Hungarian translators' perceptions of neural machine translation in the European Commission

Ágnes Lesznyák
European Commission
Directorate-General for Translation
`agnes.lesznyak@ec.europa.eu`

Abstract

This paper summarises findings from structured interviews with Hungarian translators in the European Commission's Directorate-General for Translation on their experiences with neural machine translation as a translation aid since 2017. The translators have widely divergent views on the use and usefulness of neural machine translation and varying practices when it comes to integrating it into their work. The paper concludes that human factors play a crucial role in the success of application and argues that translators' attitudes and intriguing cognitive processes merit greater scientific attention.

1 Introduction

With the emergence of neural technology, the quality of machine translation (MT) output has improved rapidly in recent years, in particular for languages with more complex morphology, such as Hungarian. Nevertheless, human intervention is still indispensable for checking and improving texts where the accuracy of information transfer is vital, such as in the legal domain, or texts intended for publication (Ive et al. 2018; Way 2018; Knowles et al. 2019).

Research in this area has focused mainly on aspects of post-editing (PE), the traditional treatment applied by translators to improve the quality of MT output.[1] A number of studies (e.g. Plitt and Masselot 2010; Koponen 2012; Guerberof 2014; Koehn and Germann 2014) have found that PE productivity and effort differ greatly between individual translators dealing with MT output. However, human factors such as professional translators' views and practices, have attracted relatively little academic attention.

In a groundbreaking study, Cadwell et al. (2017) investigate what influences the adoption of MT by professional translators in two different institutional settings, one being the European Commission's Directorate-General for Translation (DGT). They conclude that translators' sense and level of 'agency' have a crucial impact on their attitudes towards MT.

The changing nature of MT output (due to rapid technological development) may also influence translators' perceptions and work processes. Since neural machine translation (NMT) is a recent development, the body of relevant research is necessarily small (Castilho et al. 2019).

By reporting on the experience of professional translators working in DGT's Hungarian Language Department, we want to raise awareness of aspects of interaction with NMT that are highly relevant for practitioners and may require further scientific insight, in order to prompt greater acceptance and more efficient use of this tool.

2 Background

2.1 eTranslation

eTranslation[2], the successor to the European Commission's MT service MT@EC, has been developed by the DGT in the framework of the Connecting Europe Facility. It offers statistical (SMT) and neural machine translation (NMT) into all the 24 official languages of the European Union, plus Icelandic and Norwegian. At the time of writing, it can be used by officials in the

---

[1] For an overview of the practice of and research into PE, see Koponen (2016).

[2]

https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/What+is+eTranslation

EU institutions and all EU Member States, Iceland and Norway. It is intended mainly as a component of digital services, but it also offers stand-alone services for the translation of documents and text snippets[3] and is provided as an integrated service for use by DGT translators (see section 2.2). eTranslation guarantees data confidentiality and security.

Following a test phase in early 2017, the general roll-out of NMT in DGT began with the launch of eTranslation in November 2017.

The emergence of neural technology represented a breakthrough for MT in language pairs involving Hungarian (Tihanyi and Oravecz 2017). As the level of quality of SMT did not allow for efficient PE, DGT's Hungarian translators had only used it sporadically. Therefore, working with NMT output was most Hungarian translators' first encounter with MT in their professional activity.

2.2   NMT in DGT's internal workflow

DGT offers translation services to other Commission Directorates-General, who send translation requests to a central DGT service that pre-processes texts automatically using various applications. Relevant segments and documents are extracted from predefined databases and a normative memory, and subsequently made available to translators as tmx files. Since July 2018, this pre-processing has involved MT into all EU languages (except Irish). The resulting tmx file can be imported, together with other tmx files, into the CAT tool.

DGT guidelines for the use of MT set out minimum requirements in terms of translators' technical knowledge and the amount of MT output they are expected to post-edit in order to familiarise themselves with the technology.

Beyond this, individual translators decide whether to use MT output for any given assignment. In principle, they have three options:
− pre-translate the whole document with the pre-processed tmx files provided (including MT) and then post-edit the text; or
− insert the MT.tmx automatically (auto-populate) and post-edit it segment by segment; or
− use the predictive typing function to insert chunks offered from the MT.tmx.

3   Interviews

DGT's Hungarian Language Department employs 49 translators and 9 assistants[4]. The translators are supported by a terminologist and a language technology coordinator. The language department also has a quality officer responsible for quality management[5].

DGT translators have a variety of professional backgrounds. Many Hungarian translators have studied languages and been formally trained as translators. However, some have a degree in law, economics or engineering, plus a post-graduate diploma or several years of professional experience in translation.

Between June 2018 and January 2019, the quality officer carried out structured interviews for internal quality management purposes with 38 translators working in the department.[6] By that time translators had been able to use NMT in their daily work for 8 to 14 months, mainly in translating from English into Hungarian. The interviews were not intended as a survey that would produce quantifiable results. Rather, the aim was to explore translators' views on NMT, their work practices (cf. above options) and practical issues that had arisen. The findings would feed into follow-up quality assurance action.

Translators were interviewed in Hungarian and were asked the following questions:
− When is it worth using NMT?
− How do you use NMT?
− Which language version do you read first?
− In what way is NMT different from translation memories (TMs) containing human translation? What deserves special attention?
− What are the advantages of using NMT?
− As a reviser, are you aware of whether a translation was made using NMT?

Translators were also encouraged to raise and discuss any topics that they considered relevant to the use of NMT. Therefore, the interviews differed considerably in length, from around 20 to 90 minutes.

The interviewer made notes of the translators' answers[7] and analysed and summarised them in a

---

[4] The figures reflect the situation in April 2019. The number of translators and assistants in active service are subject to constant change.
[5] On quality management and the role of quality officers see Drugan et al. (2018).
[6] Time and workload constraints meant that it was not possible to talk to all the translators. No assistants were interviewed, as they do not work with NMT.
[7] For reasons of collegiality, it was not considered

report. The results were shared and discussed with the translators in a half-day workshop. This made it possible to double check the interpretation of the data collected.

For obvious methodological reasons this paper makes no claims of generalizable findings. Where ratios are mentioned, this is to highlight recurring themes in translators' reports. Issues referred to only once may turn out to be just as relevant for the translation process. The themes identified here should be verified in future research.

4 Discussion

The interviews showed that the translators' views on NMT are highly divergent and their work practices vary widely a year after the introduction of NMT into the workflow. They expressed conflicting opinions on how useful NMT was, when it was worth using and how it could be used efficiently. Their observations were so disparate that the interviewer sometimes wondered whether they referred to the same tool.

The differences were reflected not only in the details, but also in the translators' overall opinions as to the benefits of NMT. Some see NMT as a very useful, positive development, while others (having tested it on a number of documents) have stopped using it or use it only sparingly. These findings are in line with those of Cadwell et al. (2016), who found no consensus among DGT translators on some central questions relating to the use of MT.

In the following sections, we summarise and discuss recurring themes from the interviews.

4.1 Factors influencing the usefulness of NMT output

Two factors in particular seem to determine whether Hungarian translators in DGT consider NMT useful: relevance, and the quality of the NMT output, which is perceived as correlating with segment length. The latter will be discussed in section 4.2.

The question of relevance of MT output is critical for DGT translators: very often, EU documents relate closely to previous documents, such as legal bases, or concurrently translated other texts. In such cases, consistency between texts is paramount: translators must re-use previous translations and not translate the new document from scratch. As a result, they may prefer TMs

originating in DGT's multilingual database (Euramis) and judge NMT output counterproductive or a distraction in certain situations.

When gauging the usefulness of an NMT.tmx file, translators rely on 'match rates': some import the file only if the match rate for the document is under a certain percentage, typically 50% or 30%. Cadwell et al. (2016) report that DGT translators working into different languages seem to differ in this respect: while those translating to some languages use NMT output when the retrieval rate is low, others do so when it is high. The reason for this apparent disagreement may lie in the types of document translated by individual translators and an inherent contradiction in MT. Generally, NMT output seems to be useful when there is no TM available and translators would have to translate from scratch. However, NMT works better for high-retrieval, i.e. recurring documents since they were included in the training corpus of the engine with high probability. Recurrence means that there are TMs available which may or may not have priority over MT, depending on the document type. NMT output may be useful for high retrieval documents if there is no obligation to edit fuzzy matches from TMs. Therefore, this apparent disparity may require finer analysis.

Some translators import the NMT.tmx file and take a decision on the usefulness of the NMT output at segment level. They typically use NMT for 'empty' segments, i.e. where reference TMs do not return any hits under a certain match rate. This approach allows them to respect the 'relevance' principle and use NMT at the same time.

Some translators highlighted the usefulness of NMT for urgent assignments, despite the risk of the MT output not being sufficiently post-edited. They argued that a greater risk in such situations is not to have a translation at all, i.e. not to comply with the service provision requirement.

The domain and the genre of a document does not seem to directly influence the perceived usefulness of NMT in the DGT working environment. Interestingly, the interviewees held opposing views as to the usefulness of NMT for particular document types such as press material, Commission communications and legal acts, and used the same arguments for and against NMT. There are many possible explanations for this. One may be translators' varying sensitivity to or awareness of different types of error in NMT output (see section 4.2). Another may lie in their working methods. Very few reported unprompted that they do a complete read-through of their

appropriate to make audio-recordings.

translations. We have reports in another context that this is not done in times of high workload. There is some evidence (e.g. Dragsted 2006) that CAT tools direct translators' attention to the segment level. Läubli et al. (2016) report that the document-level (as opposed to segment-level) presentation of NMT output influences human raters' perception of quality.

One interesting factor mentioned by some translators as having an impact on the usefulness of NMT was their familiarity with the source language or the domain in question. They rely on NMT more when working from a language in which they do not feel confident or in a domain with which they are not familiar. This is in line with the findings of Moorkens et al. (2018). In their study, translators with less experience found MT suggestions more useful. Although the Hungarian translators reporting this benefit of MT had sufficient translation experience, they felt a certain lack of language or domain competence in the situation in question.

The next section will discuss translators' perceptions of NMT output and of the treated (i.e. pre-revision) product.

4.2 Typical errors and quality issues

The interviewees said that the unpredictable quality of NMT output is a key factor discouraging them from using it. Quality varies widely from one segment to another: some need hardly any intervention, while others have to be re-translated. Therefore, NMT output always has to be checked thoroughly and very often requires significant PE. This may explain why translators develop certain 'control' practices (see section 4.3).

As mentioned above, segment length seems to be a decisive determinant of the quality of Hungarian NMT output. The interviewees mostly agree that short segments are of much better quality than longer ones – the former only need to be revised, whereas the latter often have to be deleted and translated from scratch. The borderline between 'short' and 'long' seems to be around 30–40 words.

A correlation between sentence length and quality has previously been reported for other language pairs (for an overview, see Castilho et al. 2018). Koehn and Knowles (2017) found that NMT outperformed SMT up to a sentence length of about 60 words, but beyond that the quality fell off. The fact that 200-word sentences are not uncommon in certain types of EU document may place a serious constraint on the usefulness of NMT in the Commission.

Translators see longer sentences as problematic not only because of the potential for lower NMT quality but because their complexity prevents a quick assessment of their correctness. It takes too much time and cognitive effort to analyse the components and decide what can be used. In such cases, it seems more efficient to re-translate the segment.

Translators' responses as regards recurring errors identified in NMT output confirm the relevant findings in the literature (see e.g. Eisold 2017; Van Brussel et al. 2018; Yamada 2019). Below, we discuss these errors in the case of Hungarian.

Mistranslation and deceptive fluency emerged as the two main issues. The sources of the errors were not easy to identify, but problems mentioned included incorrect word order; misplaced attributes; inversion of subject and object; and wrong ordering of clauses. Elliptical sentences and non-literal meaning also seem frequently to give rise to mistranslations.

While these error types call for close attention to the text, intensive PE is needed to correct other typical errors, such as incorrect information structure and missing referential elements and sentence connectors. Translators have to convert sentence-level MT into a coherent text.

On the other hand, morphological errors seem to be rare in eTranslation's Hungarian NMT output. Errors mentioned were incorrect endings, definite articles and possessive structures, and non-concordance between subject and object.

A typical recurring error in the Hungarian NMT output is the translation of proper names into fictional words. Since this is a new phenomenon for translators and a challenge for automatic quality checks, this type of error constitutes a risk in the translation process. It may have serious consequences if it prevents the reader from identifying a unique referent.

A source of serious concern for many translators was incorrect or inconsistent terminology, and in general, the context-independent translation of vocabulary. This tendency in the NMT output is a serious obstacle to efficient work, since terms have to be checked one by one in a termbase. As DGT translates many legal texts, this type of error presents a high risk and increases the need for thorough quality control.

When asked about the quality of translations submitted for revision, the majority of the trans-

lators[8] maintained that they clearly recognised NMT chunks and segments even if 'automated translation' was not explicitly signaled in the CAT tool. Only three said they could not tell whether NMT had been used. When asked about 'tell-tale signs', some cited cases of colleagues who had worked reliably for years suddenly starting to translate less accurately and produce highly amateur solutions. However linguistic perfection may also be a give-away if readers realise that a text does not follow the institutional style they are used to. There were also complaints about the readability of translations, unclear references and missing logical links, i.e. document-level errors.

These remarks may indicate that revisers encounter more, or other types of, errors than before the introduction of NMT. Several said that pre-revision translations have been of a lower quality, requiring more intervention and effort from revisers. This may lead to tensions between colleagues concerned.

Generally, the errors that revisers identified as revealing the use of NMT were the same as those detected in the NMT output itself. This may suggest that at least some errors filter through to the quality check phase, indicating flaws at the translation stage.

Therefore, in the next section we discuss how translators interact with NMT output.

4.3 Technical aspects of NMT use

DGT's Hungarian translators have developed divergent practices to integrate NMT into their individual workflow in what they consider an efficient and safe way. They have shown astounding creativity in adapting a technical tool to their professional convictions and practices.

In general, we can say that many translators do not post-edit MT, but insert chunks from the NMT.tmx into their translations, in much the same way as they use other reference material. This is because they consider PE a high-risk activity: they report that if the NMT output is inserted automatically into a segment, they tend to be deceived by its linguistic fluency and oversee errors. If they first take the time to read and understand the source language sentence and create a mental structure of the equivalent sentence in Hungarian, they can safely insert elements (or whole sentences) of the NMT output. Some do not even copy/paste chunks, but re-type

them. They report processing information in their head while they are typing. Others use the CAT tool's predictive typing function to prompt suggestions in context. These findings confirm the conclusion of Cadwell et al. (2016) that, for some DGT translators, the job means 'being in control of the final outcome'.

Other translators auto-populate and post-edit their text segment by segment with NMT output when no other TM is available. But permutations of these methods were also reported.

Very few translators pre-translate the whole text with the various reference materials and post-edit the translation as a whole.

Those who pre-translate or auto-populate segment by segment reported some emerging practices to improve the efficiency of their work. They systematically delete sentences that exceed 30–40 words; they do not understand at first reading; or contain references to legal acts. However, some translators consciously try to retain long sentences and to find useful chunks.

Over time, many translators have developed a conscious strategy as to whether to read the source segment or the NMT output first. A minority read the latter first, in order to decide whether it is of sufficient quality. If so, they go on to read the source sentence and compare the meaning. They then decide whether to delete the NMT output in part or in full.

The majority read the source sentence first. They argue that this prevents them from being influenced by a wrong or unfortunate rendering of the source sentence before they have understood its meaning or formed a mental structure of the target sentence. They say that NMT output can be very misleading and, once read, is difficult to depart from.

Some interviewees were not able to say which language version they read first. They had either never reflected on the issue or assumed that they read both versions in parallel or in batches alternately.

Only one translator indicated that she read the NMT output as if she was revising a translation, whereas several stated explicitly that they do not read NMT output as a text to be revised.

In the final section we discuss the benefits and drawbacks that the translators identified when using NMT.

4.4 Advantages and risks

Two advantages of using NMT that translators mention repeatedly are that it speeds up their work and reduces typing effort.

---

[8] All translators in the Hungarian department carry out revision tasks.

However, the question of time gain is highly controversial among translators. Around half of the interviewees report that using NMT allows them to produce translations more quickly. In fact, for some, speed is the only advantage. Some add that the time gained is mitigated by a loss in quality: NMT allows them to produce lower quality more quickly.

The other half were either unsure as to whether they work faster with NMT, saying that this varies greatly depending on the document, or suggest that NMT use might speed up translation but slow down revision (for revisers' comments, see section 4.2).

Nearly half report that NMT reduces typing effort. Some stress that they either cannot touch-type or are slow typers and benefit from being able to insert whole chunks into their texts. However, some emphasise that they prefer typing: while they are typing, they are mentally preparing their translation. For them, typing is not an additional chore to translation but integral into the complex cognitive process of translation.

The perceptions of greater speed and less typing effort may be interrelated: because translators are doing less typing, they may feel that they are processing their text faster. Nevertheless, the claim of greater speed has not been tested.

A very interesting set of themes that arose from the interviews were the perceived psychological and cognitive benefits of using NMT. A quarter of the interviewees said that they found it reassuring not to have a blank segment and to have to start a translation from scratch. Some mention that the NMT output is a source of inspiration, especially when they are tired or do not know how to deal with a construction. Cadwell et al. (2016) found this to be an unusual reason for using MT.

Several translators reported that they had sometimes been highly impressed by the eloquent solutions that NMT offers.

We can only speculate as to the reasons for the readiness of DGT translators to embrace MT. One explanation may be that, as institutional translators, they are used to integrating translations from different sources.

A further benefit that was mentioned is that PE requires less cognitive effort than translation: it is easier to find and correct errors in an existing text than to create a new text. However, by no means all translators share this view. Some find PE more cognitively demanding than translation. They also argue that accepting NMT solutions is less demanding than improving them. This drives

translators into a passive role which may affect their translation skills in the long run. Cadwell et al. (2016) report that some DGT translators reject MT because of its potentially detrimental effect on their abilities.

As a disadvantage of using NMT, some translators mentioned that they might tend to skim the NMT output, instead of reading it in the depth required to judge whether it correctly renders the meaning of the original.

In this context, it was also claimed that PE was killing translators' creativity. The word 'creativity' seemed to be used here in the sense of being able to produce a new text, a process accompanied by attention, focus and an active mindset, as opposed to a passive attitude that soon leads to a lack of attention. This sense of 'creation' is a source of motivation for some translators, which they will lose if they are only required to 'clean up' a text. Herbig et al. (2019) emphasise the need to improve translators' motivation by appropriately addressing cognitive load in the PE process, in order to avoid exhaustion and boredom, the effects of which may lead to higher translation costs.

Finally, some translators note that, in the absence of explicit expectations, the availability of NMT output (of 'good enough' quality) makes it possible to produce a translation with minimal human intervention. This constitutes a high risk for the quality of the product.

5 Conclusions

Overall, many of the translators interviewed consider NMT useful, but the majority have reservations about it in their daily work. This is partly because the quality of NMT output varies greatly from segment to segment and cannot be predicted reliably. Translators tend to recognise typical errors and see NMT as a tool that helps them do their work. However, many also see significant drawbacks and know that MT is not a substitute for human translation. As a result, they have developed working methods that compensate for the perceived disadvantages and give them 'control'. In practice, these methods mean that these translators translate rather than post-edit. Unfortunately, some of the working methods create new risks. A number of errors in the NMT output do not seem to be effectively corrected in the PE/translation phase, which puts pressure on revisers. These phenomena need to be analysed for quality assurance purposes.

The wide range of (sometimes conflicting) views among translators may not be new or unique to the use of NMT, but may be a result of reflection on new circumstances. Having to adapt to a new tool makes one more aware of differences that already existed and may play a more significant role in the future.

References

Cadwell, Patrick, Sharon O'Brien and Carlos S. C. Teixeira. 2017. Resistance and accommodation: factors for the (non-)adoption of machine translation among professional translators. *Perspectives*, 26(3):301-321.

Cadwell, Patrick, Sheila Castilho, Sharon O'Brien and Linda Mitchell. 2016. Human factors in machine translation and post-editing among institutional translators. *Translation Spaces*, 5(2):222-243.

Castilho, Sheila, Joss Moorkens, Federico Gaspari, Rico Sennrich, Andy Way and Panayota Georgakopoulou. 2018. Evaluating MT for massive open online courses. *Machine Translation,* (2018)32:255-278.

Dragsted, Barbara. 2006. Computer-aided translation as a distributed cognitive task. *Pragmatics & Cognition,* 14(2):443-464.

Drugan, Joanna, Ingemar Strandvik and Erkka Vuorinen. 2018. Translation quality, quality management and agency: principles and practice in the European Union institutions. In Joss Moorkens, Sheila Catilho, Stephen Doherty and Federico Gaspari (eds.), *Translation quality assessment: from principles to practice*. Berlin: Springer, 39-68.

Eisold, Andreas. 2017. Zur Rolle der Terminologie in der maschinellen Übersetzung. Jörg Porsiel, (ed.) *Maschinelle Übersetzung: Grundlagen für den professionellen Einsatz*, Bundesverband der Dolmetscher und Übersetzer, 109-125.

Guerberof Arenas, Ana. 2014. Correlations between productivity and quality when post-editing in a professional context. *Machine Translation,* 28:165-186.

Herbig, Nico, Santanu Pal, Mihaela Vela, Antonio Krüger and Josef van Genabith. 2019. Multi-modal indicators for estimating perceived cognitive load in post-editing of machine translation. *Machine Translation,* (2019). https://doi.org/10.1007/s10590-019-09227-8

Ive, Julia, Aurélien Max and François Yvon. 2018. Reassessing the proper place of man and machine in translation: a pre-translation scenario. *Machine Translation,* 32(4):279-308.

Knowles, Rebecca, Marina Sanchez-Torron, and Philipp Koehn. 2019. A user study of neural interactive translation prediction. *Machine Translation,* (2019). https://doi.org/10.1007/s10590-019-09235-8

Koehn, Philipp and Ulrich Germann. 2014. The impact of machine translation quality on human post-editing. *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation,* Gothenburg, Sweden. Association for Computational Linguistics, 38-46.

Koehn, Philipp and Rebecca Knowles. 2017. Six challenges for neural machine translation. *Proceedings of the First Workshop on Neural Machine Translation*, Vancouver, Canada, August 4, 2017. 28–39.

Koponen, Maarit. 2016. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, 25:131-148.

Läubli, Samuel, Rico Sennrich, Martin Volk 2018. Has machine translation achieved human parity? A case for document-level evaluation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 4791–4796.

Moorkens, Joss, Antonio Toral, Sheila Castilho and Andy Way. 2018. Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces* 7:2, 240–262.

Plitt, Mirko, and François Masselot. 2010. A Productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague Bulletin of Mathematical Linguistics,* 93:7-16.

Tihanyi László and Csaba Oravecz. 2017. First experiments and results in English-Hungarian neural machine translation. *Proceedings of the 13th Conference on Hungarian Computational Linguistics (MSZNY 2017),* Szeged, Hungary, 275-286.

Yamada, Masaru. 2019. The impact of Google Neural Machine Translation on post-editing by student translators. *The Journal of Specialised Translation*, 31:87-106.

Van Brussel, Laura, Arda Tezcan, and Lieve Macken. 2018. A fine-grained error analysis of NMT, PBMT and RBMT output for English-to-Dutch. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, May, 3799-3804.

Way, Andy. 2018. Traditional and emerging use-cases for machine translation. In: Joss Moorkens, Sheila Castilho, Federico Gaspari, Stephen Doherty (eds.) *Translation Quality Assessment: From principles to practice*. Springer, Berlin, 159–178.