Ethics and machine translation

Guerberof-Arenas, Ana; Moorkens, Joss

Published in:
Towards Responsible Machine Translation

DOI:
10.1007/978-3-031-14689-3_7

Chapter 7
Ethics and Machine Translation: The End User Perspective

Ana Guerberof-Arenas and Joss Moorkens

Abstract This chapter analyses existing research on the ethical implications of using MT in translation and communication, and it describes results from usability experiments that focus on the inclusion of raw and post-edited MT in multilingual products and creative texts with an emphasis on users' feedback. It also offers suggestions on how MT content should be presented to users, readers, and consumers in general. It finally considers the ethical responsibility of all stakeholders in this new digital reality. If the ethical dimension is an ecosystem, users also have the responsibility to support products that protect language, translators, and future generations.

Keywords Ethics · Machine translation · Usability · User reception · Translation reception · Ethical responsibility · Sustainability

7.1 Introduction

In 2016, ten years after it was launched, the world's biggest machine translation (MT) producer, Google Translate, announced that it generated over 143 billion words per day (Pichai 2016). We can safely assume that this output has subsequently increased, including many text types translated for a wide variety of users. Why is it that MT use has become so widespread? There are two primary positions on this: (a) the technological determinist view that the time has come for this technology, i.e. from the natural evolution of the field at that time, and (b) the social determinist

A. Guerberof-Arenas (✉)
Guildford, UK

University of Groningen, Groningen, Netherlands
e-mail: a.guerberof.arenas@rug.nl

J. Moorkens
Dublin City University, Dublin, Ireland
e-mail: joss.moorkens@dcu.ie

H. Moniz, C. Parra Escartín (eds.), *Towards Responsible Machine Translation*,
Machine Translation: Technologies and Applications 4,
https://doi.org/10.1007/978-3-031-14689-3_7

view that circumstances (societal, technological, economical) are such that huge efforts were put into MT development. The former implies that MT is a small part of inevitable technological progress and it follows that MT should be put into use where possible without consideration of its sociocultural context. The sociotechnical counterargument is that the pros, cons, and repercussions of each new technology should be carefully considered by society before its implementation.

Kranzberg (1986) wrote that technology is "neither good nor bad; nor is it neutral" (p. 545). The view among ethicists and researchers in science and technology studies or STS (as summarised in Olohan 2017) is that science is not linear and deterministic, but rather that development is rooted in a worldview from which the decision as to what to develop, its intended audience, and its implementation are indivisible. This set of factors, in turn, influence the effects of technologies in use, as they reshape activities and their meaning, engendering new worlds of their own (Winner 1983). For example, as Larsonneur (2021) noted, the major MT providers are now big tech companies due to their access to resources and ubiquitous online offering. The university research groups that at one stage topped the leaderboards in competitive MT shared task events, particularly those for well-supported languages, have gradually been replaced by big tech research groups. This means that the perspective and motivation of big tech companies now drives much of MT development.[1] In other words, large corporations rather than all players in society are determining the use and suitability of MT for assimilation, where MT is served directly to the end user.

In Weaver's (1949) memo proposing MT, for example, he wants to enable communication and encourage peace between nations. He also sees translation as a problem and foreign languages as encrypted versions of English or an as-yet-undiscovered universal language (Raley 2003), an idea that Kenny et al. (2020, p. 1) call one of the 'most reductive. . . in translation history.' This is on the basis that translation as a communicative act is a much more complex process than coding and decoding language at the superficial level of the written word as opposed to the world of ideas and of communication between cultures. Nonetheless, this superficial view has often prevailed: Kenny et al. (2020) note that the notion of 'foreign as English' survived in MT literature well into the 2000s. We can see the evidence of superficiality and neutralisation in MT output, as discussed in recent literature about normalisation in MT (Čulo and Nitzke 2016; Toral 2019) and reduced lexical diversity (Vanmassenhove et al. 2019). And yet, MT also carries the utopian communicative intent of Weaver's memo, enabling effective communication for many people in many scenarios.

In this chapter, we attempt to systematically analyse the ethics of MT as an end-product, and to examine the world engendered by widely-available MT, a world that did not really exist before the advent of free, networked, and ubiquitous MT. If we were to take a stakeholder approach (see Fig. 7.1) in analysing the effects of MT on groups of people with different levels of involvement with MT (e.g., translators,

[1] See also Paullada (2020) on MT and power dynamics.

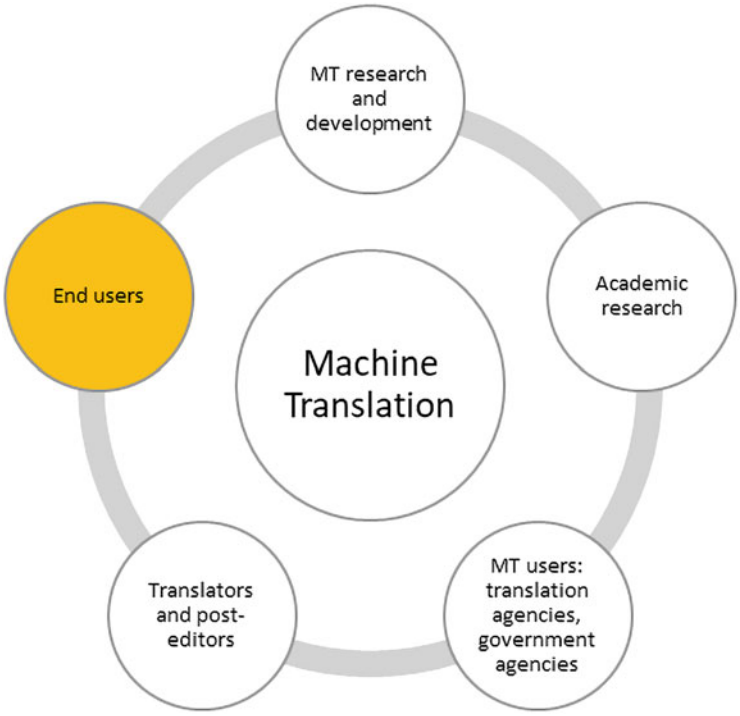

Fig. 7.1 MT stakeholder ecosystem

company shareholders, engineers, academics, end users) and to take a utilitarian position as to whether MT helped them materially or to consider whether it helped them to flourish as human beings, the results would probably be different for each user group. Here, however, we focus on end users, although we are aware that this group is multilingual and not homogeneous.

On the one hand, one could argue that the MT end user receives a lot of attention from big tech companies. Where translators are often given mixed-quality MT and expected to 'work miracles' (Thicke 2013, p. 42), end users are given up-to-date interfaces and seamless integration within internet browsers. On the other hand, there is little published user testing of these translation interfaces, and the only opportunity for feedback, if such an opportunity is offered at all, is a binary yes or no response to mark satisfaction or dissatisfaction with the translation provided, giving the impression that the text is just about worth translating, but not necessarily worth the effort of translating well. For MT providers, is no news good news? Looking at a variety of user types, Liebling et al. (2020) find that many users' needs are not catered to by contemporary mobile MT applications.

Way (2018) suggests that the level of automation for translation should relate to the perishability of the source text, but notes that new use cases are constantly being found for raw (and post-edited) MT. Canfora and Ottmann (2018) introduce a

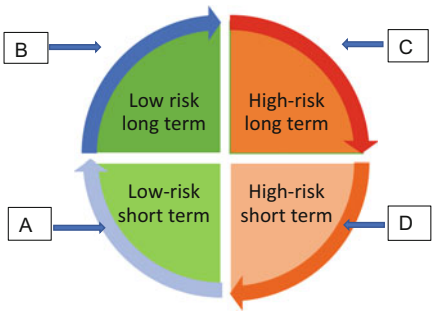

Fig. 7.2 Modelling MT use cases

second continuum of *risk*, whereby low-risk texts afford translation using MT but high-risk texts (where mistranslation may cause injury or death) require careful human revision. In Sect. 7.1 of this chapter, we consider ethics and MT as an end-product for different types of texts (see Fig. 7.2), from those that have a short shelf-life and are low risk to those that have a long shelf-life and are high risk. Of course, modelling or presuming a user perspective is not ideal, so we also hear the voices of real users testing MT in a novel long shelf-life use case in Sect. 7.2. In Sect. 7.3, we discuss the implications of our analyses and some further issues prior to conclusion.

7.2 Modelling MT Use Cases

Translation use cases with a short shelf life that present low risk (A in Fig. 7.2) are ostensibly ideal for reception of raw, 'low stakes' MT. For example, it makes little sense to hire a professional translator to translate most user-generated content such as online travel reviews or forum postings, as the reviews are likely to be superseded by newer ones within hours or days and few readers tend to read beyond the most recent postings. Similarly, online auctions are time-limited and cease to be useful as soon as the auction closes. We can only assume that most instances of low-stakes MT are positive and useful, with comprehension facilitated by MT. End users will probably get the gist of the review or auction posting, and any mistranslation will have few, if any, risks or repercussions.

The taxonomy proposed by Canfora and Ottmann (2018), categorising risks along a continuum of increasing severity, is a useful tool for evaluating the level of translation risk. These risks range from communication being impaired or impossible, loss of reputation, and financial or legal consequences, to damage to property, physical injury, and death. Following this taxonomy, there may be a possibility of communication being impaired and a loss of reputation on the part of the provider of a low-risk translation, i.e. the hosting site for reviews or sales. An example of this was the launch of an Amazon portal for Sweden, featuring mistranslations and vulgar MT errors (Hern 2020). Reputational risk is greater with mistranslation of

social media posts. An embarrassing and offensive mistranslation in 2020 led Facebook to deactivate English-Thai MT alongside a public apology (Marking 2020). Because most interactions with the average Twitter or Facebook post occur within the first 2 h or less (Rey 2014), translation needs to be timely. However, translation that maintains the original tone, content, and indexing (using the appropriate hashtags) can be difficult (Desjardins 2017), and since these posts are indefinitely searchable, they cannot be considered short shelf-life content—or low risk. If the posting has been machine translated prior to posting, the translation will be static, but if not, it will be dynamic and not controllable by the original poster, with MT applied to the post whenever it is viewed by someone using the platform in an MT-supported language (and thus may be considered part of Category B or even C in Fig. 7.2). Seemingly innocuous public social media posts may form the basis for major decisions about users that impact on the future opportunities and wellbeing. The United States (US) Citizen and Immigration Service, for example, are reported to instruct their officers to 'screen refugees' social media posts using commonly available translation tools such as Google Translate' (Doc Society v. Pompeo 2019), despite the tendency of such large language models to encode societal bias (Bender et al. 2021). The likelihood of mistranslation is greater for low-resource language pairs due to data sparsity and, according to Wang et al. (2021), possible malicious attacks on neural MT (NMT) systems to alter the output.

End users also interact with free online MT platforms to access quick, low-risk translation (A in Fig. 7.2) (Nurminen 2018). This may be to comprehend a restaurant menu, to read signs and instructions in an unfamiliar language, or for social interaction, such as fan conversations or bar and restaurant orders in the example of the use of Google Translate at the 2018 Russia World Cup (Smith 2018). In this situation, the risks are a little higher, especially for users with little MT literacy, and providers usually exclude liability (Canfora and Ottmann 2020). Bowker and Ciro (2019) introduced the concept of MT literacy to include a basic understanding of how MT works, the ability to craft appropriate input text, and to edit MT output for accuracy and readability. The user with little MT literacy may assume that any translation produced is clumsy but accurate, with possible risky results if they eat an ingredient to which they are allergic or follow a mistranslated instruction (quickly joining Category D in Fig. 7.2). With regular use of free online MT, a user might improve their MT literacy, becoming familiar with the quality to be expected of output and understanding the types of texts that are likely to be well translated. For those users, MT may provide a gateway for civic participation and for access to information (Nurminen and Koponen 2020).

This appears to be the case for the UK-based Syrian asylum seekers in Vollmer's (2020) research, who use a smartphone with MT as one of their digital literacy tools for communication and, in one case, to practice for a driving theory test. Ciribuco (2020) reports similar pragmatic use of smartphones and MT among asylum seekers in Italy. However, in the latter case most research participants appear to use MT uncritically, other than one participant who says that he will not use it again after realising that it had provided him with an inaccurate translation. The use cases presented by Ciribuco (2020), including use of MT when watching TV, conversing

with teachers, and working in a garage, probably carry little risk (A and B in Fig. 7.2), but there is a danger that uncritical use of MT will mean insufficient discrimination between low- and high-risk uses. Ciribuco (2020) writes of the 'need' for translation for 'survival', and there may be instances when the availability of online or mobile MT is crucial as a translator or informal mediator is just not available, in which case MT fulfils a communication function. This raises a risk of overreliance on MT, however. The non-English speaking immigrants to the US interviewed by Liebling et al. (2020) experienced mistranslations that were inappropriate or dangerously inaccurate, and report having lost work and struggled to build relationships due to their use of free online MT on smartphones for almost all interactions.

Translation is often necessary for survival in high risk, low shelf-life situations (D in Fig. 7.2) as addressed in work on crisis translation and on the use of free online MT engines for health and legal settings. Cadwell et al. (2019) provide examples of MT being used in response to crises, with common problems of underdeveloped technology for the language pairs in use, including insufficient data, and available data coming from an inappropriate domain. These problems exacerbate existing quality issues with MT, and Federici and O'Brien (2020) suggest preparedness and the intervention of professional translators and interpreters where possible to mitigate risk. In a crisis scenario, not all translation will be public-facing, and a digital divide may affect access to human or machine translation, whereby socio-economic factors (as highlighted by Cadwell et al. 2019) or gender inequality (as highlighted by Vollmer 2020) might limit access to technology generally. Therefore, the use of MT in these scenarios should be implemented with caution and ideally under supervision of translators or others with high MT literacy (Parra Escartín and Moniz 2019).

Example use cases for combined speech recognition and MT tools are often in medical settings (see Sumita 2017, for example), which appears to be a high-risk setting, despite the short MT shelf life. A mistranslation could have dire consequences for an individual. In high risk, long-life use cases (C in Fig. 7.2), such as translation of food ingredients, medicines and their accompanying information, or instructions for machinery, mistranslation could expose individuals to risk at the high end of Canfora and Ottmann's (2018) continuum, such as injury or death. While the argument for use of MT for assimilation in crisis scenarios might be a utilitarian effort to minimise harm, there can be little argument that the use of MT without expert human intervention in high-risk scenarios is neither wise nor ethical (Parra Escartín and Moniz 2019; O'Mathúna et al. 2020).

Aside from the use cases discussed in this section, we may have low risk, long shelf life (B in Fig. 7.2) use cases for MT, such as in the use of MT for literature or for user interface translation. The risks will come at the lower end of Canfora and Ottmann's (2018) continuum, with mistranslation risking communication being impaired or a loss of reputation. For literary translation, one could argue that there is a long-term risk to language, and reduced readability presents a possible risk to engagement with the other, to empathy. As noted by Bender et al. (2021), societal views or biases as represented in MT systems are set in aspic from the moment the

training data is harvested, whereas in society these will change over time. If literature models the way in which society thinks, feels and behaves, the consequences of poor engagement with the text could be of a high-risk nature that are currently unforeseeable. The argument inherent in the copyright waiver for developing countries in the 1971 update to the Berne Convention (see Moorkens and Lewis 2019) suggests that availability is currently considered to outweigh these risks.

In the following section, we look in detail at user interaction with raw and post-edited machine translation to bring in the direct voice of users and their perception of the issues faced.

7.3 The Voices of Users

This section gives voice to users of raw MT in technical and creative environments that are not language professionals, but the ultimate users or readers of translated texts.

Since 2017, part of our research has explored how using MT (both highly customised statistical, SMT, and NMT) engines impacts the user or reader experience when applying different translation modalities. We define translation modality as a descriptor of the process in which the translation is generated. For example, if the translation of a product or a story is generated by professional translators without the aid of MT, this is considered one translation modality that we could call "human translation", but if the translation is generated by MT and then post-edited it would be considered another modality, that we could call "MT post-editing (MTPE)", and finally, if raw MT output is used, this is labelled as MT.

7.3.1 First Experiment: Technical Environment

The first experiment involved 84 participants that were native Japanese, English, German and Spanish speakers using an eye-tracker (Guerberof-Arenas et al. 2021). The participants were frequent users of word-processing applications but had a different Microsoft Word (MS Word) literacy (varying experience when using MS Word).

We set up an intra-subject experiment where the users did six tasks, three of them using the published version of MS Word as localised for their native language (HT/MTPE as part of the content is post-edited), and, after a brief pause, the remaining three using a machine-translated version of MS Word (MT). Half of the participants in each language group follow the reverse order in the translation modality to counterbalance the order effect. The engine used for this "experimental translation" was a customised SMT engine used in production by Microsoft in the company's localization process at that time (Quirk et al. 2005), and therefore deemed of acceptable quality for post-editing for all the languages tested (Schmidtke and

Groves 2019). Based on previous experiments (Doherty and O'Brien 2014; Castilho 2016), and in order to analyse the usability of the different translation modalities, we looked at effectiveness, i.e. the number of completed tasks versus the total number of tasks; efficiency, i.e. effectiveness in relation to time; satisfaction, i.e. the level of satisfaction in completing tasks, the time, the instructions given and the language used in MS Word; and, finally, cognitive effort, i.e. the mental effort employed in completing (or not) the tasks. For a detailed methodology of this experiment, refer to Guerberof-Arenas et al. (2019, 2021).

The results show that the type of task, and hence the participants' experience and ability (MS Word literacy) was a factor in their effectiveness, but the translation modality was not a statistically significant factor. However, when it came to the combination of completed tasks and time, efficiency, and the reported users' satisfaction, the translation modality was an influencing factor, and MT scored significantly lower than the HT modality in efficiency and satisfaction. With regards to cognitive load, the results show that the English participants exert a lower cognitive effort when reading the instructions and completing the tasks in comparison to the rest of the languages, but there is no difference between the other languages or the modalities.

After the users had completed the experiment, each participant recorded a semi-structured interview while viewing their own eye-tracking data in a Retrospective Think Aloud (RTA) protocol. The interviews range from 10 to 20 min and they were conducted in English. Let us examine some of the relevant questions guiding the interviews and the responses by the participants to see the effect that MT had on their user experience.

7.3.1.1 Did You Notice a Difference in the Language in MS Word When You Came Back from the Pause?

Strikingly, only three participants said that they had noticed that the application had changed after the pause, and only one participant referred to MT "I just thought like this is a, this is something that was processed by a machine and you cannot rely on whatever you see, you have to search for it".

However, the majority did not notice the change. There were several reasons given for this: users with previous experience on that task did not look at the language in detail, they just focused on the action, because they knew the *location* of the option; others concentrated so deeply on reading the instructions and completing the tasks that they did not pay attention to the language in the application; others assumed it would be the same application, they looked at other cues in the application, or they were used to working with another version of MS Word. Having said this, the users did fixate on the words on the application, perhaps only to look for keywords or anchors without necessarily reflecting on the quality.

Most participants were not aware that they were using a different translation modality after the pause. This came as a surprise to the research group, as we were expecting that the change would be obvious because we were using a "fake" setup

with raw MT output—no post-editing at all was performed. This could have been especially problematic for the German and Japanese participants because traditionally these languages are difficult for MT, and because there were obvious errors in the text displayed within the application in all languages (Guerberof-Arenas et al. 2021).

7.3.1.2 How Did You Find the Quality of the German/Spanish/Japanese Here?

During the RTA, the participants were asked about the general quality of the language they were working on in each modality. The responses were rather mixed even from the same participants. However, from the 56 participants that commented on the quality of the MT modality, 23 mentioned, surprisingly, that the quality of the language was "Correct", "Fine", "Good" or "Very good", but at the same they were puzzled by some of the translations. "The Japanese language, I didn't recognise unnatural things in this task, but I did recognise some unnatural translation, like too informal language in some other tasks".

7.3.1.3 How Did You Find the Language in this Menu, Dialog Box, Option?

When the participants were asked about the language in certain menus, dialog boxes or options, 44 participants reported errors in the MT modality while only 5 reported errors in the HT modality (which were not actual linguistic errors). These are some examples of the errors found in MT.

> It says, 'links' so you can actually adjust the left side, but it didn't say anything for the right side, it just said '*Richting*'[2] and I was unsure about what it exactly means. (P04DE)

> I don't know if there is the original Spanish application or if there is a translation from the English. Because in the next task I was looking for the right for the column space in the right. But it was *Correcto*[3] which is the direct translation from Spanish, so I don't know if the Spanish version, I don't think it has written *Correcto* instead of right. And it was confusing for me but the rest I think it is fine. (P07ES)

The users resorted to some strategies such as using the context to understand the options or they back-translated the option to make sense of it.

Therefore, even if at the beginning we were puzzled that the change of modality was not obvious to the participants after the pause, we did realise that some of the MT options were confusing and that the language played an important role in finding

[2] *Richting* was the MT alternative instead of *Rechts* (right in German).

[3] *Correcto* was the MT alternative instead of *Derecha* because Right can have several translations, one meaning right-hand side and another one meaning correct.

an option, especially for users with little experience, and, hence, in completing the tasks.

7.3.1.4 How Did You Feel During this Task?

The participants expressed several feelings during the RTA such as confusion, confidence, concentration, disappointment, frustration, nervousness, unhappiness, and/or happiness. However, the two most frequent words (5 characters or longer) when describing their feelings were "confused" and "nervousness". On some occasions the participants were confused because of the tasks, and on other occasions they were confused because of the instructions and the difficulty to find the options, the functionality of the application, or indeed the language "I didn't see this. No *Izquierdo correcto*[4]".

In summary, in this first experiment, most users were not overtly aware when they were working with the MT modality. This could lead us to believe that using raw MT technology as part of the translation process, in the context of technical texts or even software applications as in this case, does not compromise the user experience since these texts would be considered (according to our models) low risk and long term (B in Fig. 7.2). However, the participants *did* experience difficulties with certain words which lead to inefficiency and less satisfaction when using MT despite being unaware that they were using MT, and this was more problematic with less-experienced users. Of course, there are other aspects to consider: the type of task, the experience, the MT quality for a specific language, and the MS Word setup that might also influence the user experience, but the results show that the translation modality was indeed a factor.

We see here that the user experience when dealing with the MT translation modality is not only difficult to gauge because of multiple confounding variables that make isolating language difficult, such as the task itself, the user's experience, the language combination, but also because the user has a pre-set notion of the quality of the application that is related to the status of that application and even the *historical* relationship of the user with this application. Users might be confused when they look for keywords or when they look at brief messages and this contributes to a poor experience *without* necessarily knowing why. There are insufficient studies that focus on testing the user experience when MT is involved as a translation modality, and we believe that asking the users if they found the information translated with MT useful will not reveal the real experience because this is a much more complex and nuanced phenomenon.

[4]The participant is referring here to the indentation where the MT proposal meant correct in Spanish instead of right. He understands left (*Izquierdo*) but then he sees correct (*Correcto*) instead of right (*Derecho/a*).

7.3.2 *Second Experiment: Creative Environment*

The second experiment explores the relationship between creativity in translation and different translation modalities, and the impact this could have on readers of translated fiction. The study involves 88 participants reading a short story in Catalan followed by a questionnaire in the Qualtrics online survey tool to assess three aspects of the user experience: narrative engagement, enjoyment and translation reception (Guerberof-Arenas and Toral 2020).

The experimental design this time involved three translation modalities: HT, MTPE and MT. To avoid the effect that one translator could have on the impact of the story (i.e. to avoid the translator style effect), two professional literary translators worked on the HT and on the MTPE modalities. The MT modality was provided by a NMT engine customised for the domain (literature) and language combination (English to Catalan) (Toral and Way 2018). The three modalities were analysed by a professional literary translator to identify creativity indexes (considering errors and creative shifts). Then, in a reception study, participants were presented randomly with a text translated in one of the modalities (Qualtrics automatically presents a balanced number per modality). After reading the text, they had to fill in a questionnaire that included a narrative engagement questionnaire (Busselle and Bilandzic 2009), enjoyment questions (Dixon et al. 1993; Hakemulder 2004) and a translation reception questionnaire designed for this experiment.

The results show not only that creativity is higher in HT, lower in MTPE and lowest in MT, but also that the reading experience is different depending on the modality. HT scores higher in narrative engagement and translation reception and is marginally lower than MTPE in enjoyment. However, there are no statistically significant differences between HT and MTPE for any of these variables. This means that once a professional literary translator intervenes in the process, the user experience for HT and MTPE appears to be comparable, even though we see a trend towards higher scores in HT. MT, unsurprisingly, has the lowest engagement, enjoyment and translation reception scores, and these results are statistically significantly lower than for HT and for MTPE. It is noteworthy, though, that those categories in the narrative engagement scale that are related to attentional focus, emotional engagement and narrative presence do not show statistically significant differences across the modalities.

These results seem to hint at the possibility that contemporary MT might be able to fulfil a communicative function for some genres of literary texts even if the reading experience is not as optimal as with the other modalities where professional literary translators intervene (for more information, see Guerberof-Arenas and Toral (2020))

Apart from the results from the questionnaires, what did the users have to say about the translation modalities?

7.3.2.1 If You Realised That It Was a Translation, Can You Describe How Did You Conclude This?

The participants were debriefed about the translation of the text and then were asked if they had realised they were reading a translation, 66% of the participants responded "Yes" to this question. The follow up to this was how they had realised it was a translation. In the follow up question, there was a striking difference between the modalities. In the HT modality, 75% of the participants realised that the text was a translation because it was set in the United States and so the names of the characters and places referred to this country, the remaining 25% refer to literal or unnatural expressions.

In the MTPE modality 60% of the participants refer to the USA setting, with the remaining 40% referring to unnatural words, phrases, spelling or word order (according to their own preferences).

In the MT modality, only 18% of the participants refer to the USA setting, with the remaining 82% referring to nonsensical words, literal translations, strange syntagmatic expressions, grammatical errors, wrong use of articles, lack of coherence, and incorrect word order.

These comments help us understand that this estrangement factor—these odd words and syntactic expressions—prevents the reader from engaging with or enjoying the text. If the reader is disrupted from these stylistic elements because they encounter errors, they become less focused on the text as a narration, and more focused on the unusual words and hence enjoy the experience less, even if the reader might be unaware of the translation modality.

7.3.2.2 Were There Any Paragraphs or Sentences That Were Difficult to Understand? Can You Tell Us Which Ones?

In the HT modality, 39% of the participants found that there were paragraphs or sentences that were difficult to understand because they contained too much information (for example, proper names) or the narrative voice changed from third to first person. Therefore, the difficulties were related to the structure of the ST and the narrative decisions made by the author. In the MTPE modality, 34% of the participants found paragraphs or sentences that were difficult to understand because of the change of narrative voice and because the syntax was confusing. They gave examples of sentences that were long and difficult to follow. In the MT modality, 64% of the participants responded "Yes" to this question because they found odd or nonsensical words, confusion in the gender of articles, wrong syntactical constructions, incorrect proper noun gender (*la Joan* instead of *el Joan* because in Catalan this is a name for men), or they simply thought that certain sentences were not properly translated.

In this second question we can see that in the HT and MTPE modalities the main issues are related to the original ST while in the case of MT the issues are directly related to the translation.

7.3.2.3 Was There a Sentence or a Paragraph That You Especially Liked? Can You Tell Us Which One?

In the HT modality, most participants, 61%, found paragraphs or sentences that they liked. They referred especially to the paragraphs of the story that described the way one of the main characters dies and the description of his gaze.

In the MTPE modality, 43% of the participants found paragraphs or sentences that they liked. They also referred to the paragraphs in the story where the author describes how clinical death comes about and the description of the dying man's gaze. However, most of the participants (57%) did not say that they liked a paragraph or sentences in this modality, in sharp contrast with the experience of those in the HT modality.

In the MT modality, 36% of the participants found paragraphs or sentences that they liked, mainly the first paragraph where the author explains what clinical death is, even though there were translation errors in this paragraph.

In summary, when the users were asked specifically about parts of the texts they liked, a majority liked parts of the text in the HT modality, and the three modalities make references to parts of the text that had already powerful imaginary in the ST.

7.3.2.4 Do You Want to Make Any Other Comment?

There were other comments from the participants at the end of the survey. One participant that read the HT modality said:

> I have done the exercise in half an hour, I have read the text very quickly. The writing has had an impact on me, I could see the images. For this reason, I wouldn't read something similar. The text is well written, and it transmits emotions. But I never read this type of horrible thing. It is not my genre. But yes, the translation is brilliant, it transmits everything. (P37)[5]

For this modality, the main issues reported are the genre, the topic and the narrative style of the ST. Participant P37 even refers to the translation as "brilliant".

MTPE

For this modality, again the main issues reported are aspects of the ST. Participant P68 even wants to read the whole book and P85 also praises the translation.

[5] The translations from Catalan into English are provided by the authors.

> I would like to read the whole book. (P68)
>
> I haven't read the original to correctly assess the translation. The Catalan used is very good, but I doubt that the English version would make descriptions using less colloquial expressions. (P85)

MT

In this modality, users did notice errors even though they did not disclose if they realised that they were reading MT, and this seemed to influence their reading experience.

> There are words in the text that do not make any sense in the context, such as "jihad" or "thone". (P03)
>
> It is difficult to know the quality of the translation without reading the original text. (P06)
>
> As I advanced in my reading of the text the translation deficiencies have become less problematic. (P23)
>
> I've been in lock down for about fifty days and maybe this has influenced the fact that I had a hard time concentrating. (P49)

We see here that once the readers are immersed in the narrative, they might *compensate* for the lack of coherence, lexical accuracy and cohesion, so the context and the narrative help to decipher a low-quality text. We wonder about the additional cognitive effort of doing this, especially for a longer story. Recent research confirms that the cognitive effort is higher when reading MT in literary texts (Colman et al. 2021). Finally, something that appears obvious, but given the current world situation seems even more relevant, the personal circumstances of readers influence their perception of language and their engagement. If participants already have difficulties reading because of their personal circumstances (such as a pandemic), shouldn't a translation facilitate the engagement and enjoyment of the text instead of making it more cognitively demanding?

In summary, we see that in a creative environment, the MT modality has a strong effect on the reader experience. Readers show significantly less engagement, enjoyment and diminished reception than those who read a version where a professional translator intervened. We also see a pattern of higher values in the HT modality in relation to the MTPE. We are aware that at present MT is not used in the publishing sector, or at least its use is not publicised. However, MT is becoming an intrinsic part of the translation workflow in the audiovisual sector through platforms that might simplify structures to obtain better MT output (Mehta et al. 2020). Are viewers then exposed to the best possible version of their language? There are some studies that look into the productivity gains and quality of subtitles and conclude that this is a viable solution (Bywood et al. 2017; Matusov et al. 2019; Koponen et al. 2020), but we feel that analysing productivity and final quality in a more "traditional" way leaves out an important aspect: the impact that MT has on the viewers. Based on this research, it is important that streaming platforms make viewers and readers aware of

the use of MT, but also that translation reception studies are prioritised if technology is to be used in any creative domain. The implications of these results for society are various: on the one hand it shows that MT diminishes the reader's experience and that using MT as a tool constrains the translators' creativity. On the other hand, the long-term effects of using technology might be worrying such as loss of lexical richness, style simplification, loss of reputation for authors, and, thus, the minimization of the transformative effect that literature and fiction has on society.

7.4 Discussion on Ethical Implications

In Sect. 7.1 we saw that risk for low-stakes MT is minimised (but not negligible), rising along with the shelf life of the text. In Sect. 7.2 we looked in detail at two experimental use cases for raw MT—so-called MT for assimilation—where we can see that, although the risk is not high, there are nuanced implications for translation modalities that affect the end user or reader experience. Participants in the studies described in Sect. 7.2 were not aware of the translation modality chosen to produce the text that they engaged with, as their preconceptions may have altered their behaviour or responses.[6]

It is commonly assumed that MT should help users to communicate and that this means of communication is improving as the technology improves, e.g. the perception that NMT is "harmless" in short-term and low-risk scenarios, as opposed to high-risk long-term scenarios that involve mainly health, crisis and/or legal settings (Vieira et al. 2020). This stems from the logical perspective that when users are only trying to gist for content, skimming a website, a document, or a message, misleading or even inaccurate translations are not as "important" as when users are trying to understand or carry out an action that involves their health, their legal status, or indeed their survival. We are aware, as users of public and private MT technology and researchers, that indeed MT helps users to communicate in a language other than their own, especially if they have not mastered that second language.

By looking at the data from our research, we see several common patterns when examining the user experience in the context of raw MT output in several models considered low-risk: (a) users do not necessarily recognise that they are exposed to MT, as indeed the technology improves, if they are not explicitly informed; (b) nevertheless, users might be confused, frustrated or (in the case of biased output) misled by the information found and are likely to encounter errors, awkward style, and unintelligible words that will result in lower efficiency, satisfaction, enjoyment or engagement scores; and (c) end users are affected by the translation modality, especially if their experience of the translated application or knowledge of the source

[6] Our ethics committees agreed that these were low-risk settings for use of MT, but in a high-risk setting this should change. At what level of risk does a study using MT without informing participants become research involving deception?

Table 7.1 Summary of risks discussed

Information duration	High risk	Low risk
Short term	• Overreliance on MT may cause miscommunication • MT might facilitate timely and cheap sharing of information • Health hazards • Legal implications • Loss of work opportunities	• Overreliance on MT may cause misunderstanding • Facilitates timely and cheap communication • Might cause confusion and frustration • Loss of efficiency • Loss of engagement
Long term	• Financial consequences • Legal actions • Loss of health • Loss of legal status • Loss of privacy • Loss of working opportunities	• Higher cognitive effort • Loss of satisfaction • Impoverishing of language • Loss of reputation (brands or authors)

language is low, and they either fail to achieve what they set out to do or compensate for errors by looking at the context, back-translating, or even compensating at later stages of their interaction with the "product". We conclude that the user experience is not a binary issue resolved by asking if this "information was helpful or unhelpful in your language" or by counting the "number of translation errors in the target language" or by calculating a similarity score with a gold reference translation (as happens in automatic MT quality evaluation). User experience research that considers MT should look at a broader picture that considers experience not only as a static and isolated event, but as part of a communicative process in the short and long term. See Table 7.1 for a summary according to levels of risk and length of time.

All of this indicates that there are implications and risks inherent in the use of MT that has not been reviewed, suggesting that readers or users should be made aware when text has been machine translated via a note or, keeping in mind the legal implications of mistranslation, a disclaimer. Usability, privacy, and (cyber-)security do not always integrate well, as evidenced by the 'accept cookies' popups that are mandated by GDPR to appear when visiting many websites from within the European Union, but a mandatory label or disclaimer for raw MT should not be distracting or user-unfriendly. We should also clearly indicate what standard of pre-publication review is necessary to avoid such a disclaimer.

The motivation for low-stakes MT, as described in Sect. 7.1, is for fast or immediate translation of highly perishable low-risk text. As the shelf-life and risk level of texts increase, producers or content providers may want to use MT in translation workflows or even use raw MT to increase productivity and reduce costs. There is a decision to make in this case, balancing potential risk against speed and savings. Sometimes, as in Massidda (2015) and many multimedia translation production networks, consumer demand for simultaneous releases in all locales—what used to be called simship—pushes turnaround speed. But is it really

the case that consumers cannot wait for the new game, software, or product, even if this means a compromise on quality and an introduction of risk? Clearly labelling MT published without review would make this clearer.

7.5 Conclusions on Ethical Implications

Therefore, in the MT ecosystem, the stakeholders involved have certain ethical responsibilities for the communication to be optimal and fair, if indeed this is the objective of MT. A well-informed translation ecosystem could better self-regulate, although it is naïve to assume that this would end unethical practices.

Firstly, the stakeholders in charge of improving or creating MT engines have a responsibility to create algorithms and treat data in a way that not only is not discriminatory or harmful, but also so that the data is harvested in a way where the "owners" of this data are acknowledged and reap the benefits of their work, and that the final product preserves the richness of the user experience. This might mean curating and documenting the data at deeper levels, such as lexical, syntactical and stylistic levels, as well as removing elements of gender, racial or national bias, but also in acknowledging the ownership of that data, and highlighting where issues might arise in that data (racial, gender and other biases for sure, but also impoverishment of the language). This can be done by engaging professionals in the translation field and not simply "annotators" found in online mechanical torques and by understanding translation processes and user experience as opposed to taking data only as an engineering and numerical exercise. As Bender et al. (2021) suggest, this requires greater transparency and accountability on the part of MT systems and services.

Secondly, companies or government agencies that make use of raw MT without human intervention for reasons such as "improving productivity" or "disseminating information" need to make this fact visible to users so that they know the content at hand has not been reviewed and understand the consequences mistranslation might have. Profit might be the primary motivation for a business and information dissemination for government agencies, but the ethical responsibility towards its users and society should be equally considered. In the case of MT, business and governments could consider the immediate effect of its usage (can the user understand/enjoy the product/text) and also the long-term effect (such as the impact of MT on language style/richness in the target culture). This means that the responsibility not only lies with the user in deciding to use public MT engines or translators using MT integrated in CAT tools, but also with businesses and governments to investigate this impact and to make it public so that users/society can choose whether to utilise a text with enough accurate information at hand. There needs to be a legal framework (Yanisky-

Ravid and Martens 2019)[7] that enforces the advertising of MT usage as part of the translation process so that users are aware and proceed with caution, and this warning cannot be in small print hidden somewhere within the legal documentation of the product or the text, but visible in the application, the text, or the leaflet that spells out the existing issues with the technology and the possible risks the users might encounter while using it.

Thirdly, translators and post-editors are also stakeholders in this ecosystem, as they are obviously involved in the MT workflow as creators of training data when post-editing, translating, and even when using MT as an additional tool. They have a responsibility to acquire enough knowledge about the technology, to be aware of the types of errors and biases found in the raw output and how best to fix them. Also, more importantly, they need to be aware of copyright infringements and possible interferences in producing a high-quality final product. We wonder how long a translator can be exposed to a simplified version of a language without being influenced by it. We understand that translators are not always well compensated for their work, but there is a belief in some parts of the translation community that, because they are dealing with MT post-editing, the effort required and the responsibility towards the content, the user and reader, and the final quality is not as high as without MT, and that the knowledge of a language is something immutable that will not be influenced by processing poor translations and reading low-quality references. The suggestion from Parasuraman et al. (2000) is that reasonably (but not entirely) reliable automation can lead to skill degradation and to complacency in trusting the automated output. In a similar way that good journalism does not occur by cutting and pasting news from different sources, good translation practice does not involve just cutting and pasting references from different technologies and that process could result in a serious impoverishment of the language and the profession.

Fourthly, academics need to bring to light the aspects that we have mentioned in this paper, mainly use and propagation of curated data, responsible use of that data, MT literacy for translators and users, analysis of user and reader reception of translated work, including different translation modalities in different languages, and different genres. Academia has been better at analysing the use and effect of MT than that of other tools (such as translation memories), possibly because the impact of MT has also been greater. However, there is a need to engage more often and more deeply with the final users of MT and within different branches of knowledge. Interdisciplinary research cannot be a mere box to be ticked in grant applications, it must be a real endeavour where academics value all aspects of a field. The more mathematical or engineering sides of research cannot have precedence over the more humanistic side of technology: the effect on the people using the technology. For research on user reception and user experience, interdisciplinary research needs to

[7] The authors offered a detailed description of copyright laws, translation and how AI might be infringing these international laws that protect authors and creators, and they suggest including work generated by AI systems.

happen at all levels of the research system, and not as an afterthought regarded as a lesser science.

And finally, we should also consider the ethical responsibility of the users/readers. If the ethical dimension is an ecosystem, users also have the responsibility to buy products that protect language, translators, and those in the text supply chain in the source and target culture. In the same way that some consumers might not buy certain brands or in certain shops or from web portals or shopping malls because of their operational practices or because in doing so they are destroying the local, social, and commercial fabric of their city, region or country, users should have enough information about how a text is produced to choose not to buy/see/hear a given product or, at least, to know the effects this will have on their user experience, their language, and their culture in the long term. This can only happen if readers and users are informed by the whole ecosystem and if the ecosystem promotes transparent information sharing.

References

Bender EM, Gebru T, McMillan-Major A, Shmitchel S (2021) On the dangers of stochastic parrots: can language models be too big? In: FAccT '21, March 3–10, 2021, Virtual Event, Canada

Bowker L, Ciro JB (2019) Machine translation and global research: towards improved machine translation literacy in the scholarly community, 1st edn. Emerald Publishing, Bingley

Busselle R, Bilandzic H (2009) Measuring narrative engagement. Media Psychol 12:321–347. https://doi.org/10.1080/15213260903287259

Bywood L, Georgakopoulou P, Etchegoyhen T (2017) Embracing the threat: machine translation as a solution for subtitling. Perspectives 25:492–508. https://doi.org/10.1080/0907676X.2017.1291695

Cadwell P, O'Brien S, DeLuca E (2019) More than tweets: a critical reflection on developing and testing crisis machine translation technology. Transl Spaces 8:300–333

Canfora C, Ottmann A (2018) Of ostriches, pyramids, and Swiss cheese: risks in safety-critical translations. Transl Spaces 7:167–201

Canfora C, Ottmann A (2020) Risks in neural machine translation. Transl Spaces 9:58–77

Castilho S (2016) Acceptability of machine translated enterprise content. Ph.D. Thesis, Dublin City University

Ciribuco A (2020) Translating the village: translation as part of the everyday lives of asylum seekers in Italy. Transl Spaces 9:179–201

Colman T, Fonteyne M, Daems J, Macken L (2021) It is all in the eyes: an eye-tracking experiment to assess the readability of machine translated literature. In: The 31st meeting of computational linguistics in The Netherlands, Ghent

Čulo O, Nitzke J (2016) Patterns of terminological variation in post-editing and of cognate use in machine translation in contrast to human translation. In: Proceedings of the 19th annual conference of the European association for machine translation, pp 106–114

Desjardins R (2017) Translation and social media. In: Theory, in training and in professional practice, 1st edn. Palgrave Macmillan, London

Dixon P, Bortolussi M, Twilley LC, Leung A (1993) Literary processing and interpretation: towards empirical foundations. Poetics 22:5–33. https://doi.org/10.1016/0304-422X(93)90018-C

Doc Society v. Pompeo (2019) Doc Society v. Pompeo: a lawsuit challenging the State Department's social media registration requirement. In: 1st Night Amendment Institute at Columbia University. https://knightcolumbia.org/cases/doc-society-v-pompeo

Doherty S, O'Brien S (2014) Assessing the usability of raw machine translated output: a user-centered study using eye tracking. Int J Hum Comput Interact 30:40–51. https://doi.org/10.1080/10447318.2013.802199

Federici FM, O'Brien S (2020) Cascading crises: translation as risk reduction. In: Federici FM, O'Brien S (eds) Translation in cascading crises. Routledge, Abingdon, pp 1–22

Guerberof-Arenas A, Toral A (2020) The impact of post-editing and machine translation on creativity and reading experience. Transl Spaces 9:255–282

Guerberof-Arenas A, Moorkens J, O'Brien S (2019) What is the impact of raw MT on Japanese users of Word: preliminary results of a usability study using eye-tracking. In: Proceedings of XVII machine translation summit. European Association for Machine Translation (EAMT), Dublin, pp 67–77

Guerberof-Arenas A, Moorkens J, O'Brien S (2021) The impact of translation modality on user experience: an eye-tracking study of the Microsoft Word user interface. Mach Transl. https://doi.org/10.1007/s10590-021-09267-z

Hakemulder J (2004) Foregrounding and its effect on readers' perception. Discourse Process 38:193–218. https://doi.org/10.1207/s15326950dp3802_3

Hern A (2020) Amazon hits trouble with Sweden launch over lewd translation. The Guardian

Kenny D, Moorkens J, de Carmo F (2020) Fair MT: towards ethical, sustainable machine translation. Transl Spaces 9:1–11

Koponen M, Sulubacak U, Vitikainen K, Tiedemann J (2020) MT for subtitling: user evaluation of post-editing productivity. In: Proceedings of the 22nd annual conference of the European association for machine translation. European Association for Machine Translation, Lisboa, pp 115–124

Kranzberg M (1986) Technology and history: "Kranzberg's Laws". Technol Cult 27:544–560

Larsonneur C (2021) Neural machine translation: from commodity to commons? In: Desjardins R, Larsonneur C, Lacour P (eds) When translation goes digital: case studies and critical reflections. Springer, Cham, pp 257–280

Liebling DJ, Lahav M, Evans A et al (2020) Unmet needs and opportunities for mobile translation AI. In: Proceedings of the 2020 CHI conference on human factors in computing systems. ACM, Honolulu, pp 1–13

Marking (2020) Thai mistranslation shows risk of auto-translating social media content. Slator

Massidda S (2015) Audiovisual translation in the digital age: The Italian fansubbing phenomenon, 1st edn. Palgrave Macmillan, London

Matusov E, Wilken P, Georgakopoulou Y (2019) Customizing neural machine translation for subtitling. In: Proceedings of the fourth conference on machine translation, vol 1. Association for Computational Linguistics, Florence, pp 82–93

Mehta S, Azarnoush B, Chen B, et al (2020) Simplify-then-translate: automatic preprocessing for black-box machine translation. arXiv:200511197 [cs]

Moorkens J, Lewis D (2019) Research questions and a proposal for governance of translation data, p 24

Nurminen M (2018) Machine translation in everyday life: What makes FAUT MT workable? In: TAUS eLearning blogs. https://blog.taus.net/elearning/machine-translation-in-everyday-life-what-makes-faut-mt-workable. Accessed 25 Aug 2020

Nurminen M, Koponen M (2020) Machine translation and fair access to information. Transl Spaces 9:150–169

O'Mathúna DP, Escartín CP, Roche P, Marlowe J (2020) Engaging citizen translators in disasters: virtue ethics in response to ethical challenges. TIS 15:57–79. https://doi.org/10.1075/tis.20003.oma

Olohan M (2017) Intercultural faultlines: research models in translation studies: v. 1: textual and cognitive aspects. Routledge, London

Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern 30:286–297. https://doi.org/10.1109/3468.844354

Parra Escartín C, Moniz H (2019) Ethical considerations on the use of machine translation and crowdsourcing in cascading crises. In: Translation in cascading crises, 1st edn. Routledge, London

Paullada A (2020) How does machine translation shift power? In: Resistance AI workshop at NeurIPs 2020, Virtual Event, Canada

Pichai S (2016) Google I/O 2016 - keynote

Quirk C, Menezes A, Cherry C (2005) Dependency treelet translation: syntactically informed phrasal SMT. In: Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05). Association for Computational Linguistics, Ann Arbor, pp 271–279

Raley R (2003) Machine translation and global English. Yale J Crit 16:291–313

Rey B (2014) Your tweet half-life is 1 billion times shorter than Carbon-14's. In: Wiselytics. https://www.wiselytics.com/blog/tweet-isbillion-time-shorter-than-carbon14/. Accessed 3 May 2021

Schmidtke D, Groves D (2019) Automatic translation for software with safe velocity. In: Proceedings of machine translation summit XVII volume 2: translator, project and user tracks. European Association for Machine Translation, Dublin, pp 159–166

Smith R (2018) The google translate world cup. The New York Times

Sumita E (2017) Social innovation based on speech-to-speech translation technology targeting the 2020 Tokyo Olympic/Paralympic Games Presentation at MT Summit XVI, Nagoya, Japan

Thicke L (2013) Post-editor shortage and MT. Multilingual Magaz 2013:42–44

Toral A (2019) Post-editese: an Exacerbated Translationese. arXiv:190700900 [cs]

Toral A, Way A (2018) What level of quality can neural machine translation attain on literary text? arXiv:180104962 [cs]

Vanmassenhove E, Shterionov DS, Way A (2019) Lost in translation: loss and decay of linguistic richness in machine translation. In: Proceedings of machine translation summit XVII volume 1: research track. European Association for Machine Translation, Dublin, pp 222–232

Vieira LN, O'Hagan M, O'Sullivan C (2020) Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. Inf Commun Soc 1:1–18. https://doi.org/10.1080/1369118X.2020.1776370

Vollmer SM (2020) The digital literacy practices of newly arrived Syrian refugees: a spatio-visual linguistic ethnography. PhD Thesis, University of Leeds

Wang J, Xu C, Guzman F, et al (2021) Putting words into the system's mouth: a targeted attack on neural machine translation using monolingual data poisoning. arXiv:210705243 [cs]

Way A (2018) Quality expectations of machine translation. In: Moorkens J, Castilho S, Gaspari F, Doherty S (eds) Translation quality assessment: from principles to practice. Springer, Berlin, pp 159–178

Weaver W (1949) Translation. UNESCO memo. Rockefeller Foundation

Winner L (1983) Technologies as forms of life. In: Cohen RS, Wartofsky MW (eds) Epistemology, methodology and the social sciences. Reidel, Dordrecht, pp 249–263

Yanisky-Ravid S, Martens C (2019) From the myth of babel to google translate: confronting malicious use of artificial intelligence – copyright and algorithmic biases in online translation systems. SSRN J. https://doi.org/10.2139/ssrn.3345716