Rivista Internazionale di Tecnica della Traduzione / International Journal of Translation 2023 (25), 147-165
DOI: 10.13137/2421-6763/35502 ISSN 1722-5906 (print) ISSN 2421-6763 (online)
https://www.openstarts.units.it/dspace/handle/10077/2503

"To Use or Not to Use MT?": Some Insights into Trust and Reliance From the LeMaTTT Project

Carla Quinci
University of Padova
carla.quinci@unipd.it

Abstract

This paper explores translation trainees' attitudes toward MT in terms of trust and reliance. The data were collected within the LeMaTTT project, an empirical investigation of the potential impact of neural machine translation (NMT) on the development of info-mining and thematic competences in legal translator trainees. The sample consisted of MA-level trainees with different educational backgrounds. Besides completing either a post-editing or a from-scratch translation task, they responded to a pre- and a post-task questionnaire investigating whether and how they use MT for specialised translation tasks and their trust in and perception of its output. The analysis considers within- and cross-group tendencies and takes account of the differences in (a) the types of tasks and (b) the participants' experience and training in specialised translation and post-editing. The trends concerning the use and perception of MT are also correlated with perceived task difficulty and self-assessment, as reported in the questionnaires. Results suggest that MT is perceived as a reliable tool which speeds up the translation process and provides candidate terminological equivalents, though revision is required. It appears to generally reduce the perceived difficulty of the ST while increasing the perceived quality of the TT in less experienced and competent trainees.

Keywords

neural machine translation, legal translation, translator training, trust, questionnaire data.

1. Introduction

For the translators who started working or training after the so-called "technological turn" (O'Hagan, 2013), the use of software assisting during the translation process in various ways is common practice. Machine translation (MT) – particularly the neural paradigm (NMT) – is only the most recent technology to be implemented in the translation workflow but possibly the one which has encountered the greatest resistance (Olohan, 2011; Cadwell, O'Brien and Teixeira, 2018) on the part of its professional users. However, since the early 2010s, "the image of technology-averse translators treating MT as a threat [was said to be] replaced by that of translators coexisting with an increasing integration of technology into their work environments" (O'Hagan, 2013, p. 513). The increasing quality of more recent architectures based on artificial intelligence has contributed to overcoming the stigma that has long been associated with MT and fostered the adoption of this technology by practitioners (ELIS, 2021, 2022, 2023).

The increasing use of NMT in professional translation is not always coupled with positive attitudes by qualified practitioners and still poses a series of ethical, professional, and quality issues that can undermine their trust or support their critical stance (e.g. Läubli and Orrego-Carmona, 2017; Vieira, 2020). On the other hand, the general high quality and fluency of NMT raw outputs can earn or increase the trust in this technology, especially that of less- or non-experienced translators and non-expert or less-educated users – e.g. trainee translators and non-professional users (cf. Scansani *et al.*, 2019; Scansani, 2020; Kasperė *et al.*, 2021) – who might fail to identify inaccuracies, have lower quality standards, or simply value fluency over accuracy (Martindale and Carpuat, 2018, p. 21).

The issue of trust in MT is thus key as both scepticism and overtrust – or disuse and misuse (Lee and See, 2004, p. 50) – can negatively affect performance by, respectively, neutralising the potential benefits of MT in terms of efficiency and reducing the quality of the translated text.

This paper explores the issue of trust in MT with a focus on translator trainees. Drawing on the data gathered within the LeMaTTT (Legal Machine Translation in Translator Training) research project, it seeks to gain some insights into the trust that legal translation trainees with different levels of experience and competence put in MT, their perceptions concerning its reliability and quality, and its positioning among the external resources used in specialised translation.

2. Trust and MT

Following Lee and See (2004, p. 51), trust is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability". Drawing on Ajzen and Fishbein, Lee and See (2004, p. 53) explain that

attitudes result from beliefs and perceptions and can determine specific intentions which, in turn, lie behind specific behaviour.

As an attitude, trust can affect – though not determine in itself (Lee and See, 2004, p. 51) – the user's reliance on specific systems, e.g. MT, as "[p]eople tend to rely on automation they trust and tend to reject automation they do not" (Lee and See, 2004, p. 51; Scansani, 2020). Trust in automation is based on three main factors: (a) process, i.e. understanding the functioning and internal mechanisms of a system, (b) performance, i.e. observing the behaviour of the system, and (c) purpose, i.e the reason why it is used (Lee and See, 2004, pp. 59, 67).

Trust based on process is less crucial in the use of structured, stable, and simple technology, while it is fundamental to rely on complex systems whose functioning is obscure or unknown (Lee and See, 2004, p. 52). This scenario is particularly relevant to NMT, the current state-of-the-art paradigm based on neural networks, whose outputs tend to be unpredictable because of the very nature of the system. Unlike previous paradigms, which were mostly dependent on human intervention for both development and training, NMT can autonomously improve its performance by learning from the translations it helps produce. Most importantly, what happens in its hidden layers of nodes remains substantially unknown. Hence, more than other paradigms, NMT requires a certain level of trust from its users.

The fact that trust is higher and less necessary to achieve reliance when the user understands the functioning of a system is indirectly confirmed by Scansani et al. (2019, p.78) and Scansani (2020), who correlate translation trainees' positive attitude toward MT and their ability to interact with such technology to the knowledge – and experience – they have acquired during their academic path. Other supporting evidence is provided by research on DGT translators' attitudes. Rossi (2019, p. 189) observed "a significant correlation between fear (i.e. a low degree of security) and knowledge of MT [as] the translators who perceived MT as a threat were regularly those with the lowest scores for MT knowledge". Cadwell, O'Brien and Teixeira (2018) found that both DGT translators and in-house translators working at Alpha tended to trust more translation memories (TMs) than MT. Translators at Alpha trusted more their TMs than MT system because these were fed by their own translations and "every segment comes with information about when the entry was created, who created it, which TM it comes from" (Cadwell, O'Brien and Teixeira, 2018, p. 315). DGT translators were instead more trustful towards MT, but mainly because they knew that the engine was based on their previous works. Furthermore, poor quality appeared to be more tolerated in TMs than MT as the types of errors found in TMs were considered more consistent and discrepancies in TM matches are highlighted, while MT errors were deemed to be "unpredictable, inconsistent and foster distrust" (Cadwell, O'Brien and Teixeira, 2018, p. 314). DGT translators' higher trust in MT can also be connected to the fact that, unlike Alpha translators, they are somehow involved in the development process and use one engine per language pair (Cadwell, O'Brien and Teixeira, 2018, p. 315), which ensures greater customisation and quality,

and ultimately leads to a greater adoption of MT segments (Cadwell, O'Brien and Teixeira, 2018, p. 312).

The same study provides evidence also on how trust can be derived from performance, rather than process. The investigation revealed that the DGT translators who had previously used lower-quality MT systems based on older paradigms were surprised by the advances of new MT technologies and were open to their implementation (Cadwell, O'Brien and Teixeira, 2018, p. 315). This proves that trust is not a static attitude, but rather one that dynamically interacts with automation performance (Yang *et al*., 2017). Precisely, bad interactions more significantly impact trust than positive ones, and "trust is more resilient if automation reliability starts high and declines than if it starts low and increases" (Lee and See, 2004, p. 72; cf. Yang *et al*., 2017, p. 409). Moreover, irrespective of their magnitude, faults are more tolerated if consistent, i.e. if predictable. This is supported by Cadwell, O'Brien and Teixeira's investigation (2018, p. 312), in which DGT translators' greater trust in and reliance on MT might be explained also by the reported predictability of terminological errors by the EU MT systems.

Since trust grows with information and experience, it is generally first connected to purpose and process, but eventually derives mostly from performance, i.e. from direct experience. The information at the basis of trust can be acquired through analytic, analogical, and affective processes, i.e. through, respectively, "rationally derived assessment of costs and benefits", "analogical judgments that link levels of trust to characteristics of the agent and environmental context", and "emotional responses" (Lee and See, 2004, p. 61).

Analytic processes are made visible when translators express their stance on the use of MT, its benefits and drawbacks. For instance, increased efficiency and the presence of a draft target-language version kickstarting the translation process or providing alternative wordings are regarded as beneficial, while the potential reduction of the translator's creativity or ability to assess quality, laziness, insufficient quality of the raw output, and reduced rates are considered as MT's main downsides (Cadwell, O'Brien and Teixeira, 2018, p. 312; Rossi, 2019, p. 190; Vieira, 2020, p. 14; Liu *et al*., 2022, pp. 8, 11, 13). An analytic process is exemplified in Dorst, Valdez and Jongste (2023, p. 55), when a participant claimed to determine whether to use MT or not "by looking at the syntax, grammar and tone of the MT output: if there's a correct/good base to work on, then it is worth the effort of adjusting it stylistically/terminologically".

Analogical judgments emerge, for instance, when translators stress that MT is not suitable for all types of tasks and/or language pairs (Cadwell, O'Brien and Teixeira, 2018, p. 311; Rossi, 2019, p. 190; Vieira, 2020, p. 15; Pastor, 2021, p. 55) and can be used for domains that are generally considered less suitable for automation, e.g. law, only if the source text (ST) is highly repetitive or standardised (Dorst, Valdez and Jongste, 2023, p. 54).

Emotional responses were also attested in previous research. In the study by Cadwell, O'Brien and Teixeira (2018, p. 313), Alpha translators, who had proven

more sceptical and resistant to the use of MT, also appeared "concerned about the lack of humanity and lower levels of enjoyment that working more with MT". Positive affective processes were also observed, e.g. when a participant acknowledged how MT made him/her faster and more powerful (Cadwell, O'Brien and Teixeira, 2018, p. 316). MT was also found to serve as a sort of safety net for translator trainees without which they feel less confident or more stressed (Liu *et al.*, 2022, p. 14), but it is perceived as less rewarding than human translation, which generally remains trainee and professional translators' favourite working method (Gaspari *et al.*, 2014; Daems, 2016, pp. 159–160).

Naturally, emotional responses can also be irrational and based on incorrect perceptions, which can affect attitude, and consequently behaviour, toward MT. Several studies found that translators do not perceive a change, or even perceive a decrease, in their productivity when post-editing. The study by Gaspari et al. (2014, p. 70) is a case in point as, despite the proven productivity gains of post-editing vs from-scratch translation, it revealed a bias in favour of from-scratch translation in different translation directions and levels of perception, i.e. speed, effort and favourite working method. Analogously, 45% of the professional translators participating in the study by Guerberof Arenas (2013, pp. 77, 83) did not perceive any increase in their productivity, with one translator reporting an alleged decrease, which was however countered by empirical evidence. A similar trend was also observed by Daems (2016, p. 160), as before participating in her empirical study, both trainee and professional translators believed post-editing to be as fast as, or even slower than, human translation, which was again proven untrue by the data. This false perception is also reported by Cadwell, O'Brien and Teixeira (2018, pp. 303, 311), with special reference to the most reluctant MT users in their study, i.e. Alpha translators.

Alongside emotions and perceptions, predisposition and self-confidence deriving from education and/or expertise can combine with trust and influence the choice to use or not to use MT. "If a user is self-confident and has a low level of trust in automation, he/she is likely not to rely on the latter and vice versa" (Scansani, 2020, p. 31); conversely, a low level of self-confidence might increase the perceived risk of lower quality and consequently increase the trust in the MT system. This might explain the greater openness toward MT observed in trainee translators as opposed to professionals, who "are presumably confident they are capable of delivering a high-quality translation, regardless of translation method" (Daems, 2016, pp. 159–160). Self-confidence was also discriminating for trusting MT in non-professional users: Kasperė et al. (2021, p. 13) observed a correlation between non-professional users' level of education and their reliance on MT as the respondents having (incomplete) secondary or higher education reportedly used MT mostly as a support for their own translations and tended to edit the translation more than the respondents with vocational training, who mostly use the output with no edits.

3. The potential influence of trust in MT on the development of thematic and info-mining competences

The inherent differences that distinguish translation from revision and post-editing – and, parallelly, translation competence from revision and post-editing competences – in terms of information needs and expertise or experience in a subject field can result in different research patterns and use of external resources by translators (Witczak, 2021, pp. 165–167). Unlike unaided translation, post-editing implies the presence of a machine-translated text to be verified. Experienced post-editors can thus focus their attention only on the segments, phrases or terms which appear faulty or inadequate and limit their searches to checking procedures. The number and types of searches – as well as the number and types of resources – necessary for such task might thus be more limited than those needed in unaided translation (Daems, 2016, p. 156; cf. Quinci, forthcoming), especially for non-experts, e.g. trainee translators.

Using MT in a training setting would imply that students are exposed to specific information needs, which call for the use of selected reference material and less varied types of queries as compared to unaided translation. The use of lexicographic resources, e.g. specialised dictionaries, termbases or glossaries, can generally suffice to determine the accuracy of machine-translated specialised terms and phrases, even when these are unknown. If the target-language term is associated with the source-language term also in one or more of such resources, the solution is likely to be considered acceptable. Further reference materials and more sophisticated searches should be avoided so as not to neutralise the efficiency gains implied in post-editing and would only be necessary in the case of a mismatch between the equivalents proposed by lexicographic resources. The choice to accept or edit the raw output would thus depend solely on external *linguistic* support and knowledge.

The need for finding – as opposed to verifying – a target-language equivalent for a specialised term or phrase can instead involve a more varied approach to info-mining. First, research can be needed to understand the source text (ST) as specialised terms and phrases can often be opaque, particularly to non-experts. In this case, monolingual specialised dictionaries can be used, but research often extends to other text-based (vs lexicographic) content-oriented resources, e.g. specialised encyclopaedias and other background texts. One might argue that this practice is not common among novice trainees (Way, 2012, p. 44); yet, should the non-expert translator fail to perform such preliminary research and immediately turn to lexicographic resources to look up for target-language equivalents, text-based resources, e.g. parallel and background tests, would be needed to select the most appropriate one or verify conceptual matching, adequacy, and suitability for purpose. It is precisely the need to select one among the different options that prompts content-oriented searches aimed at gaining the necessary thematic knowledge for determining the solution to be implemented.

Given the increasing use of MT by professional translators (Pielmeier and O'Mara, 2020; ELIS, 2021, 2022), the implementation of MT in translator training and the acquisition of post-editing competence are necessary to ensure that students are well-

trained to enter the translation market (see Romaniuk-Cholewska, 2021, p. 209). Yet, since over-trust in and overreliance on MT are typical of users with scarce knowledge of the process and lower levels of education and expertise (Section 2), prospective translators at an early stage of their training might overtrust MT and accept machine-generated solutions after making very little or no research work. Using MT prior to acquiring background knowledge in the subject field and info-mining skills might thus hamper the development of thematic and info-mining competences in trainee translators. However, evidence in this field is rather limited and contradictory, with the range of resources being either associated with (Daems, 2016, p. 156; Kuznik, 2017) or unrelated to (Witczak, 2021) the participants' competence and/or the task type.

4. The LeMaTTT project: objectives, design, and methods

The LeMaTTT project is a simulated longitudinal empirical study aimed at investigating the potential influence of MT on the translation processes of legal translator trainees. Precisely, it seeks to observe whether experience and competence in the subject field and the use of automation have an impact on research patterns and translation quality.

The empirical phase consisted of the post-editing or from-scratch translation of a legal text by a heterogenous sample of MA translation trainees at the University of Padova, which includes a cohort of 48 first-year students (G1) with limited experience in specialised translation and no machine translation literacy, and a cohort of 104 second-year students (G2) having completed a one-semester course in legal translation and possessing basic training in translation technologies, including MT and post-editing. Each cohort was divided into two groups depending on the assigned task, i.e. post-editing (MT group) or from-scratch translation (FS group).

The assignments differed in the translation procedure (i.e. PE vs FS translation) but involved the same language combination (English to Italian) and ST, i.e. a power of attorney. The trainees' screen activity was recorded via *Flashback*, and their translations were collected and stored in *Google Drive*. Cross-cohort and cross-group analyses allow us to determine the weight that thematic knowledge and training in legal translation, on the one hand, and the type of task, on the other, have on trainees' processes and products.

The assignment was preceded and followed by two self-administered questionnaires via *Google form*, which provided the data analysed in this paper. Given the different types of tasks, two versions were developed for each questionnaire. The pre-task questionnaire for FS groups was comprised of 24 questions – yes/no, multiple-choice, open-ended, and five-point Likert scale – which were also answered by MT groups together with one additional question about the expected reliance on MT. Besides their identification code, which was used for anonymisation purposes, the respondents had to indicate whether they possessed previous experience in legal translation (Yes/No), and for which text types, and knowledge in the legal field (Yes/No) by specifying if

this had been acquired through academic courses or working/personal experiences. Further, they had to indicate how often they generally use specific resources when translating specialised texts; these included: dictionaries, which were distinguished as mono- vs bilingual and general vs specialised, glossaries, corpora, parallel texts, machine translation, and others to be specified. The respondents also had to specify to what extent each resource would presumably be used during the assignment. These questions aimed to investigate the role of automation in their translation processes as compared to that of more traditional resources. Their use of MT was further investigated by asking whether they generally check the MT output before using it – and if not, why. The additional question for MT groups asked them to indicate to what extent they expected to rely on MT when completing the assignment.

The two versions of the post-task questionnaire were instead more tailored to the specific task and included 9 questions for FS students and 10 for MT students; these were yes/no, multiple-choice, open-ended, five-point or ten-point Likert scale questions. The questions common to both versions were about the respondents' identification code and perception of the adequacy of the time allowed, the text difficulty, and self-assessment; trainees were also asked whether, to what extent, and in which respect screen-activity recording had affected their behaviour. The remaining questions inquired about MT use and were therefore different for the two types of tasks. FS students were asked whether having a machine-translated target text would have made the task easier and, if so, to what extent, and whether it would have speeded up the research work needed to find correct terminology and phraseology. On the other hand, MT students had to evaluate the MT output in terms of accuracy, indicate whether they had found error spotting easy, with special reference to terminological errors, and provide their opinion about the role of MT during the task with a focus on reliability and higher speed.

5. Analysis, discussion, and triangulation of data

The following sections will provide an analysis of questionnaire data concerning the respondents' experience in legal translation and knowledge in law (5.1), the frequency with which external resources are generally used for translating specialised texts and the extent to which they assumed to resort to each of them for the specific task at hand (5.2), their perceptions about the adequacy of the time allowed for the task, the difficulty of the ST, and self-assessment (5.3), the impact that MT had or might have had on their performances (5.4), and the extent to which these were influenced by screen-activity recording (5.5). The analysis is restricted to 40 participants – i.e. 20 per cohort and 10 per group (G1FS, G1MT, G2FS, and G2MT) – so as to triangulate the data concerning trust and reliance with those concerning the use of time and external resources by the same participants as analysed in Quinci (forthcoming).

The initial profiling questions were meant to confirm the supposed difference in the two cohorts' experience and knowledge in the legal field. The data in Figure 1 show that 80% of first-year students (in light and dark yellow) had never translated a legal text and possessed no previous knowledge in the legal field. 20% of them (N5, N8, N10 from G1MT and N30 from G1FS) had instead already translated legal texts, but these belonged to genres other than the one selected for the assignment, i.e. contracts, certificates, and judicial documentation. Two of these students (N5 and N30) had also attended academic law courses, as two other trainees in G1 also indicated (N13 from G1MT and N35 from G1FS). Given the general inexperience in legal translation of G1 trainees, the data produced by these students are unlikely to affect overall results.

Conversely, G2 students (in light and dark blue) generally had previous experience in legal translation; the only 2 students from G2MT (I4 and I7) indicating zero experience had attended the course in legal translation but possibly failed to complete the training activities assigned during the course. The range of legal documents translated by G2 trainees was also wider as compared to G1 and included contracts, certificates, last wills and testaments, affidavits, privacy policies, and orders, i.e. the text types that had been covered during the course. Almost half of G2 also possessed some knowledge in the legal field: six students had attended academic law courses, two of them had had previous working experience in the field, and one had acquired such knowledge outside both the academic and work settings. The remaining 11 students had no knowledge in the field except that acquired during the legal translation course.

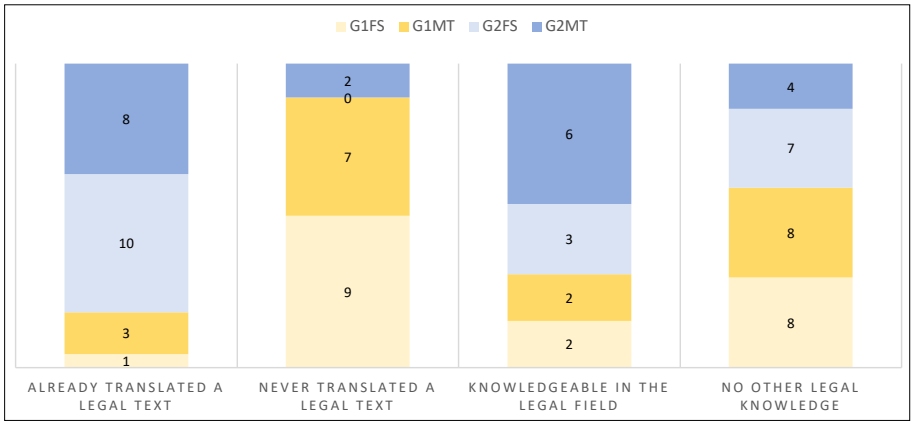


Figure 1. Previous experience in legal translation and knowledge in the legal field

As for the use of external resources, the respondents had to specify (a) how often they generally resort to dictionaries, parallel texts, corpora, glossaries, and MT when translating specialised texts, and (b) how often they expected to use of each of them for the assignment at hand on a five-point Likert scale. Figure 2 shows the weighted means calculated per cohort and group; the closer the value to 5, the higher the frequency with which that resource is generally used in specialised translation assignments.

The results concerning the general frequency of use (in solid colours) show similar frequencies for specific resources while suggesting different approaches to the use of others. Mono- and bilingual general dictionaries appear to be used equally frequently by G1 ($\bar{x}_w$= 3.50 for G1FS and 4.15 for G2MT) and G2 ($\bar{x}_w$= 3.50 for G2FS and 4.05 for G2MT), with bilingual dictionaries proving to be the most frequently used resource, as evidenced also by other studies (Krings, 1986; Künzli, 2001; Torrejón and Rico, 2013). Similarly, MT is equally used by G1($\bar{x}_w$= 3.50) and G2 ($\bar{x}_w$= 3.35), but the prominence of such technology within the range of external resources is different for the two groups. For G1, it also represents the second most frequently used resource together with general monolingual dictionaries, while for G2 it is the least frequently used tool given that G2's weighted mean is equal to or lower than the ones scored by the other resources. In the remaining cases, the two cohorts display differing tendencies, with G2 always outscoring G1, especially as concerns corpora (+1.00), specialised bilingual dictionaries (+0.80), and parallel texts (+0.70).

This suggests a more diversified approach to info-mining by G2, which is more evident when conducting a more fine-grained analysis of the students' selected options. The distributions of G1's and G2's values across the five points of the Lickert scale (Figure 3) are both asymmetric, but G2's is considerably more negatively skewed than G1's. Hence, G2 trainees have more often indicated a higher frequency of use for the different resources, with 4, 5 and 3 being the most selected weights, while G1's most preferred options were 4, 3, and 2, which suggests a more limited use of resources during the translation process. This is in line with the analysis of process data (Quinci, forthcoming) showing that G2 made on average a higher number of searches by using a more diverse set of resources.

Attention also deserves the cross-group variation in the two cohorts as concerns general dictionaries (both mono- and bilingual), which appear to be more frequently used by FS groups, and glossaries, which would be used more frequently by G1MT and G2FS. These reportedly different attitudes appear to have only partially influenced FS trainees' behaviour during the task, i.e. only with reference to general bilingual dictionaries, which have been consulted far more frequently by FS than MT students. General monolingual dictionaries were instead used more frequently by G1FS than G1MT, but equally frequently by G2FS and G2MT, and glossaries were only used by MT groups (Quinci, forthcoming). This might indicate (a) that general research patterns are only partially connected to the trainees' habits but are

largely moulded by the specific task or (b) that trainees are only partially aware of the extent to which they actually use specific resources.

The former hypothesis appears to be backed up by the expected frequency of use of external resources during the legal translation assignment (Figure 2, bars in shaded colours). When contextualised to the task at hand, the students' responses partially changed, with the two cohorts largely scoring comparable values. Specifically, the frequencies of use of general dictionaries (both mono- and bilingual) of all cohorts and groups decreased, while those concerning the other resources increased with the minor exceptions of G1MT and G2FS as concerns glossaries.

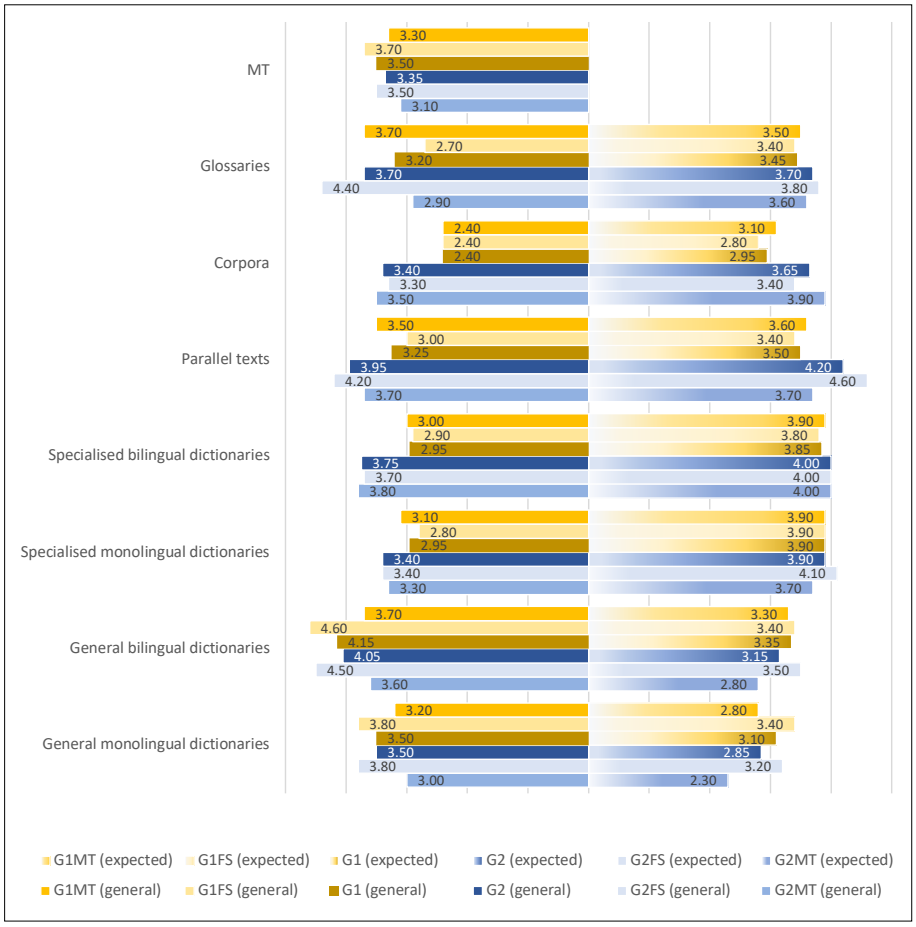
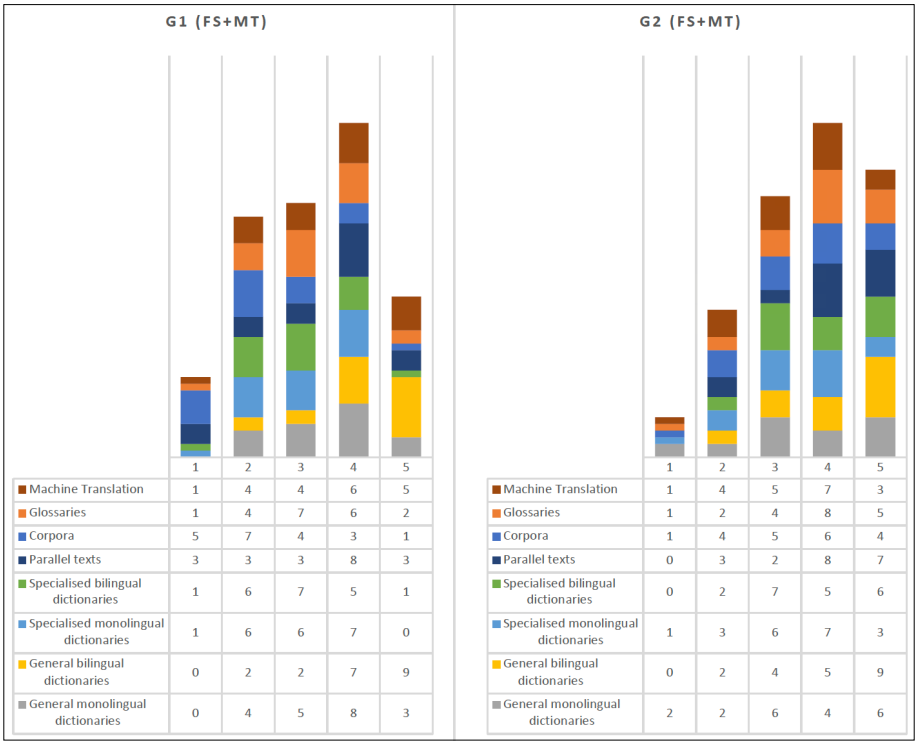


Figure 2. General frequency of use of external resources vs expected frequency of use during the assignment.



	1	2	3	4	5
Machine Translation	1	4	4	6	5
Glossaries	1	4	7	6	2
Corpora	5	7	4	3	1
Parallel texts	3	3	3	8	3
Specialised bilingual dictionaries	1	6	7	5	1
Specialised monolingual dictionaries	1	6	6	7	0
General bilingual dictionaries	0	2	2	7	9
General monolingual dictionaries	0	4	5	8	3

	1	2	3	4	5
Machine Translation	1	4	5	7	3
Glossaries	1	2	4	8	5
Corpora	1	4	5	6	4
Parallel texts	0	3	2	8	7
Specialised bilingual dictionaries	0	2	7	5	6
Specialised monolingual dictionaries	1	3	6	7	3
General bilingual dictionaries	0	2	4	5	9
General monolingual dictionaries	2	2	6	4	6

Figure 3. General frequency of use of external resources by cohort

On the other hand, hypothesis (b) above is substantiated by the combined analysis of questionnaire and process data. When asked whether they generally resort to other resources aside from those mentioned in the questionnaires, all the respondents but two answered no. Only I56 and I65 from G2MT indicated that they also use, respectively, video resources for terminological issues and operators in Google searches. Process data (Quinci, forthcoming) indicate instead G2 trainees used a wider array of different resources, including language fora, law firm websites, institutional websites, concordancers, general and specialised encyclopaedias, and background texts from websites of professionals or entities operating in the legal field. This partially holds also for G1 as the other resources they consulted are limited to concordancers, language fora, and background texts from websites relating to the legal domain. Yet, when asked about the potential use of other resources for the legal translation assignment, more students suggested other options, i.e. "texts on the topic" (N10, G1MT), machine translation (N4, G1MT), encyclopaedias (I64, G2FS), operators in Google searches (I65, G2FS), general research on the web and the notes taken during the course (I11, G2MT). Interestingly, a student from G1MT (N7) replied that the use of other resources would be determined by the needs implied in the task, which points back to our hypothesis (a). Both hypotheses thus seem to be plausible.

The impact of MT on the students' performance is highlighted by the data concerning the perceived adequacy of the time allowed, the perceived ST difficulty, and – though less remarkably – self-assessment scores. As shown in Figure 4, MT groups experienced a reduced time pressure than FS groups irrespective of the cohort as their mean and median values are higher than the other groups' and, most importantly, their interquartile ranges (IQRs) are considerably wider and cover higher weights of the Likert scale (G1MT and G2MT=3.00-4.25) than those of FS groups (G1FS and G2FS=3.00-3.25). This supports the assumption suggested by process data whereby the higher time pressure perceived by FS groups' was inferred from their longer drafting phases and shorter orientation and revision phases, and higher number of searches in approximately the same time as compared to MT groups (Quinci, forthcoming).

MT also appears to reduce the ST perceived difficulty as testified by MT groups' lower means, with special reference to G2MT (G1MT: $\bar{x}$=4.30; G1FS: $\bar{x}$=4.40; G2MT: $\bar{x}$=4.00; G2FS: $\bar{x}$=4.40), and G1MT's wider IQR as compared to both FS groups.

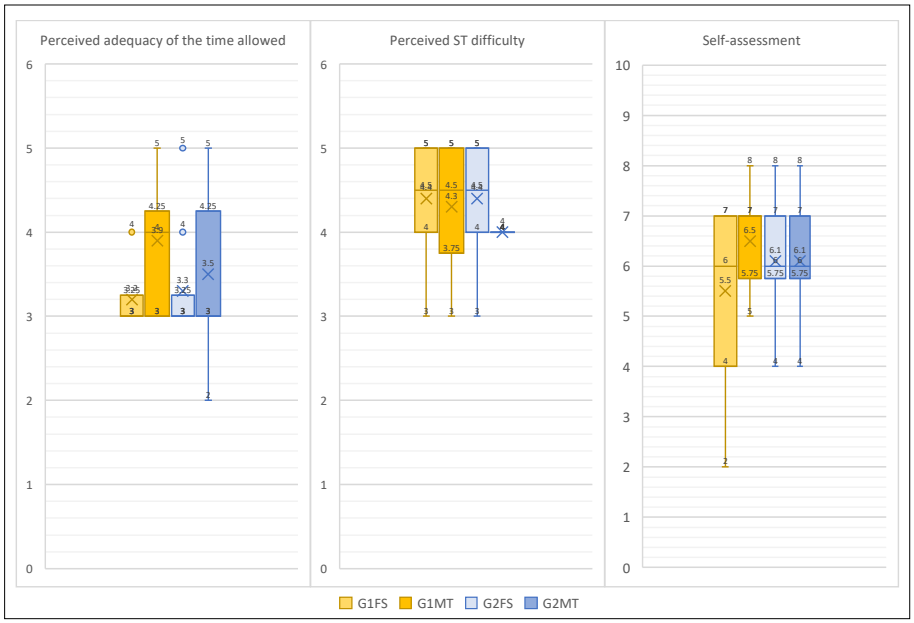


Figure 4. Questionnaire data about time, text difficulty, and self-assessment

The tempering effect of MT on the perception of time and difficulty is only partially mirrored by self-assessment, since it applies to G1 (G1FS: IQR= from 4.00 to 7.00 and $\bar{x}$=5.50; G1MT: IQR= from 5.75 to 7.00 and $\bar{x}$=6.50), while G2FS and G2MT have the same distribution (IQR= from 5.75 to 7 and $\bar{x}$=6.10). As suggested by previous research (Daems, 2016, pp. 159–160; Scansani, 2020a, p. 31; Kasperė *et al.*, 2021, p. 13; cf. Section 2), education and self-confidence also play a role in the users' reliance on and trust in automation, which can explain why MT does not affect per-

ceived quality in the participants who had received training in legal translation, while it substantially raised unexpert trainees'. This might have serious consequences for training as inexperienced trainees might feel capable of performing tasks that are out of their reach given that they can achieve (perceived) high quality with no or little effort also in unknown and challenging domains, e.g. legal translation. Unawareness might also reinforce their trust in and reliance on automation, whose performance they are unable to knowledgeably assess due to their lack of thematic competence.

5.4 Expected, actual and potential impact of MT

The questions concerning the use of MT were tailored to the specific tasks and explored MT students' opinions about its actual impact on their performances and FS students' views about the potential assistance they might have received from it.

The opinions of both MT groups (Figure 5) appear to be largely overlapping as concerns the difficulty of error spotting and the overall positive evaluation of MT (cf. Scansani *et al.*, 2019, pp. 75, 78). For 60-70% of students in either group (G1MT=7; G2MT=6), detecting translation errors – and terminological errors – in the raw output was easy only in some cases, 1 student per group found errors difficult to spot, and 2-3 students per group found that most errors required extensive research work to be identified. Both groups generally perceived MT as "a reliable resource which speeded up the translation process" (G1MT=90%; G2MT=70%). Only two trainees, one per group, found MT misleading and would have preferred to translate from scratch. Finally, two G2MT trainees selected the option "other": one specified that MT is "a good starting point but it needs a lot of additional work carried out by the translator" (I5), while the other found it "maybe misleading, but [he/she] would have not preferred to translate from scratch since […] with the help of MT it is possible to save time" (I2). This raises G2MT's percentage of positive answers to 80% and that of negative to 20%, making this group slightly more critical about the role of MT than G1MT.

In previous research (Yang and Wang, 2019, p. 122) perceived usefulness (in our case, the reliability associated with efficiency) appeared to be a stronger predictor of behaviour intention than perceived ease of use (in our case, the ease with which errors are spotted). Thus, G2MT's trust in MT might be considered as slightly lower than G1MT's, as substantiates the analysis of perceived accuracy.

Despite defining MT as generally reliable, G2MT was less satisfied with the accuracy of the output than G1MT, with 60% of G2MT respondents finding it "good, although it required some editing" and 40% claiming it was "poor as it required extensive editing". Interestingly, prior to the task, G2MT trainees were slightly more trustful in MT than G1MT as 90% supposed that they would rely on MT either moderately (3 on the five-point Likert scale; 70%) or a lot (4; 20%), while all G1MT students opted for 2 (20%) or 3 (80%). G2MT trainees' greater thematic knowledge and legal translation competence enabled them to detect more errors and thus negatively influenced their perception of MT accuracy, which ultimately made them more critical toward MT as compared to less experienced and competent first-year trainees (cf. Section 2).

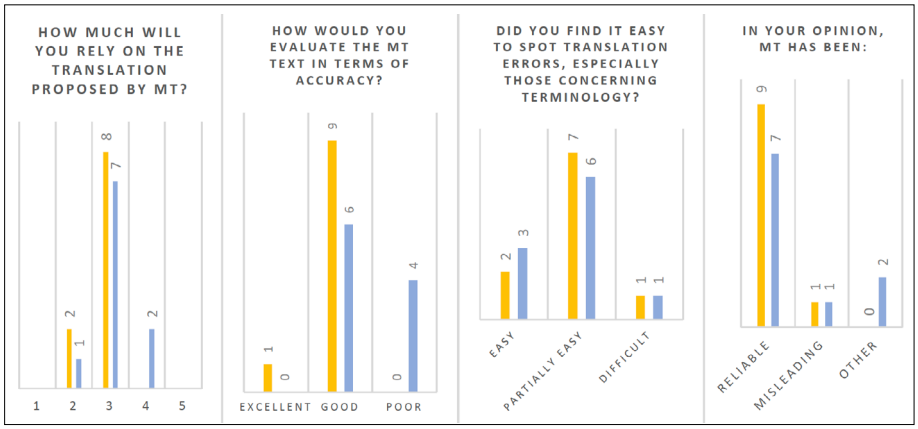


Figure 5. Data concerning the expected and actual impact of MT on the assignment (G1MT in yellow and G2MT in blue).

FS students were asked about the potential impact MT might have had on their performance. Both FS groups indicated that MT would have made the task easier, but its impact was believed to be greater by G2FS students, with 70% of them opting for "much easier" vs 40% of G1FS trainees. Its perceived potential impact thus parallels thematic competence and experience in legal translation: MT is considered as an aid by both trained and untrained legal translators, but one that is unable to compensate for the lack of education and experience in the legal field.

Both groups agree on the efficiency ensured by MT during terminological research, which is perceived as substantially high by most students (G1FS= 60%; G2FS= 50%). Presumably, as highlighted by a G2MT student above, the presence of target-language terminological equivalents in the text is viewed as "a good starting point" (I5) prompting and guiding further research.

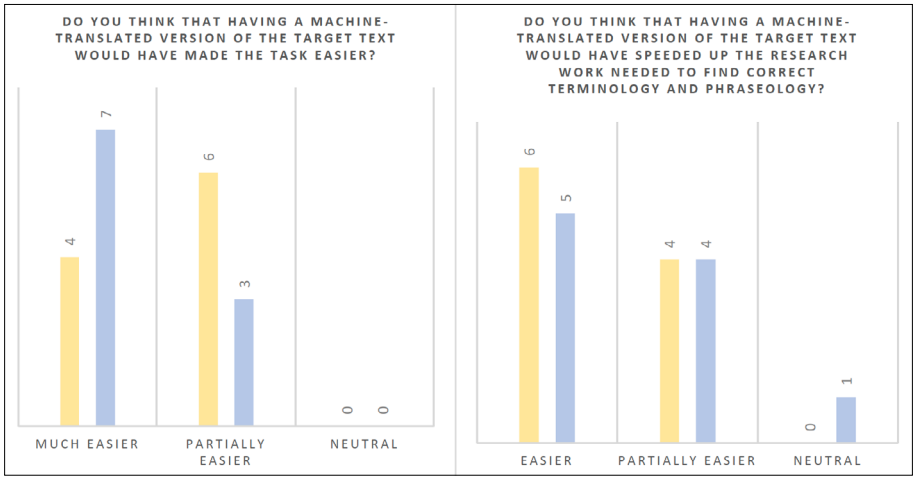


Figure 6. Data concerning the potential impact of MT on the assignment (G1FS in light yellow and G2FS in light blue).

The last questions of the survey sought to determine to what extent screen-activity recording might have affected the participants' performances. The medium to low means and medians of all groups, which range from 2.5 to 5.6 out of 10 (Figure 7), suggest that the impact of recording procedures was minor. Interestingly, it is particularly reduced for MT vs FS groups, and especially for G1MT ($\bar{x}$=3.00; M=2.50; SD=2.26), which scored considerably lower than the other groups ($\bar{x}$=4.80-5.60; M=5.00-5.50; SD=1.81-2.26). Conversely, FS groups display the highest maximum values (9) and lower and upper quartiles (Q2= 3.5 for G1FS and 4 for G2FS; Q3= 6.5 for G1FS and 7.25 for G2FS), with G2FS scoring always highest.

These results suggest that MT increased the participants' self-confidence and trust. It might be perceived as a shield against any criticism over the translator's textual and strategic choices, which are ultimately largely determined by the pre-translated text. In from-scratch translation, trainees were instead more exposed and possibly feared for their reputation, particularly G2FS trainees, who had been trained by the researcher conducting the experiment and presumably felt they had somehow to prove their competence.

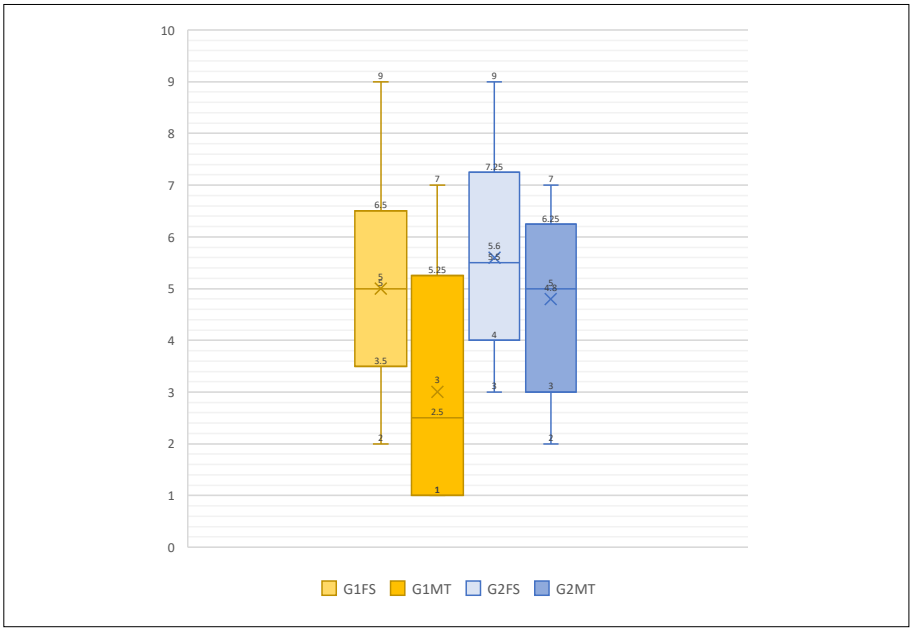


Figure 7. Perceived influence of recording on a scale of 1 to 10.

These conclusions are supported by the responses to the open-ended questions concerning the impact of screen-recording, as:

- MT students reportedly experienced general anxiety and psychological pressure due to the test-like setting (N13, I5, I7, I8, I11) or the fear of making mista-

kes in the experiment procedure (N13), while FS students were more explicitly concerned with possible translation errors (N29, N31, N33) and especially the choice of reliable and suitable external resources (I59, I60, I61, I62);
- most G2 students were concerned about their info-mining processes (I1, I10, I59, I60, I61, I62) or time management (I59) and put extra effort and attention into the task (I65).

6. Conclusions and future work

This paper explored the impact of trust and reliance on MT in MA-level (legal) translation trainees by drawing on the questionnaire data produced within the LeMaTTT project. The analysis mapped the trends emerged from the trainees' responses onto their levels of experience and training in legal translation and post-editing, on the one hand, and onto the specific task they completed, i.e. post-editing vs from-scratch translation, on the other hand. The findings suggest that MT is generally used in specialised translation assignments but for the least experienced trainees (G1), it represents the most frequently used resource after dictionaries, while for more competent trainees (G2), it is just one of the various tools they resort to.

MT had a visible impact on the perception of the time allowed and the difficulty of the ST. MT students worked under a lower pressure and perceived the ST as (slightly) less difficult than FS students. Self-assessment was instead only partially affected by MT, which apparently played a more crucial role for inexperienced trainees by considerably raising the self-confidence of G1MT as opposed to G1FS. The scores of G2FS and G2MT were instead comparable, which confirms the relation between lower trust and greater expertise observed in the literature.

MT trainees generally saw MT as a reliable tool, even if second-year students were initially more trustful than first-year trainees when indicating their expected reliance on MT during the task, but later proved less satisfied with its quality, probably due to their higher thematic competence and G1 trainees' inability to knowledgeably assess quality. Even if, reportedly, errors were not always easy to spot, only two students out of 20 ultimately considered MT misleading and would have preferred to translate from scratch. FS students also thought that MT would have been a useful support, albeit first-year students were slightly more sceptical about its potential impact on their performances, possibly because they believed that MT could hardly compensate for their lack of thematic competence.

Finally, the results concerning the impact of recording practices on the trainees' performances suggest a high trust in MT, which enhances trainees' self-confidence and acts as a shield against the potential criticisms raised about their textual and/or info-mining choices. FS trainees felt instead greater anxiety and were naturally more exposed, being responsible for the full translation and info-mining processes.

Future work will extend the analysis to the full sample and further triangulate these findings with the other variables investigated in the LeMaTTT project to study the impact of trust in MT on translation quality and editing distance.

References

Cadwell, P., O'Brien, S. and Teixeira, C.S.C. (2018) "Resistance and Accommodation: Factors for The (Non-) Adoption of Machine Translation Among Professional Translators", *Perspectives: Studies in Translatology*, 26:3, pp. 301–321.

Daems, J. (2016) *A Translation Robot for Each Translator? A Comparative Study of Manual Translation and Post-Editing of Machine Translations: Process, Quality And Translator Attitude*. University of Gent.

Dorst, A.G., Valdez, S. and Jongste, D. (2023) "Professional translators' and project managers' perceptions of machine translation and post-editing: a survey study", in *Proceedings of The New Trends in Translation and Technology Conference - Nettt 2022*. Ed. by S. Castilho et al., pp. 50–59.

ELIS (2021) *European Language Industry Survey 2021*. https://commission.europa.eu/system/files/2021-07/elis_2021_european_language_industry_survey.pdf, accessed on 10 October 2023.

ELIS (2022) *European Language Industry Survey 2022 - Trends, Expectations and Concerns of the European Language Industry*. https://elis-survey.org/wp-content/uploads/2022/03/ELIS-2022-report.pdf?utm_source=elis-repository&utm_medium=website&utm_campaign=elis-report22&utm_id=elis-report-22, accessed on 10 October 2023.

ELIS (2023) *European Language Industry Survey 2023 - Trends, Expectations and Concerns of the European Language Industry*. https://elis-survey.org/wp-content/uploads/2023/03/ELIS-2023-report.pdf, accessed on 10 October 2023.

Gaspari, F., Toral, A., Kumar Naskar, S., Groves, D. and Way, A. (2014) "Perception vs Reality: Measuring Machine Translation Post-Editing Productivity", in *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas (2014)*, pp. 60–72.

Guerberof Arenas, A. (2013) "What do professional translators think about post-editing?", *The Journal of Specialised Translation Issue*, 19, pp. 75–95.

Kasperė, R., Horbačauskienė, J., Motiejūnienė, J., Liubinienė, V., Patašienė, I. and Patašius, M. (2021) "Towards sustainable use of machine translation: Usability and perceived quality from the end-user perspective", *Sustainability*, 13:23, pp. 1–17.

Krings, H.P. (1986) "Translation Problems and translation Strategies of Advanced German Learners of French (L2)", in *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*. Ed. By J. House and S. Blum-Kulka, Tübingen, Narr, pp. 263–275.

Künzli, A. (2001) "Experts vs. novices: L'utilisation de sources d'information pendant le processus de traduction", *Meta*, 46:3, pp. 507–523.

Kuznik, A. (2017) "Use of Instrumental Resources", in *Researching Translation Competence by PACTE Group*. Ed. By A. Hurtado Albir, Amsterdam/Philadelphia, John Benjamins, pp. 219–241.

Läubli, S. and Orrego-Carmona, D. (2017) "When Google Trans-

late is better than Some Human Colleagues, those People are no longer Colleagues", in *Proceedings of the 39th Conference Translating and the Computer*. AsLing, pp. 59–69.

Lee, J.D. and See, K.A. (2004) "Trust in Automation: Designing for Appropriate Reliance", *Human Factors*, 46(1), pp. 50–80.

Liu, K., Kwok, H. L., Liu, J. and Cheung, A. K. F. (2022) "Sustainability and Influence of Machine Translation: Perceptions and Attitudes of Translation Instructors and Learners in Hong Kong", *Sustainability*, 14, pp. 1–29.

Martindale, M.J. and Carpuat, M. (2018) "Fluency Over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT", in *Proceedings of AMTA 2018, vol. 1: MT Research Track*, pp. 13–25.

O'Hagan, M. (2013) "The Impact of New Technologies on Translation Studies: A Technological Turn?", *The Routledge Handbook of Translation Studies*. Ed. by C. Millán and F. Batrina, Abingdon and New York: Routledge, pp. 503–518.

Olohan, M. (2011) "Translators and Translation Technology: The 'Dance of Agency'", *Translation Studies*, 4:3, pp. 342–357.

Pastor, D.G. (2021) "Introducing Machine Translation in the Translation Classroom: A Survey on Students' Attitudes and Perceptions", *Revista Tradumatica*, 19, pp. 47–65.

Pielmeier, H. and O'Mara, P. (2020) *The State of the Linguist Supply Chain: Translators and Interpreters in 2020*, *CSA Research*.

Quinci, C. (forthcoming) "The Impact of Machine Translation on the Development of Info-Mining and Thematic Competences in Legal Translation Trainees: A Focus on Time and External Resources", *The Interpreter and Translator Trainer*.

Romaniuk-Cholewska, D. (2021) "Machine Translation Post-Editing (MTPE) from the Perspective of Translation Trainees: Implications for Translation Pedagogy", in *Proceedings of the 18th Biennial Machine Translation Summit Virtual USA, August 16 - 20, 2021, Volume 2: MT Users and Providers Track*, pp. 200–210.

Rossi, C. (2019) "Uses and perceptions of machine translation at the European Commission", *JosTrans*, 31, pp. 177–200.

Scansani, R., Bernardini, S., Ferraresi, A. and Bentivogli, L. (2019) "Do translator trainees trust machine translation? An experiment on post-editing and revision", *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, pp. 73–79.

Scansani, R. (2020) *Machine Translation for Institutional Academic texts: Output, Quality, Terminology Translation and Post-editor Trust*. University of Bologna.

Torrejón, E. and Rico, C. (2013) "Skills and Profile of the New Role of the Translator as MT Post-editor", *Tradumàtica: tecnologies de la traducció*, 10, p. 166–178.

Vieira, L.N. (2020) "Automation anxiety and translators", *Translation Studies*, 13:1, pp. 1–21.

Way, C. (2012) "A Discourse Analysis Approach to Legal Translator Training: More Than Words", *International Journal of Law, Language & Discourse*, 5:2, pp. 39–61.

Witczak, O. (2021) *Information Searching in the Post-Editing and Translation Process*. University of Poznań.

Yang, X.J., Unhelkar, V. V., Li, K. and Shah, J. A. (2017) "Evaluating Effects of User Experience and System Transparency on Trust in Automation", in *ACM/IEEE International Conference on Human-Robot Interaction*. IEEE Computer Society, pp. 408–416.

Yang, Y. and Wang, X. (2019) "Modeling the Intention to Use Machine Translation for Student Translators: An Extension of Technology Acceptance Model", *Computers and Education*, 133, pp. 116–126.