 and 2.02, respectively). Finally, the statements on time-saving for improving quality (A2) and the negative impact on translation products (B2) drew the most diverse reactions. While the averages (1.79 and 1.80, respectively) point to stronger agreement than disagreement, a significant proportion of respondents mostly or fully disagreed with A2 (31.87%) and B2 (40.66%).

Table 6: Assessment of the impact of MT use (overall scores)

		Fully disagree (0)	Mostly disagree (1)	Mostly agree (2)	Fully agree (3)	Average
Translation process	**A1**	5.78%	12.67%	47.78%	33.78%	**2.10**
	A2	10.33%	21.54%	47.03%	21.10%	**1.79**
	A3	5.15%	14.99%	52.57%	27.29%	**2.02**
Translation product	**B1**	1.75%	12.04%	34.57%	51.64%	**2.36**
	B2	5.33%	35.33%	33.78%	25.56%	**1.80**
Translation competence	**C1**	2.22%	10.42%	32.15%	55.21%	**2.40**

Figure 4: Assessment of the impact of MT use (overall scores)
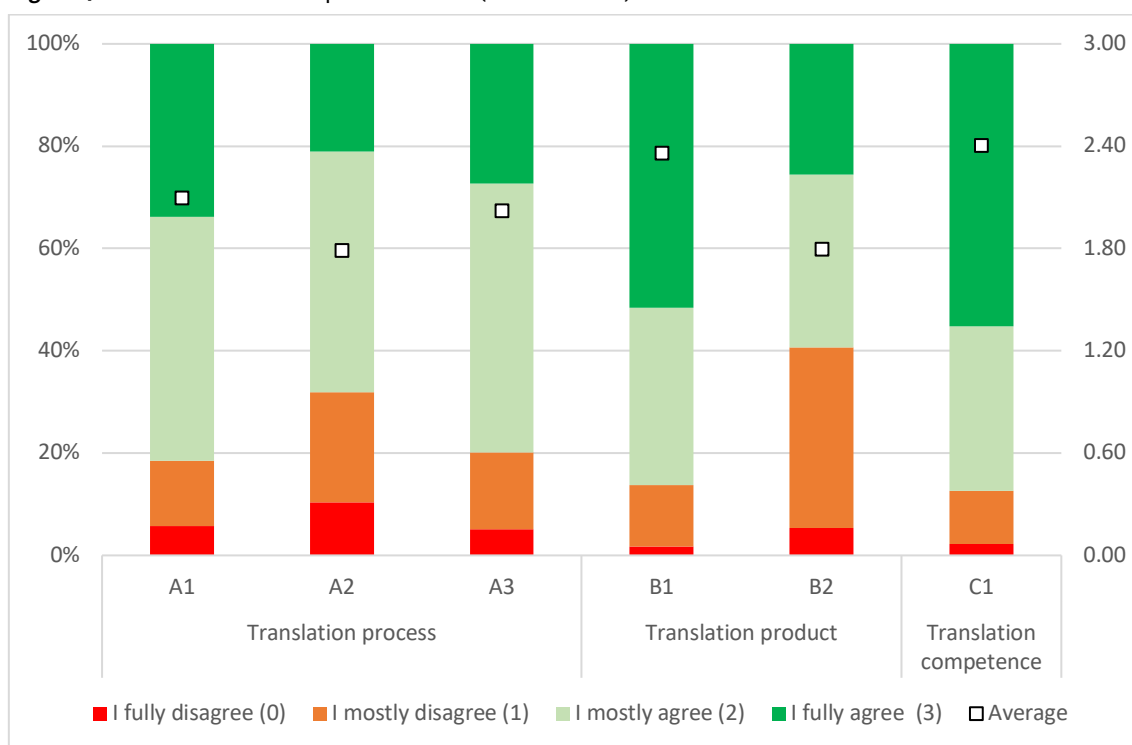

The issue of impact on translation product and target language (B2) was expected to generate mixed results, as it would be difficult for some respondents to admit negative views despite their efforts to ensure quality, or simply to assess the longer-term effects

or the impact for all text types, considering the various levels of quality assurance currently implemented to mitigate risks. The distribution of replies also suggests that statement A2 was equally divisive, with 68.13% of agreement replies, but also the highest proportion of full disagreement across the entire series (10.33%). This implies that, despite efficiency or productivity gains (statement A1), individual experiences varied as to how MT may or may not have saved them time and, if it did, as to what activities it may have allowed them to focus more on. For many, as also elicited in sections 3 and 4, translation time had to be used primarily for editing or correcting errors.

Table 7: Assessment of the impact of MT use per institution

	A1	A2	A3	B1	B2	C1
EC	2.35	1.69	2.22	2.58	2.06	2.42
EP	2.27	1.81	2.06	2.26	1.67	2.40
COEU	2.14	1.75	2.07	2.41	1.69	2.19
CJEU	2.16	2.01	2.15	2.27	1.49	2.28
Other EU inst.	2.08	1.89	2.03	2.40	1.76	2.39
Overall – EU	**2.18**	**1.84**	**2.10**	**2.38**	**1.72**	**2.33**
UN	1.82	1.58	1.82	2.26	2.08	2.66
WTO	2.17	1.72	1.83	2.72	2.22	2.61
Other IGOs	1.76	1.63	1.85	2.21	1.80	2.56
Overall – IGOs	**1.86**	**1.63**	**1.83**	**2.32**	**1.99**	**2.61**
Average	**2.10**	**1.79**	**2.02**	**2.36**	**1.80**	**2.40**

Figure 5: Assessment of the impact of MT use per institution
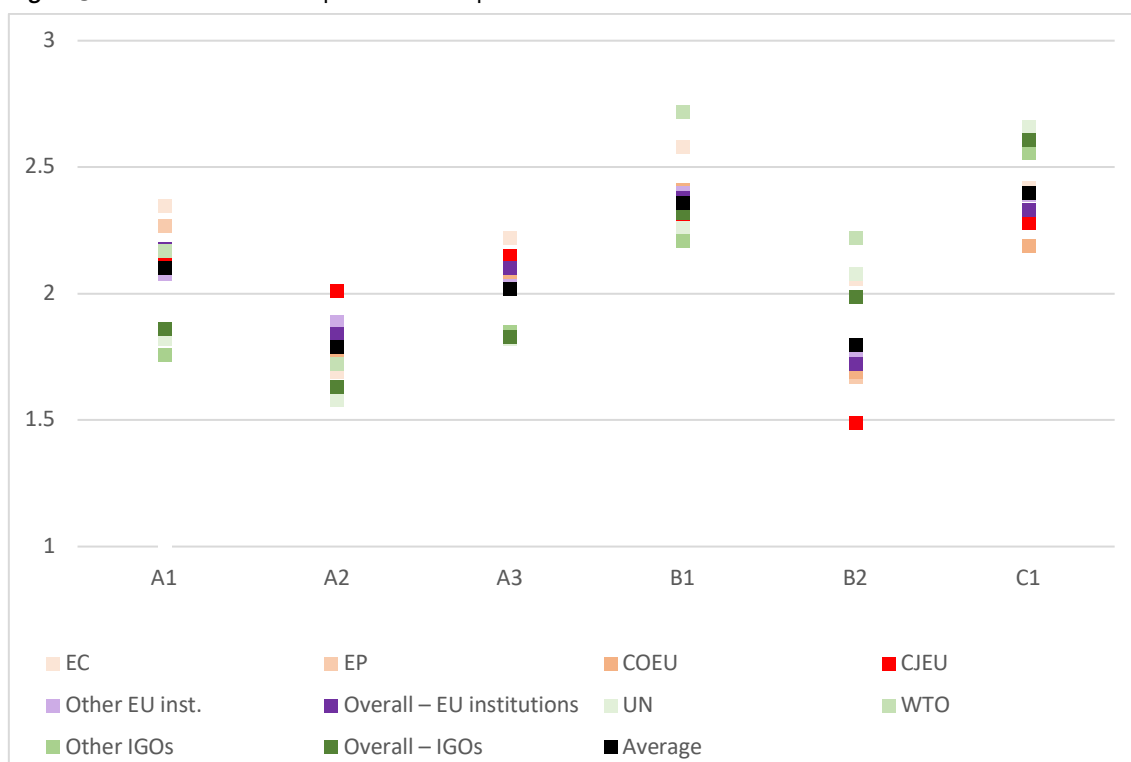
Given these global findings, it was particularly interesting to examine variations per institutional setting and years of experience. The overall scores for the EU *institutions* were slightly above the global averages for all the statements, except for B2 and C1 (see Table 7 and Figure 5). The reverse applies to the overall IGO scores. Yet, the fluctuations between the lowest and the highest scores per statement are not very significant, approximately +/-0.50, with the exception of B2 (+/-0.73). The most remarkable deviation from the patterns of each group of institutions is found at the EC, with the highest agreement rates of all the EU institutions, except for A2, rated below the global average and lower than all the other EU institutions. This means that positive views on productivity (A1) and advantages of MT use (A3) at the EC's DGT co-exist with high awareness of its risks (B1) and quality implications (B2), and the need for solid translation competence (C1), while potential efficiency gains might not entail more time for quality improvements (1.69 score for statement A2). A combined reading of these findings implicitly aligns with previous research where DGT translators suggested that NMT can be particularly efficient for producing lower quality translations (Lesznyák, 2019: 21).

However, it is at the UN that A2 scored the lowest (1.58), together with the least agreement of all the institutions with the advantages of MT use (1.82 for A3), the strongest approval of the need for translation expertise (2.66 for C1) and the second highest score for the negative impact on quality (2.08 for B2). In fact, a correlation can be established between more frequent exposure to some error types (i.e. at the EC, the UN, the WTO and other IGOs –see Table 5 and Figure 3–) and comparatively lower scores for statement A2 (i.e. time gains not systematically used for quality improvements, but possibly for

detecting and fixing errors) and higher for B2 and C1 (i.e. greater awareness of the impact on quality and the need for translation competence). While this correlation is also supported by common sense, it would require further empirical testing in each setting.

As for variations per *experience levels,* which are generally associated with translators' age groups (i.e. 5 first years of experience associated with early career or more junior staff, 10 additional years for mid-career professionals, and longer for more senior staff), although the differences in perceptions are not dramatic, three patterns seem to emerge (see Table 8 and Figure 6). First, positive views in A1 to A3 statements were more highly rated by less experienced translators (0-5 years), especially productivity gains (2.32) and overall advantages (2.16), which gradually decreased at higher experience levels (down to 1.99 and 1.95, respectively, among the more senior staff).

Second, the pattern is the opposite for the negative views on risks and impact on quality, with the lowest B1 and B2 agreement rates among the junior translators. Even though the averages for practically all the statements are within the "mostly agree" range in all experience levels, B1 and B2 scores in particular point to lower awareness of translation quality issues among the less experienced translators. In this regard, it is interesting to note that previous research also connects lesser experience with lesser awareness of MT limitations (Moorkens et al., 2018) and lower quality performance (Quinci, 2023: 191). The comparative difference between group perceptions of the impact of MT use on translation quality and target language (B2) is actually the most divergent of all the statements validated, 1.56 among junior translators (i.e. no strong agreement or disagreement) *versus* 1.93 (mostly agreement) among the most experienced and 1.73 for the mid-career group. This can be linked to the broader perspective that experience brings to assessing the impact of technological changes on product quality. Finally, the statement on translation competence needs (C1) was the most similarly rated among the three groups.

Table 8: Assessment of the impact of MT use per experience level

	0-5 years	5-15 years	≥15 years	Overall
A1	2.32	2.12	1.99	**2.10**
A2	1.92	1.75	1.77	**1.79**
A3	2.16	2.05	1.95	**2.02**
B1	2.23	2.40	2.39	**2.36**
B2	1.56	1.73	1.93	**1.80**
C1	2.37	2.42	2.41	**2.40**

Figure 6: Assessment of the impact of MT use per experience level
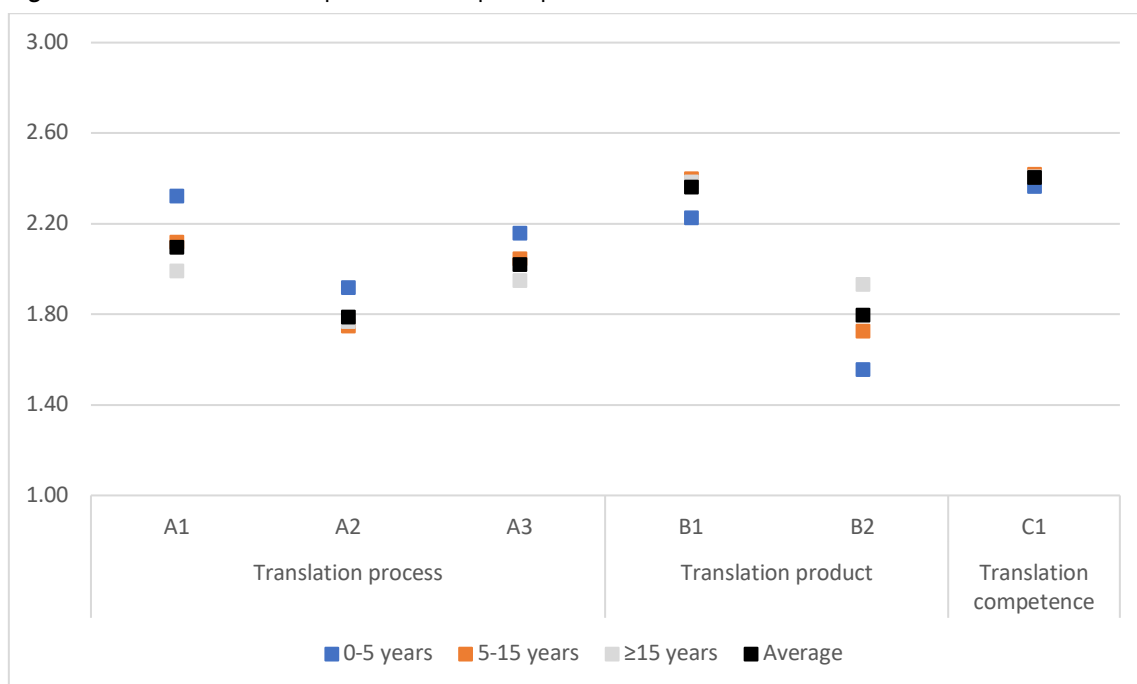
Open-ended comments in connection with the statements assessed often combined several aspects, and tended to focus on methodological changes, effort and competences required, for example:

▪ Focus on translation processes:

MT involves double work: overhauling the machine takes much more time than translating from scratch, and always depending on the correct functioning of the system, which is not usually the case.

MT is a boon, but requires extreme vigilance.

The working methods have dramatically changed. Delivering always the same high quality with languages we are not able to speak is a daily challenge. To master the IT tools modifies our job deeply and requests huge investments.

- Focus on translation products and target language:

When revising MT matches there is more noise; the translation tends to be more literal and often the structure of sentences does not sound natural, even if they are correct. Language is impoverished.

Due to time constraints, revising MT suggestions might lead to an impoverishment of the target language. The translator's linguistic vigilance might be reduced. We already see wrong syntax given by the machine being approved by translators.

- Focus on translation competence:

Using MT properly and avoiding the blunders to which it is prone requires not only that translators and revisers have a more thorough knowledge of both the source language and the subject matter, but also that they be confident enough in that knowledge to override without hesitation the translations generated by the software.

[...] What's definitely more important than before is proper understanding of subject matter (ability to read between the lines and avoid word for word translation) and ability to critically review previously translated text. And having the time to properly review MT output.

[...] MT output tends to look good at first sight; this means that translators need to be even more proficient in methodology, thematic and textual competence, and also instrumental competence, to understand how the tool works. The real danger of MT is if its use were to pollute the data quality on which all the accumulated efficiency gains to date rely, both for the reuse of translation memory segments and MT training.

In order to gain a more detailed understanding of the impact of NMT on specific translation *competence components*, respondents were also asked to evaluate whether each of five competences, as described in Prieto Ramos (2024b) ((1) language or communicative and textual, (2) strategic or methodological, (3) thematic and cultural, (4) instrumental and (5) interpersonal or professional management competences) was more, equally or less important (values of 2, 1 or 0, respectively) to ensure translation quality in institutional settings since the advent of MT.

Table 9: Change in importance of competences to ensure translation quality in institutional settings since the advent of MT (overall scores)

	Less important than before (0)	Equally important as before (1)	More important than before (2)	Average
Instrumental	3.16%	42.83%	54.01%	**1.51**
Language / communicative & textual	1.90%	53.80%	44.30%	**1.42**
Thematic & cultural	0.63%	57.38%	41.98%	**1.41**
Strategic / methodological	1.90%	61.39%	36.71%	**1.35**
Interpersonal & professional management	7.38%	77.43%	15.19%	**1.08**

Figure 7: Change in importance of competences to ensure translation quality in institutional settings since the advent of MT (overall scores)
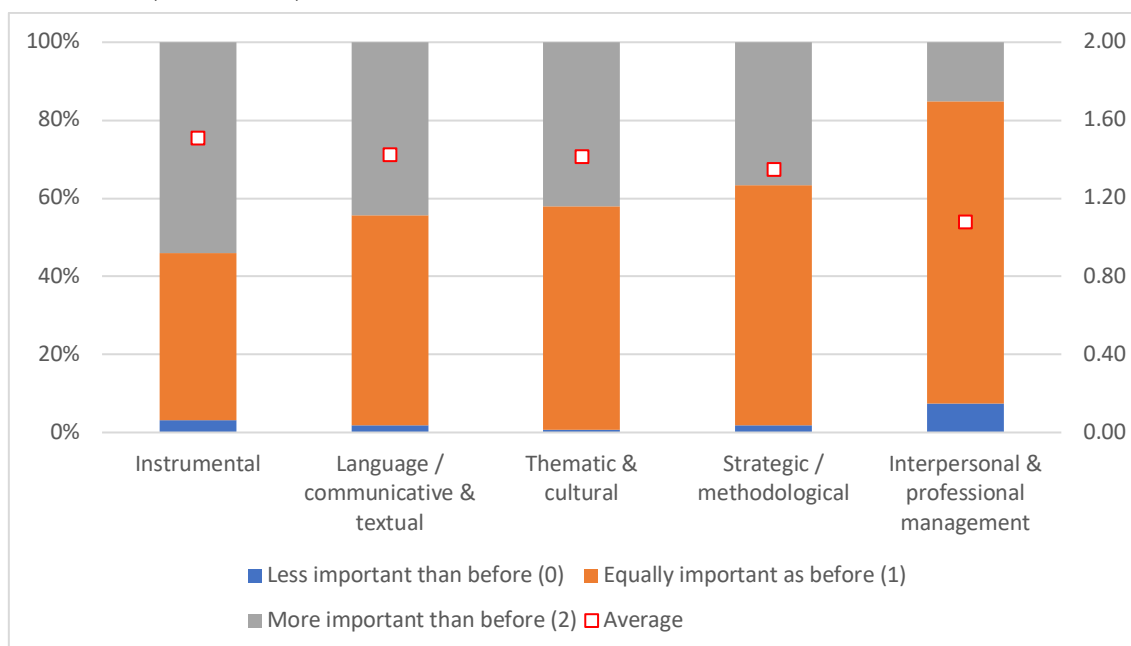
According to the overall scores (Table 9 and Figure 7), all competences were considered as important as before the integration of MT or even more important than before, with averages between 1.08 and 1.51. Unsurprisingly, instrumental competence (i.e. use of tools and information management) registered the most marked increase, with more top rating replies than for the other competences, but closely followed by language or communicative and textual, thematic and cultural, and strategic or methodological competences.

Despite the growing relevance of mastering translation tools, the three core language, thematic and strategic competences were still perceived as more important than instrumental competence to ensure translation quality (see more detailed results of this section of the survey in Prieto Ramos, 2024b). Among the more junior group, however, instrumental competence ranked second in importance only after language competence and closely followed by strategic competence, while thematic and cultural competence ranked fourth. This compares with the perceptions of more experienced translators, who placed thematic and cultural competence as the second most relevant together with strategic competence (Prieto Ramos, 2024b: 166–167). These findings, when read in conjunction with the other variations across the experience levels referenced in this section, suggest a less critical view of MT reliability for retrieving domain-specific content and a trend towards underestimating subject matter knowledge in comparison with more experienced translators, as well as a comparatively lower awareness of the risks associated with this reliance on machine-generated suggestions to detect MT errors and achieve translation quality.

Skills related to thematic or domain competence, such as knowledge of the organization and subject-matter knowledge to understand complex topics, were among the most marked increases in competence deficits among new recruits in IOs over the past decade (Lafeber, 2022: 39). A few of the comments made in the survey section on MT impacts actually referred to the experience factor and the potentially harmful effects of (over-)reliance on technology from an early stage, for example:

> The competences are equally important. They are just used differently. The important thing is that people have to develop these skills properly before they can start translating effectively with MT.

> It is very harmful to engage young translators in post-editing. They need a more solid experience in translation.

6. Discussion and Conclusions

The quantitative and qualitative findings of our study on the features, issues and impacts of MT use, as perceived by a large cohort of institutional translators, make an empirical contribution to our understanding of what machine-aided professional translation entails today. The insights gathered from 24 IOs are generally in line with previous smaller-scale studies on the differences and overlaps between translation, revision and PE. As opposed to some of these studies, however, our results leave little doubt as to the translational competence core of PE. As elicited in the survey, the core decision-making competence is aimed at ensuring the communicative adequacy of the final target text, whether building it from scratch or drawing on machine-generated inputs to varying degrees. It is the diversity and variable nature of these inputs that has reshaped the concept of translation, as there is no optimal one-size-fits-all combination of tools for every translation scenario, and translation competence requires increasing versatility to engage with these tools and meet both efficiency and quality expectations. This demands the ability to discern the type and level of human-machine interaction relevant to each purpose, while also adapting to the challenges of such interaction.

The evidence outlined above supports a revised concept of translation that considers an increasingly hybrid and dynamic mix of "translational actions" encompassing a wider range of inputs, applications and types of attention into what Holz-Mänttäri (1984) broadly defined as the core process of reading a source text and conveying its content into a target text. In practice, and despite recent advances in MT, the paradigm of recycling translation segments that became commonplace since the popularization of TMs still prevails, especially in institutional settings, where revision in a broad sense (Pym 2013: 496), inter- and intratextual consistency and data quality are primary concerns. As already noted by Pasteur (2013: 297) a decade ago, institutional "[t]ranslators' foremost

concern and main demand remains the relevance and the validity of the background materials provided upstream, whether it is MT outputs, TM-based pre-translations, terminology data or lists of previously translated fragments".

As illustrated by our survey respondents, simplistic constructs of translation as PE of drafts pre-translated by machines do not accurately reflect current practices. Institutional translation professionals also globally agree on the diversity and great unpredictability of NMT errors, and the effort required to detect and fix these issues, as the main distinctive features of PE. They also strongly acknowledge the need for solid translation competence and subject matter knowledge to ensure that NMT errors are not overlooked. A majority of respondents see more advantages than disadvantages in using MT, especially with regard to productivity gains, but also higher risks of lower translation quality and negative effects on the target language. The time saved on typing or routine searches when using NMT cannot always be devoted to improving translation quality, as translators may be occupied with detecting and fixing issues in an effort to deliver an error-free translation. These perceptions of relative efficiency and risks of negative effects are slightly more pervasive among translators of institutions where the frequency scores for certain error types are higher.

Nonetheless, the results per institutional setting are highly similar. A broad pattern emerged when crossing quantitative and qualitative data, regardless of the MT systems, languages and text types involved: NMT risks were often linked to the fact that NMT suggestions can read as fluent, since grammar and linguistic or factual errors are less frequent, but may carry a number of other issues, particularly recurrent inconsistencies (the most prevalent error type indicated), idiomaticity or style issues, inappropriate terminology or phraseology, non-conformity to genre or institutional conventions and content inaccuracies.

As for the "experience factor", translation professionals of all experience levels showed similar perceptions of the translational nature and specificities of PE and the need for translation expertise to ensure quality when using MT. However, the more senior groups perceived MT as having a more negative impact on quality than less experienced groups, while the latter perceived MT as posing less risk and providing advantages more often. The less experienced translators also rated the relevance of instrumental competence more highly, only after language competence, and revealed a tendency towards underestimating thematic competence in comparison with more experienced translators. In turn, this can be related to a higher risk of not detecting NMT limitations with domain-specific content, as found in previous research.

Several implications derive from these conclusions. We will only mention some of the most relevant. Firstly, rather than devaluing the role of the human translator, MT use has actually reinforced the relevance of translation expertise and the need for excellent and increasingly-adaptive human translators to ensure quality and manage the new risks of augmented translation workflows. Given the hybridization of translation and the increased human-machine interaction, this involves mastering the use of tools and

their critical assessment, for which the core translation competences remain critical to achieve both process efficiency and product adequacy. The narrative comments from survey respondents illustrate how they develop nuanced views of the tools they use and when, as they embrace new technologies.

Secondly, professional development and quality assurance practices must take account of the changing translation processes not only to promote data quality analysis, risk awareness and mitigation strategies at institutional and individual levels, but also to factor in the implications for cognitive load and motivation. Thirdly, the increasing adaptability expected of human translators and the heightened attention to automation in translator training, including the integration of PE, should not interfere with the development of the core competences required to achieve quality in translation, including proper domain specialization (see also Prieto Ramos & Guzmán, 2024). It is precisely domain specialization and targeted research skills that can make one of the greatest differences in enhancing quality and customizing MT tools. Last but not least, the evidence-based knowledge generated through this kind of research can be critical to empower translators in their role in the digital era and contribute to further educating clients and the general public. While technological innovations are swiftly integrated and "demystified" in professional practice, an emerging mismatch between perceptions of MT among translation experts and non-experts represents an additional challenge in the field, one that resonates with translators beyond institutional settings.

Acknowledgments

References

Arnejšek, Mateja & Unk, Alenka (2020). Multidimensional assessment of the eTranslation output for English–Slovene. In Martins et al. (Eds.), *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation* (pp. 383–392). Lisbon: European Association for Machine Translation.

Cadwell, Patrick; Castilho, Shela; O'Brien, Sharon & Mitchell, Linda (2016). Human factors in machine translation and post-editing among institutional translators. *Perspectives*, 26(3), 301–321. DOI: 10.1075/ts.5.2.04cad.

Daems, Joke & Macken, Lieve (2021). Post-editing human translations and revising machine translations: impact on efficiency and quality. In Koponen, Mossop, Robert & Scocchera (Eds.), *Translation Revision and Post-Editing: Industry Practices and Cognitive Processes* (pp. 50–70). Abingdon: Routledge.

Do Carmo, Félix & Moorkens, Joss (2021). Differentiating editing, post-editing and revision. In Koponen, Mossop, Robert & Scocchera (Eds.), *Translation Revision and Post-Editing: Industry Practices and Cognitive Processes* (pp. 35–49). Abingdon: Routledge.

Farrell, Michael (2018). Machine translation markers in post-edited machine translation output. In Chambers et al. (Eds.), *Proceedings of the 40th Conference Translating and the Computer* (pp. 50–59). London: Asling.

Guzmán, Diego & Prieto Ramos, Fernando (2024). Human and machine translation of legal terminology in international institutional settings: A case study. *De Europa - European and Global Studies Journal*, special issue, 61–81.

Herbig, Nico; Pal, Santanu; Vela, Mihaela; Krüger, Antonio & van Genabith, Josef (2019). Multi-modal indicators for estimating perceived cognitive load in postediting of machine translation. *Machine Translation*, 33(1), 91–115. DOI: 10.1007/s10590-019-09227-8.

Holz-Mänttäri, Justa (1984). *Translatorisches Handeln: Theorie und Methode [Translational Action: Theory and Method.]*. Helsinki: Suomalainen Tiedeakatemia.

Jiang, Yue & Niu, Jiang (2022). A corpus-based search for machine translationese in terms of discourse coherence. *Across Languages and Cultures*, 23(2), 148–166. DOI: 10.1556/084.2022.00182.

Koehn, Philipp & Knowles, Rebecca (2017). Six challenges for neural machine translation. In Luong et al. (Eds.), *Proceedings of the First Workshop on Neural Machine Translation*. Vancouver: Association for Computational Linguistics. DOI: 10.18653/v1/W17-3204.

Konttinen, Kalle; Salmi, Leena & Koponen, Maarit (2021). Revision and post-editing competences in translator education. In Koponen, Mossop, Robert & Scocchera (Eds.), *Translation Revision and Post-Editing: Industry Practices and Cognitive Processes* (pp. 187–202). Abingdon: Routledge. DOI: 10.4324/9781003096962-15.

Lafeber, Anne (2022). Skills and Knowledge Required of Translators in Institutional Settings. In Svoboda, Biel & Sosoni (Eds.), *Institutional Translator Training* (pp. 30–48). New York: Routledge. DOI: 10.4324/9781003225249-4.

Lesznyák, Ágnes (2019). Hungarian translators' perceptions of neural machine translation in the European Commission. In Forcada et al. (Eds.), *Proceedings of MT Summit XVIII: Translator, Project and User Tracks* (pp. 16–22). Dublin: European Association for Machine Translation.

Lommel, Arle (2018). Augmented translation: A new approach to combining human and machine capabilities. In Cherry & Neubig (Eds.), *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas* (Volume 2: User Track) (pp. 5–12). Boston: AMTA.

Macken, Lieve; Prou, Daniel & Tezcan, Arda (2020). Quantifying the effect of machine translation in a high-quality human translation production process. *Informatics*, 7(2). DOI: 10.3390/informatics7020012.

Moorkens, Joss; O'Brien, Sharon; Da Silva, Igor; De Lima Fonseca, Norma & Alves, Fabio (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation*, 29(3), 267–284. DOI: 10.1007/s10590-015-9175-2.

Moorkens, Joss; Toral, Antonio; Castilho, Sheila & Way, Andy (2018). Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces*, 7(2), 240–262. DOI: 10.1075/ts.18014.moo.

Nitzke, Jean; Hansen-Schirra, Silvia & Canfora, Carmen (2019). Risk management and post-editing competence. *The Journal of Specialised Translation*, 31, 239 –259.

Olohan, Maeve (2021). Post-editing: a genealogical perspective on translation practice. In Bisiada (Ed.), *Empirical Studies in Translation and Discourse* (pp. 1–25). Berlin: Language Science Press. DOI: 10.5281/zenodo.4450077.

Pasteur, Olivier (2013). Technology at the service of specialized translators at international organizations. In Borja Albi & Prieto Ramos (Eds.), *Legal Translation in Context: Professional Issues and Prospects* (pp. 283–297). Bern: Peter Lang.

Pouget-Abadie, Jean et al. (2014). Overcoming the curse of sentence length for neural machine translation using automatic segmentation. In *Proceedings of SSST-8* (pp. 78–85). Doha: Association for Computational Linguistics. DOI: 10.48550/arXiv.1409.1257.

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*, 28(1), 11–30. DOI: 10.1007/s11196-014-9390-9.

Prieto Ramos, Fernando (2024a). Patterns of human-machine interaction in legal and institutional translation: from hype to fact. *Polissema*, 24, 1–27.

Prieto Ramos, Fernando (2024b). Revisiting translator competence in the age of artificial intelligence: the case of legal and institutional translation. *The Interpreter and Translator Trainer*, 18(2), 148–173. DOI: 10.1080/1750399X.2024.2344942.

Prieto Ramos, Fernando & Guzmán, Diego (2024). The impact of specialised translator training and professional experience on legal translation quality assurance: an empirical study of revision performance. *The Interpreter and Translator Trainer*, 18(2), 313–337. DOI: 10.1080/1750399X.2024.2344948.

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

Quinci, Carla (2023). *Translation Competence: Theory, Research and Practice*. Abingdon & New York: Routledge.

Rico, Celia & Torrejón, Enrique (2012). Habilidades y perfil del nuevo rol del traductor como poseditor de traducción automática. *Tradumàtica*, 10, 166–178. DOI: 10.5565/rev/tradumatica.18.

Robert, Isabelle S.; Remael, Aline & Ureel, Jim J. J. (2017). Towards a Model of Translation Revision Competence. *The Interpreter and Translator Trainer*, 11(1), 1–19. DOI: 10.1080/1750399X.2016.1198183.

Robert, Isabelle S.; Schrijver, Iris & Ureel, Jim J. J. (2023). Comparing L2 translation, translation revision, and post-editing competences in translation trainees: An exploratory study into Dutch–French translation. *Babel*, 69(1), 99–128. DOI: 10.1075/babel.00307.rob.

Rossi, Caroline & Chevrot, Jean-Pierre (2019). Uses and perceptions of machine translation at the European Commission. *The Journal of Specialised Translation*, 31, 177–200.

Sánchez-Gijón, Pilar; Moorkens, Joss & Way, Andy (2019). Post-editing neural machine translation versus translation memory segments. *Machine Translation*, 33, 31–59. DOI: 10.1007/s10590-019-09232-x.

Scocchera, Giovanna (2017). *La Revisione Della Traduzione Editoriale Dall'inglese All'italiano. Ricerca, Professione, Formazione*. Rome: Aracne.

Stefaniak, Karolina (2020). Evaluating the usefulness of neural machine translation for the Polish translators in the European Commission. In Martins et al. (Eds.), *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation* (pp. 263–269). Lisbon: European Association for Machine Translation.

Stefaniak, Karolina (2023). Machine translation and terminology: the experience of the European Commission. In Castilho, Caro Quintana, Stasimioti & Sosoni (Eds.), *Proceedings of the New Trends in Translation and Technology Conference - NeTTT 2022* (pp. 134–141). Shoumen: Incoma.

Strandvik, Ingemar (2020). Digital transformation and institutional translation – change and challenges. In Della-Zuanna & Kurz (Eds.), *Translation Quality in the Era of Digitalization* (pp. 466–494). Berlin: Elanders GmbH.

Teixeira, Carlos (2014). *The Impact of Metadata on Translator Performance: How Translators Work with Translation Memories and Machine Translation. Doctoral thesis*. Universitat Rovira i Virgili & Katholieke Universiteit Leuven.

Toral, Antonio & Sánchez-Cartagena, Víctor (2017). A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. In Lapata, Blunsom & Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 1063–1073). Valencia: Association for Computational Linguistics.

Vanmassenhove, Eva; Shterionov, Dimitar & Gwilliam, Matthew (2021). Machine translationese: effects of algorithmic bias on linguistic complexity in machine translation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 2203–2213). Kyiv: Association for Computational Linguistics. DOI: 10.48550/arXiv.2102.00287.

Vardaro, Jennifer; Schaeffer, Moritz & Hansen-Schirra, Silvia (2019). Translation Quality and Error Recognition in Professional Neural Machine Translation Post-Editing. *Informatics*, 6(3). DOI: 10.3390/informatics6030041.

Vega Expósito, Miguel (2024). Análisis cuantitativo de traducciones del inglés al español poseditadas por traductores institucionales. *Entreculturas*, 14, 80–99. DOI: 10.24310/ertci.14.2024.17678.