POST-EDITING OF MACHINE TRANSLATION OUTPUT WITH AND WITHOUT SOURCE TEXT

Matej Kovačević

Abstract

Post-editing of machine translation output is a practice which aims to speed up translation production. There is still no agreement on whether post-editors should have access to the source text of the translations they are post-editing. The aim of this paper is to see how access to source text influences post-editors' quality of work and their speed. An experiment was conducted among 22 graduate students of English, who post-edited two translations produced by Google Translate. The subjects were divided into two groups, with access to the source text for only one of the translations. Task duration and the number of corrected errors in the MT output were measured. Types of errors were further analysed. Contrary to expectations, access to source text was found not to have a considerable impact on speed. As expected, it did have an impact on the quality of the final translation.

1. Introduction

Machine translation (MT) systems have been available to translators for 60 years, but they still cannot produce perfect translations. This is the reason why people are still apprehensive when it comes to using such systems. The development of MT systems has introduced a practice called post-editing of MT output whereby a machine translates the text and the translator (post-editor) revises the translation. This paper will explore that practice, more specifically the issue of whether the post-editor should have access to the source text (ST) when post-editing a translation. The research questions were based on the following statement by Rico and Torrejón (2012: 168): "In the translation industry, the question of whether the post-editor should get access to the source text is still under consideration as in some contexts it is deemed as a barrier to reaching optimal productivity."

Just like any other industry, the translation industry aims to produce as much content as possible in the shortest amount of time and with suitable quality. It is therefore important that the post-editor does not waste too much time going back to the source text. On the other hand, MT can sometimes be impossible to understand without access to the source text, which makes it impossible to post-edit such a translation. The aim of this paper is therefore to explore to what extent having access to the source text influences the quality of the post-edited translation, that is, the final translation, and the time that post-editors need to do their job.

The paper begins with an overview of the practices and concepts which were important for choosing the research questions. After that, the hypotheses and the methodology will be explained and, finally, the results will be presented. In the conclusion, some ideas for further research are outlined.

2. Machine translation

Machine translation is the process of translating a text from one natural language into another by a computer, without any human involvement. The development of machine translation and MT systems began in the 1930s and it has continued until today. Early MT systems produced translations using only bilingual dictionaries and paid little or no attention to syntax. In the 1980s advanced technology enabled MT systems to analyse sentences with regard to syntax, morphology, and even semantics (Dovedan, Seljan and Vučković 2002). At that time, MT systems were primarily developed by governments for military and diplomatic purposes. The US Air Force used the Systran system for translating important documents from Russian, and the European Commission used the same system for translations from French. In the 1990s, MT systems were gradually introduced into the commercial sector, mostly for translating all types of manuals into as many languages as possible. The same decade brought about an increased use of MT systems on personal computers, and in 1997 LANT launched the first online MT system intended for translating e-mails and webpages. Since then, MT systems have constantly been developed and updated (Hutchins 1999).

According to the Systransoft (2014) webpage, there are three major approaches to MT:

1. rule-based – such MT systems use built-in linguistic rules and a great number of bilingual dictionaries to create translations. They analyse the sentences of the source text, after which they transfer their grammatical structures into the target language. They usually offer greater quality of translation but they have high initial and maintenance costs;

2. statistical – these MT systems generate translations using statistical models based on corpora that consist of translations done by human translators. They analyse the texts from the corpora, interpret the connections and offer solutions. Initial costs for such systems are low, but they require large multilingual corpora, extensive hardware and excellent programmers' knowledge in order to provide good-quality solutions;
3. example-based – these MT systems also contain corpora, but in their case source text sentences and sentence elements are compared to sentences from the corpora, and translations are created based on existing sentences with similar elements (Duh 2005).

Each approach has some advantages and disadvantages, and there is still no system which can consistently produce high quality translations in any field.

Some MT systems use sublanguages, which Luckhardt (1991: 306) claims are good "for solving some of the notorious problems in MT such as disambiguation and selection of target language equivalents". Sublanguages are essentially natural languages with adaptations and limitations applied to grammar, vocabulary, syntax and semantics, which then facilitate MT of texts written in those languages. Translations produced in this way are of higher quality (they are up to 95 % correct) than translations from ordinary languages, but they do require some post-editing (Seljan 2000: 17). According to Luckhardt (1991: 308), sublanguages can be best applied when there is a good terminological database for the field in question and when there is a significant amount of similar texts for translation from a specific field.

MT software and systems have advantages over traditional, fully human translation, but they also have numerous disadvantages. They are faster and they can be reliable, depending on the type of text they are translating. On the other hand, most of them are still very expensive, which makes them unavailable to individuals, while the quality of translation depends on the language pair in question (Hutchins 1999). Free online MT systems are an alternative which is available to anyone, but the quality of their translations is still too low to be used commercially without human post-editing. According to Hutchins' statement from 2001, which is still valid today, "all current commercial and operational [MT] systems produce output which must be edited (revised) if it is to attain publishable quality. Only if rough translations are acceptable can the output of MT systems be left unrevised." Since the first steps in developing MT systems, the ultimate goal has been to get a system capable of producing fully automated high-quality translation (better known as FAHQT), but Krings and Koby (2001: 15) think that this dream is "just as difficult to achieve today as it was in the 1960s". Whatever the case may be, it is certain that MT systems still depend heavily on human translators and other language experts.

2.1 Google Translate

Provided by Google Inc., Google Translate is a multilingual online MT system. It appeared in 2006 and at first it could only translate between English, Arabic, German, French and Spanish, while at the moment of writing this paper it can translate between 80 different languages, using English as the pivot language if necessary (Google 2014). Research has shown that the system works best with translations from French and Italian into English, and it generally provides very good translations from all languages of the European Union into English. The reason for this is most probably the fact that its corpus contains all the documents that the EU has published, which is an excellent base for providing good-quality translations (Wikipedia 2014).

The system translates using a statistical approach and it was based on Systran until 2007, when Google introduced their own translation service. Like any other statistical MT system, Google Translate uses a corpus which "includes all the paper put out since 1957 by the EU in two dozen languages, everything the UN and its agencies have ever done in writing in six official languages, and huge amounts of other material" (Bellos 2011). Apart from using the corpus, Google Translate welcomes user feedback on its translations. When it offers a translation, users can change whatever they think is necessary and submit the "post-edited" translation back to Google. The system will then use the feedback in similar future translations. In effect, Google Translate uses millions of documents translated by humans in order to provide fast and free translations to whoever needs them.

In addition to the MT service, Google offers what it calls a "translator toolkit", which includes a translation memory system, term base system and a tool for translating websites. Furthermore, it offers translator applications for smartphones with an optical character recognition option which enables them to translate text in photographs taken with the phone. It is a user-friendly tool which can help Internet users understand at least the gist of foreign language websites and texts instantly.

3. Post-editing of machine translation output

The term post-editing refers to the practice of revising translations that have been produced by a MT system. MTs are mostly considered to be unfinished, which is why this practice evolved (Allen 2003: 297-8). Post-editing as a profession was first mentioned by Vasconcellos and León (1985), after which it was recognised and it became a common term. Allen (2003: 297) describes a post editor as a person whose task is "to edit, modify and/or correct pre-translated text that has been processed by an MT system from a source language into (a) target language(s)." According to TAUS (2006), post-editing "involves linguistic more than subject area skills and is performed best by alert translators, familiar with machine output, working in a standard translation environment." Other scholars agree on the issue of the qualifications of a post-editor. For example, Krings and Koby (2001: 12) say that "the post-editor must be a translator [because] only a translator can judge the accuracy of a translation." They further say that a post-editor needs "linguistic, technical and problem-solving skills" (2001: 16), while Rico and Torrejón (2012: 169) use the terms linguistic skills, instrumental competence, and core competences. Regardless of different terminology, it is clear that they agree on the competences a good post-editor should have in order to produce maximum output of desired quality.

In 1985, Vasconcellos and León (1985: 122) reported that such a post-editor can produce 4,000–10,000 words of translation a day, which is two to three times more than the average output of a human translator. This question has been studied more recently as well. Thus Thicke (2011: 39) found that an average post-editor can produce 5,600 words of translation a day (compared to an average human translator who produces 2,500 words a day), while Flournoy and Duran (2009) found that post-editors can produce the benchmark 2,500 words a day in as much as two hours, which would make them four times faster than human translators. From this data it is clear that the productivity of post-editors makes post-editing of MT output a much better option whenever such translations would satisfy the users' needs.

According to Allen (2003), there are two main reasons for using MT in combination with post-editing. The first is increased focus on globalisation, which primarily refers to corporations and smaller companies. They can no longer rely on only one language in doing business because that way they would not be as successful as possible. Since human translation is sometimes too slow for their needs, they use post-editing of MT output in order to be able to publish information in as many languages as possible in the shortest amount of time. The second reason is the fact that some texts and documents do not have to be translated perfectly, which is why a rough translation is often enough. Sometimes users only need to see what the text is about or what the main information in the text is, and MT output with some or no post-editing serves that purpose well. Since there are free online MT systems, this can now be done quickly and easily without having to pay human translators.

The extent of post-editing of MT output can be very different depending on the purpose of a translation. Doherty and Gaspari (2013, emphasis in the original) say that post-editing serves different needs than revision of human translation and that "the aim of [post-editing] is to improve the output, not necessarily to make it perfect." Vasconcellos (1987) states that "with MT postediting, the focus is on adjusting the machine output so that it reflects as accurately as possible the meaning of the original text." In line with this, according to most authors (e.g. Allen 2003; Krings and Koby 2001) there are three types of post-editors' interventions: no post-editing, minimum post-editing and full post-editing. The approach is chosen on a case-by-case basis depending on the users' needs. For example, MT output is not post-edited at all when a user needs only the gist of a given text, while full post-editing is chosen when a translation is intended for publication for a wide audience. Minimum post-editing is most commonly used because that way the most important errors are corrected, but the post-editor does not spend too much time fine-tuning the translation (Allen 2003: 302-6).

Every post-editor should receive clear guidelines in order to do their job the best they can and to satisfy the clients' needs. Many authors report on post-editing guidelines or give some of their own (e. g. Allen 2003; Krings and Koby 2001; Rico and Torrejón 2012). From their work it is easy to see that there is no common set of guidelines which could apply to all translations and users. Allen (2003: 307-311) gives several examples of instructions for post-editors from corporations and institutions. Post-editors at General Motors use the SAE J2450 standard metric for translation quality which provides seven categories and two subcategories of errors by order of priority. On the other hand, post-editors at the European Commission Translation Service get their instructions in the form of "dos" and "don'ts". Since these are only guidelines, efficient post-editing training is of utmost importance in order to get translations which correspond to the guidelines and to achieve the highest possible productivity.

When discussing post-editing, it is important to note the comparisons of MT output with fully human translations. Bellos (2011) commented on the translations made by Google Translate, but this can apply to all MT systems:

> Of course, [Google Translate] may also produce nonsense. However, the kind of nonsense a translation machine produces is usually less dangerous than human-sourced bloopers. You can usually see instantly when GT has failed to get it right, because the output makes no sense, and so you disregard it. […] Human translators, on the other hand, produce characteristically fluent and meaningful output, and you really can't tell if they are wrong unless you also understand the source – in which case you don't need the translation at all.

Similar claims have been made by other authors with regard to revision and post-editing (e.g. Krings and Koby 2001). They say that misunderstandings of human translators can influence the whole text and ultimately lead to a completely incorrect translation, with the possibility of a reviser never noticing the errors and correcting them. In MT output, errors are more local and, while certain errors may be repeated several times in a translation, they are easier to spot and thus they are more often successfully corrected.

Regarding the best conditions for post-editing, Hutchins (2001) claims that "it is now widely accepted that MT proper works best in domain-specific and controlled environments", and other authors agree with this (Allen 2003; Torrejón and Rico 2002). Those environments are adapted to the way an MT system "thinks" and source texts are written in a way which makes them easiest to translate well. Krings and Koby (2001: 5) suggest three conditions that texts need to fulfil so that MT systems could produce high-quality output: they should be restricted to a specific domain of knowledge, they should conform to a formal syntax and semantics, and the MT system should be adapted to that domain and language. The language used in such situations is called controlled language. According to Torrejón and Rico (2002: 108), controlled languages "improve the readability of the documents by imposing clear and direct writing, they reduce syntactic and lexical ambiguities by applying grammatical and lexical constraints, and they also increase the translatability of the text, making it amenable to MT". The primary purpose of those languages is to be as consistent and as clear as possible, so that an MT system could achieve the highest possible quality. Torrejón and Rico (2001) give some examples of well-known controlled languages: AECMA Simplified English, which was one of the first, Boeing Technical English, and Controlled Automotive Service Language, among others.

As concluded above, MT systems and post-editing are mostly used by corporations, companies, institutions and services which can employ controlled languages and which can finance such systems. Examples of such users are the European Commission, the Pan American Health Organization, the US Air Force, Caterpillar Inc., General Motors, and companies that deal with language localisation in general (Krings and Koby 2001; Allen 2003; Vasconcellos and León 1985, among others). All of them have been using the services of post-editors for more than 20 years for different types of translations and with different views on how the final translation should look. One of the more recent users of MT is the European Patent Office, which has cooperated with Google and developed "a translation service optimised for patent documents" (European Patent Office 2013). The system is called "Patent Translate" and it can translate between English and 31 other languages, and between French and German and 27 other languages. Users looking for patent documents in other languages can use the system to search through patent databases and the system gives them instant translations of the patent documents. This system can produce good MT output because of the fact that patents are very similar and consistent documents, which can then be translated well by a machine. All of these are environments in which MT systems can produce output of decent quality and thus enable post-editors to work quickly and be as productive as possible. The goal of the MT industry is to achieve this in as many fields as possible and for as many languages as possible, but to simultaneously reduce the costs of maintaining and improving MT systems. It remains to see if, or maybe when, this goal will be achieved.

4. Research design

The following chapter will present the research that was conducted. It will begin with the aims of the research and the hypotheses, which were formed based on previous research. After that, it will present the methodology of the research, including subchapters about the test subjects, the texts that were used, the description of the experiment, and the details about the analysis of data obtained through the experiment.

4.1 Rationale, aims and hypotheses

As quoted above, Rico and Torrejón (2012: 168) conclude that there is still significant debate regarding the question of whether to provide post-editors with access to the source text of the machine translation that they are editing because it is feared that this might negatively influence their productivity. Krings (in Krings and Koby 2001) conducted a study which examined that issue using translations from English into German made by SYSTRAN and translations from German into English made by METAL machine translation system. He had test subjects who rated the raw MT output and the final translations sentence by sentence on a scale from 1 to 5 (applying whatever criteria they thought were the best). In the second part, he quantified the errors in the MT output and analysed how many of those errors were corrected by post-editors who did not have access to source texts. All of this was done using a think aloud protocol, and the results showed that "four-fifths [(79%)] of all machine translation errors could thus be repaired, even without the availability of a source text" (2001: 273). Ratings for most post-edited sentences were higher, and only several were rated lower than the original machine-translated sentences. However, the ratings were still very low for those sentences which were translated poorly in the first place. Krings' study showed that it is possible to recognize and correct most errors in MT output without having access to the source text, but it also showed that the 20% of the errors which were not corrected proved to be very problematic and often could not be detected in any way (2001: 273)

Krings only wanted to see if decent post-editing was possible without having access to the source text and he studied only those translations which were made under such conditions. He did not make a comparison to translations produced with access to the source text. Furthermore, he did not measure the time necessary for post-editing, which means that he could not make any claims as to the differences in productivity between the two methods. This is precisely what the present study aims to do. The goal is to explore the differences in quality between translations post-edited with access to the source text and those without access to the source text. In addition, the study sets out to examine the time difference between the two conditions in order to determine whether access to source text during post-editing spells lower productivity.

Based on Rico and Torrejón's (2012: 168) statement that having access to the source text might negatively influence post-editors' productivity, it was expected that post-editing without the source text would be faster. On the other hand, there are reasons to fear that without access to the ST post-editors may have problems comprehending sections of MT output if the meaning is severely distorted. For this reason, the following hypotheses were formulated:

Hypothesis 1: Post-editors work faster when they do not have access to the source text, i.e. their productivity is higher in such conditions;

Hypothesis 2: The quality of the final translation – post-edited MT output – produced without access to the source text is poorer than the quality of MT output post-edited with access to the source text. This means that, without access to the source text, fewer errors made by the machine are detected, or even that new errors are introduced due to comprehension problems.

4.2 Methodology

4.2.1 Subjects and environment

The study was conducted in the form of an experiment that involved twenty-two subjects who were all graduate students taking the Translation Track in the Department of English at the University of Zagreb. At the time of the experiment, they had all taken at least three translation courses in which they had worked regularly on translation projects from English into Croatian and vice-versa. For most of them, one of those courses had been on translation for the European Union institutions, which means they were familiar with the subject matter. All of the subjects' L1 was Croatian, while their L2 was English, and they were all approximately the same age and had similar education backgrounds.

The experiment was conducted in a neutral environment in a computer laboratory. The subjects worked on desktop computers using MS Office 2010. It was impossible for all the subjects to be there at the same time, so two sessions were organized, five days apart from each other. This problem was explained to the subjects and they were asked not to tell their colleagues who were not there anything about the experiment, which they agreed to do.

4.2.2 Texts

The texts that the subjects had to post-edit were about the European Union and its functioning. The Croatian source text (see Appendix A) was from Entereurope, a Croatian website which provided information before the country's accession to the European Union in July 2013. The English source text (see Appendix B) was selected from the European Union's official Europa server and it was about decision-making in the EU. The texts were comparable in length (the Croatian text had 113 words, and the English text had 127 words), in the number of errors present in Google Translate translations of the texts (the translation into English had 22, while the translation into Croatian had 29 errors), and in the topic they dealt with. The MT output that had to be post-edited comprised translations of those texts made by Google Translate (see Appendices A and B).

4.2.3 Experiment

The subjects were divided into two groups – Group A and Group B. Before they arrived, the documents they would have to work with were saved on the computers in the laboratory. Each of the subjects received a sheet with detailed instructions. They were not aware of what they would have to do before the experiment itself. When they arrived, they were asked to read the instructions carefully, then the instructions were repeated orally, and finally they got the chance to ask questions. They were instructed to change anything they thought was necessary in the translations, but without any help from print dictionaries or the Internet. The texts they had to post-edit were specifically chosen to cover a topic which the subjects were familiar with so that they would not have any problems understanding and post-editing them without using any additional resources.

Each of the subjects received two documents – the first one contained the translation from Croatian (HR) into English (EN) (Translation 1), while the second one contained the EN -> HR translation (Translation 2). Each group had access to the source text for only one of the translations – Group A had it for Translation 1, while Group B had it for Translation 2. The order of the tasks was reversed in order to counter the "retest effect". The division of tasks is shown in Figure 1. The subjects worked in MS Word using the Track Changes feature. Since they had all done translation tasks during their studies, they were familiar with the feature and they knew how to use it.

The subjects were instructed to turn on Track Changes, write "START" at the beginning of the documents as soon as they opened them, and write "END" at the end of the documents just before closing them after finishing their work. This was deemed to be the easiest way to measure the time it took to complete the tasks as each change applied with Track Changes leaves a timestamp. When the subjects had completed each task, they saved the final translations under codes and the documents were collected. This process was repeated in the second session and there were no problems in either of the two sessions.

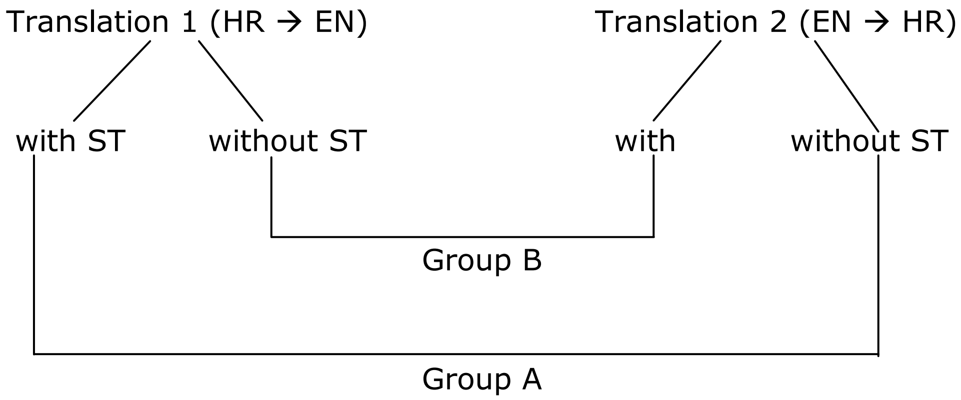

Figure 1 Division of tasks

4.2.4 Data analysis

The data obtained from the experiment was in the form of forty-four MS Word documents, two per subject. This is not a large sample, but it should show general trends regarding the research questions which are explored in this study. Since the subjects used the Track Changes feature, every change they made was marked within the text and time-stamped, so we know what they changed and when exactly they did that. In order to test the first hypothesis, the first set of data that was analysed was the time the subjects needed to post-edit the MT output. The data were quantified by simple calculation because the subjects marked when they started and finished working.

The second set of data was more difficult to analyse. The criterion which the second hypothesis was based on was the amount of errors that were corrected in the MT output. In order to compare this data, it was necessary to analyse all forty-four documents and quantify the errors which were corrected, the errors which went undetected, and possibly the errors introduced by the post-editors. Only linguistic rather than style-related errors were quantified, because of difficulties related to assessing style. In his study, Krings (Krings and Koby 2001: 267) divided the errors in post-edited translations into eleven categories, but some of them could not be applied to the MT output used in this study and some were too narrow for this situation. They were therefore adapted to six categories, with the addition of errors on the textual level, as used by Pavlović (2007: 83-4). The resulting categories were as follows:

1. Textual errors – errors on the level of the text (changes in sentence boundaries, deictic words and other devices providing cohesion among the sentences);
2. Semantic errors – errors in interpreting and translating meaning;
3. Lexical errors – wrong choice of words and parts of speech;
4. Syntactic errors – errors in establishing proper relations on the level of the sentence;
5. Morphological errors – errors in word formation;
6. Orthographical errors – errors in spelling, punctuation and capitalization;
7. Other errors – missing words, unnecessarily added words, typographical errors.

These categories were chosen with the aim of covering all the errors that Google Translate made and all the errors that the post-editors might have made during their work. Examples of errors from each category can be found in Table 1 (ST elements, GT elements, literal translations of errors in GT elements in square brackets and reference translations, REF).

Table 1 Examples of error types

Error Type	Example
Textual (TEX)	ST: U veljači <u>2002. godine Europska</u> je unija sazvala međuvladinu konvenciju… GT: In February <u>2002. The</u> European Union has convened an intergovernmental convention… [U veljači <u>2002. (sentence break) Europska</u> je unija sazvala međuvladinu konvenciju…] REF: In February <u>2002, the</u> European Union...
Semantic (SEM)	ST: … correspond to the needs of <u>those most concerned</u>… GT: … odgovaraju potrebama <u>onih koji najviše brine</u>… […correspond to the needs of <u>those which worry the most</u>…] REF: ... odgovaraju potrebama <u>onih kojih se najviše tiču</u>...
Lexical (LEX)	ST: … it <u>assesses</u> the potential economic, social and environmental consequences… GT: … <u>ocjenjuje</u> potencijalne gospodarske, socijalne i ekološke posljedice… [… it <u>rates</u> the potential economic, social and environmental consequences…] REF: ... <u>procjenjuje</u> potencijalne gospodarske, socijalne i ekološke posljedice…
Syntactic (SYN)	ST: Konvencija <u>je tijekom jednoipolgodišnjeg rada</u>… GT: Convention <u>is over one and a half of work</u>… [Konvencija <u>je tijekom jednog i pol rada</u>…] REF: <u>Over the course of a year and a half</u>, the Convention...
Morphological (MORPH)	ST: <u>Groups of experts</u> give advice on technical issues. GT: <u>Grupe stručnjaci</u> daju savjete o tehničkim pitanjima. [<u>Groups experts</u> give advice on technical issues.] REF: <u>Skupine stručnjaka</u> daju savjete o tehničkim pitanjima.
Orthographical (ORT)	ST: Before the <u>Commission</u> proposes… GT: Prije <u>komisija</u> predlaže… [Before the <u>commission</u> proposes…] REF: Prije nego što <u>Komisija</u> predloži...

Other (OTHER)	ST: … to deal with an issue at national rather than EU level. GT: … da se bave nekom pitanju, na nacionalnoj razini, a ne EU. [… to deal with an issue, at national level, rather than the EU.] REF: ... se nekim pitanjem bave na nacionalnoj razini, a ne na razini EU-a.

The data obtained through this analysis helped to determine how successful both groups of post-editors were in correcting the errors made by Google Translate, as well as to examine the differences between the corrections of those post-editors who had access to the source text and those who did not. The results were quantified in terms of total values and mean values in order to see the exact relations between them.

5. Findings

5.1 Hypothesis 1 – duration of post-editing

The first hypothesis had to do with the time that the subjects needed for post-editing. The post-editing was expected to be faster when the post-editors worked only with the MT output without access to the source text. Table 2 shows how much time each subject needed for post-editing the first and second translation.

Table 2 Duration of post-editing (minutes)

	Translation 1 (HR → EN)		Translation 2 (EN → HR)	
Condition / Subject	With ST (Group A)	Without ST (Group B)	With ST (Group B)	Without ST (Group A)
Subject 1	12	12	15	7
Subject 2	15	15	11	13
Subject 3	22	22	12	11
Subject 4	11	18	11	9
Subject 5	16	17	13	12
Subject 6	17	19	16	10
Subject 7	25	12	10	13
Subject 8	23	20	14	16
Subject 9	7	14	8	15
Subject 10	14	16	11	15
Subject 11	14	15	17	10
Total	***176***	***180***	***138***	***131***
Mean	*16*	*16.36*	*12.54*	*11.9*

As is visible from the results provided above, although post-editing times vary considerably between different subjects, total times per group and text are very similar, as are the mean values. In both translations, subjects from Group A were slightly quicker on average – in Translation 1 they were quicker by 21.6 seconds (which is a 2.25% difference), while in Translation 2 they were quicker by 38.4 seconds (which is a 5.38% difference). The results also show that the subjects were almost 25% quicker when they were post-editing translations into Croatian, which is their mother tongue.

The results obtained from the experiment correspond to the hypothesis that the subjects without the source text would be faster only for Translation 2, but in both cases the mean values are very close. For this reason, it would be far-fetched to say that either option was really quicker and that the hypothesis has been confirmed or refuted.

5.2 Hypothesis 2 – quality of final translations

The second hypothesis had to do with the quality of final translations. The quality was assessed by comparing the number and types of errors in the Google Translate machine translation output to the number and types of errors undetected or added by the post-editors. The tables below show the overall number of errors in the MT output and the number of errors that were still present (or new errors introduced) in the post-edited translations of every subject. In this way it was possible to see how successful the subjects were in correcting the errors in the MT output and whether having access to the source text had an impact on their work. Error analysis for Translation 1 will be presented first. The results for both groups are shown in Tables 3 and 4. Subjects' mean values that are greater than the number of errors in the MT output are presented in bold.

Table 3 Error analysis for Translation 1 (HR → EN), Group A, with ST

	Type of error							
Subject	TEX	SEM	LEX	SYN	MORPH	ORT	OTHER	**Total**
MT output	1	2	2	6	6	2	3	22
Subject 1	/	/	1	1	1	1	3	7
Subject 2	/	/	3	4	2	1	/	10
Subject 3	/	/	2	2	5	/	3	12
Subject 4	/	/	2	2	3	5	1	13
Subject 5	/	/	4	2	2	2	2	12
Subject 6	1	1	2	6	1	2	2	15
Subject 7	/	/	3	5	1	2	1	12
Subject 8	/	1	3	3	2	1	1	11
Subject 9	/	1	3	5	3	1	2	15
Subject 10	/	/	2	5	1	/	/	8
Subject 11	/	/	2	1	1	3	2	9
Total (subjects)	1	3	27	36	22	18	17	124
Mean (subjects)	0.09	0.27	**2.45**	3.27	2	1.63	1.54	11.27

Table 4 Error analysis for Translation 1 (HR → EN), Group B, without ST

	Type of error							
Subject	TEX	SEM	LEX	SYN	MORPH	ORT	OTHER	**Total**
MT output	1	2	2	6	6	2	3	22
Subject 1	/	4	4	1	2	2	1	14
Subject 2	1	2	3	3	2	5	4	20
Subject 3	/	7	3	1	/	5	1	17
Subject 4	/	5	2	3	1	1	/	12
Subject 5	/	5	4	3	1	5	2	20
Subject 6	/	2	3	1	/	2	3	11
Subject 7	/	2	3	5	2	2	2	16
Subject 8	1	4	2	4	/	/	3	14
Subject 9	/	3	2	3	2	/	1	11
Subject 10	/	2	4	2	2	1	/	11
Subject 11	/	3	2	6	/	2	1	14
Total (subjects)	2	39	32	32	12	25	18	160
Mean (subjects)	0.18	**3.54**	**2.9**	2.9	1.09	**2.27**	1.63	14.54

The results above show that the subjects from Group A, who had access to the source text for Translation 1, were more successful in correcting the errors in the MT output. On average, their final translations had 3.27 errors less (29%) than the translations post-edited by Group B, who worked without access to the source text. Compared to the number of errors in the MT output, Group A had 49% less errors, while Group B had 34% less errors.

The largest and probably the most important difference between the two groups in Translation 1 was in the number of semantic errors. As was explained above, those are the errors in interpreting and translating meaning, which means that they are the most serious type of error because they change the meaning of individual sentences, and maybe even the entire text. While the final translations produced by the subjects working with access to the ST contained only 3 such errors in total, the final translations produced by the subjects working without access to the ST contained a total of 39 semantic errors, which is 3.54 per subject. The final translations of seven of the eleven subjects working without access to the ST contained more semantic errors than were present in the MT output they were asked to edit. The remaining four translations in this group contained the same number of errors as the MT output. The reason for this is the fact that most of these subjects did not detect the errors made by Google Translate, and they even misunderstood other parts of the translation and applied corrections which changed the meaning of the text.

As for the other types of errors, the subjects working with access to the ST made fewer textual, lexical, and orthographical errors, while the subjects working without access to the ST made fewer syntactic, morphological and 'other' errors. The one textual error made by Google Translate in this text was fairly obvious even without the source text – the engine divided one sentence into two at a point where a year was mentioned because in Croatian years are written with a dot. Because of this, neither of the sentences made any sense, but almost all of the subjects noticed this error and corrected it. Google Translate made two lexical errors, while both groups of subjects made more on average – the subjects who worked with the source text made 2.45 errors, while the subjects who worked without the source text made 2.9 errors. Almost none of the subjects detected the two lexical errors made by Google Translate, while some even introduced additional ones, often by changing the name of the *Treaty establishing a Constitution for Europe*, which was correctly translated in the MT output.

Syntactic errors were a problem for Google Translate, as well as for both groups of subjects. Even though the subjects corrected approximately half of those errors in the MT output, they still made 2.9 and 3.27 syntactic errors per subject. However, considering that this category included errors with articles, which are fairly common, this number is not surprising. Both groups were able to correct most of the morphological errors (66% for the subjects working with access to the ST, and 80% for the subjects working without access to the ST), and almost half of 'other' errors. As for the orthographical errors, the subjects working with access to the ST had 18.5% less than the MT output, while the subjects working without access to the ST had 13.5% more. On average, the subjects working with access to the ST made additional errors only in the lexical category, while the subjects working without access to the ST made them in the semantic, lexical and orthographical categories.

Translation 1 was done from Croatian into English, which is the subjects' second language. Translation 2 was done into Croatian, which is their mother tongue, and the results are presented in Tables 5 and 6. Once again, subjects' mean values that are greater than the number of errors in the MT output are presented in bold.

Table 5 Error analysis for Translation 2 (EN → HR), Group B, with ST

	Type of error							
Subject	TEX	SEM	LEX	SYN	MORPH	ORT	OTHER	**Total**
MT output	1	5	2	1	10	6	4	29
Subject 1	1	1	4	1	3	3	1	14
Subject 2	1	1	1	/	4	3	/	10
Subject 3	/	2	3	/	7	4	/	16
Subject 4	1	1	5	/	2	3	/	12
Subject 5	/	/	3	/	3	2	2	10
Subject 6	/	1	/	/	/	/	/	1
Subject 7	1	1	3	/	4	2	2	13
Subject 8	1	1	1	/	1	4	1	9
Subject 9	1	1	2	/	2	1	1	8
Subject 10	1	2	1	/	/	1	/	5
Subject 11	1	2	/	/	4	/	/	7
Total (subjects)	9	13	23	1	30	23	7	105
Mean (subjects)	0.81	1.18	**2.09**	0.09	2.72	2.09	0.63	9.54

Table 6 Error analysis for Translation 2 (EN → HR), Group A, without ST

	Type of error							
Subject	TEX	SEM	LEX	SYN	MORPH	ORT	OTHER	**Total**
MT output	1	5	2	1	10	6	4	29
Subject 1	1	1	1	/	2	6	1	11
Subject 2	1	2	2	1	3	3	3	15
Subject 3	/	2	2	1	4	6	3	18
Subject 4	1	/	2	/	/	1	2	6
Subject 5	1	1	2	/	4	/	/	8
Subject 6	/	4	2	/	1	2	1	10
Subject 7	1	3	2	/	4	1	1	12
Subject 8	1	2	1	/	5	2	/	11
Subject 9	1	4	1	/	5	1	1	13
Subject 10	1	1	3	1	1	4	/	11
Subject 11	1	1	2	/	2	3	/	9
Total (subjects)	9	21	20	3	31	29	12	124
Mean (subjects)	0.81	1.91	1.81	0.27	2.81	2.63	1.09	11.27

As the results above indicate, the subjects who had access to the source text (Group B) were again more successful than those who did not (Group A). The former group’s final translations had, on average, 1.73 fewer errors (15%) than the translations made by the latter group. Compared to the translation produced by Google Translate, which contained 29 errors, Group B had 67% less errors, and Group A had 61% less errors.

As in Translation 1, the difference in the average number of errors between the two groups was largest for semantic errors: the final translations produced by subjects working with access to the ST contained a total of 13 errors, compared to 21 in the translations produced by subjects working without access to the ST. However, in Translation 2, the difference is considerably smaller and both groups were more successful in correcting the semantic errors made by Google Translate than in Translation 1. The MT output contained five semantic errors and subjects working with access to the ST corrected 76.4% of them, while subjects working without access to the ST corrected 61.8% of them.

In Translation 2, Group B, who had access to the source text, had fewer errors than Group A on average in all the categories except for the textual and lexical categories. Google Translate made one textual error with a demonstrative pronoun, which most of the subjects did not correct. In this category, both groups had the same average result. Surprisingly, the final translations of the group with access to the source text contained more lexical errors (2.09) than the group working without access to the ST (1.81), and even more than Google Translate (2). Once again, the subjects mostly failed to detect the two lexical errors made by Google Translate, while some of them even introduced additional errors by changing correct solutions to incorrect ones.

As for the syntactic errors, both groups were successful in correcting the single error which Google Translate made in its translation. Morphological errors, which were expectedly the largest group in the MT output, mostly due to the relative complexity of the Croatian morphology, were also successfully corrected. Google Translate made 10 morphological errors, of which subjects working with access to the ST corrected on average 7.28, while subjects working without access to the ST corrected 7.19. The subjects also successfully corrected more than half of the orthographical and 'other' errors. In Translation 2, additional errors were made by one of the groups in one category: it was Group B (subjects with access to the ST) with lexical errors.

6. Conclusion

The aim of this study was to determine whether having access to the source text influences the work of post-editors. Based on other authors' work and knowledge about translation processes, it was assumed that post-editors who had access to the source text would be slower and that they would produce translations with fewer errors. The first hypothesis, which said that post-editing would be faster if post-editors did not have access to the source text, was proven to be inconclusive because it was correct for only one of the two translations and the difference was slim (just over 21 seconds). This indicates that speed/productivity might not be negatively affected if post-editors have access to the source text. Of course, more research should be done to confirm this tentative conclusion.

The second hypothesis, which said that the quality of the final translations would be higher if the post-editors had access to the source text, was confirmed for both translations. The subjects who had access to the source text were more successful in correcting the errors, especially the semantic ones, made by Google Translate. This indicates that the source text is most useful for correcting such errors, and they are probably the most important type. The results have also shown that some types of errors (syntactic, morphological, and orthographical) can be successfully corrected without access to the source text. The margin between the numbers of errors in the two groups of subjects was larger for Translation 1 (11.27 with ST and 14.54 without ST), which indicates that the source text was more useful in that case. This might be because Translation 1 was done into the subjects' second language, suggesting that the subjects who did not have access to the source text might have struggled with identifying the errors made by Google Translate. In the same direction of translation the subjects introduced additional errors in several categories, most likely because they were working without access to resources, and their level of second language competence is lower than their first language competence. In Translation 2, both groups were more successful and the margin between them was smaller (9.54 with ST and 11.27 without ST), while the percentage of errors they corrected was higher. This indicates that even the subjects who did not have access to the source text were able to deduce what the source text said and correct the translation accordingly. Also, there were almost no additional errors made by the subjects in Translation 2.

The results of this study also suggest something which was not in its focus. As it was mentioned above, both groups were considerably faster when they were post-editing the translation into the subjects' mother tongue. Also, both groups were more successful in correcting the errors made by Google Translate in that translation and the margins between them were smaller. The latter finding is in line with studies comparing translation quality in different directions (e.g. Pavlović 2007). This might indicate that post-editors should work in their mother tongue whenever possible, not only because they would be faster (more productive), but also to achieve greater quality. Further studies are also needed in this respect.

This experiment was not conducted with a text written in a controlled language, which might be another topic for further research. It would be interesting to see how big the differences would be between the two groups of subjects if they were post-editing texts in a controlled language. Furthermore, Google Translate is not a translation tool intended for professionals, and it is a statistical MT engine, so it might be useful to carry out a similar experiment with a different engine and see what the results would be. Other studies might use a different method or combination of methods, for instance screen-recording, keystroke logging or verbalizations to explore the post-editing process in more depth. All in all, there is still a significant amount of possible topics for research in the area of post-editing.

References

Allen, Jeffrey. 2003. "Post-editing". In: Somers, Harold (ed.), *Computers and Translation. A translator's guide.* Amsterdam/Philadelphia: John Benjamins. pp. 297-317.

Bellos, David. 2011. "How Google Translate Works". http://www.independent.co.uk/life-style/gadgets-and-tech/features/how-google-translate-works-2353594.html. Accessed on: 29 May 2014.

Doherty, Stephen and Gaspari, Federico. 2013. "Effective Post-Editing in Human & Machine Translation Workflows: Critical Knowledge & Techniques". http://www.qt21.eu/launchpad/sites/default/files/QTLaunchPad%20Post-Editing%20Webinar.pdf. Accessed on: 29 May 2014.

Dovedan, Zdravko, Seljan, Sanja and Vučković, Kristina. 2002. "Machine Translation as Help in the Communication Process - Strojno prevođenje kao pomoć u procesu komunikacije". *Informatologia* 35(4): pp. 283-91.

Duh, Kevin. 2005. "Example-based machine translation". http://ssli.ee.washington.edu/people/duh/projects/ebmt.pdf. Accessed on: 29 May 2014.

Entereurope. 2006. "Budućnost Europske unije". http://www.entereurope.hr/page.aspx?PageID=143. Accessed on: 29 May 2014.

Europa. 2014. "How EU decisions are made". http://europa.eu/eu-law/decision-making/procedures/index_en.htm. Accessed on: 29 May 2014.

European Patent Office. 2013. "Patent Translate". http://www.epo.org/searching/free/patent-translate.html. Accessed on: 29 May 2014.

Flournoy, Raymond and Duran, Christine. 2009. "Machine Translation and Document Localization at Adobe: From Pilot to Production". http://www.promt.ru/press/pdf/ADOBE_Flournoy2009.pdf. Accessed on: 29 May 2014.

Google, Inc. 2014. "Inside Google Translate". https://translate.google.com/about/intl/en_ALL/. Accessed on: 29 May 2014.

Hutchins, John. 1999. "The development and use of machine translation systems and computer-based translation tools". http://www.foreignword.com/technology/art/Hutchins/hutchins99.htm. Accessed on: 29 May 2014.

Hutchins, John. 2001. "Machine translation and human translation: in competition or in complementation?". *International Journal of Translation*, 13(1-2): pp. 5-20.

Krings, Hans and Koby, Geoffrey (ed.). 2001. Repairing Texts. Empirical Investigations of Machine Translation Post-Editing Processes. Kent: The Kent State University Press.

Kučiš, Vlasta, Seljan, Sanja and Klasnić, Ksenija. 2009. "Evaluation of Electronic Translation Tools through Quality Parameters". In: Stančić, Hrvoje, Seljan, Sanja, Bawden, David, Lasić-Lazić, Jadranka and Slavić, Aida (ed.), *INFuture2009: Digital Resources and Knowledge Sharing*. Zagreb: Odsjek za informacijske znanosti. pp. 341-51.

Luckhardt, Heinz-Dirk. 1991. "Sublanguages In Machine Translation". In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics. pp. 306-8.

Pavlović, Nataša. *Directionality in Collaborative Translation Processes*. Unpublished PhD dissertation. Tarragona: Universitat Rovira i Virgili. http://www.tdx.cat/bitstream/handle/10803/8770/THESIS.pdf?sequence=1. Accessed on: 29 May 2014.

Rico, Celia and Torrejón, Enrique. 2012. "Skills and Profile of the New Role of the Translator as MT Post-editor". *Revista tradumàtica* 10. http://revistes.uab.cat/tradumatica/article/view/18/pdf. Accessed on: 29 May 2014.

Seljan, Sanja. 2000. „Sublanguage in Machine Translation". In: Biljanović, Petar, Skala, Karolj, Ribarić, Slobodan and Budin, Leo (ed.), *Proceedings of 23rd International Convention MIPRO 2000: Computers in Intelligent Systems CIS + CTS*. Rijeka: Liniavera. pp. 17-20.

Seljan, Sanja, Brkić, Marija, and Kučiš, Vlasta. 2011. "Evaluation of Free Online Machine Translations for Croatian-English and English-Croatian Language Pairs". In: Billenness, Clive, Hemera, Annette, Mateljan, Vladimir, Banek Zorica, Mihaela, Stančić, Hrvoje and Seljan, Sanja (ed.), *INFuture2011: Information Sciences and e-Society*. Zagreb: Department of Information Sciences. pp. 331-44.

Systransoft. 2014. "What Is Machine Translation?" http://www.systransoft.com/systran/corporate-profile/translation-technology/what-is-machine-translation/. Accessed on: 29 May 2014.

TAUS: Translation Automation User Society. 2006. "Putting Machine Translation to Work". https://www.taus.net/downloads/finish/56-public-reports/16-taus-putting-machine-translation-to-work-april-2006. Accessed on: 29 May 2014.

Thicke, Lori, 2011. "Improving MT Results: A Study". http://lexworks.com/cms/wp-content/uploads/2012/07/Improving_MT_Results_Multilingual_01-11.pdf. Accessed on: 29 May 2014.

Torrejón, Enrique and Rico, Celia. 2002. "Controlled Translation: A New Teaching Scenario Tailor-Made for the Translation Industry". http://www.mt-archive.info/EAMT-2002-Torrejon.pdf. Accessed on: 29 May 2014.

Vasconcellos, Muriel and León, Marjorie. 1985. "SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization". In: *Computational Linguistics* 11, 122-136. http://www.aclweb.org/anthology/J/J85/J85-2003. Accessed on: 29 May 2014.

Vasconcellos, Muriel. 1987. "A comparison of MT postediting and traditional revision". In Newman, Patricia (ed.), *Proceedings of the 28th Annual Conference of the American Translators Association*. Medford: Learned Information. pp. 409-16.

Wikipedia: The Free Encyclopedia, 2014. "*Google Translate".* http://en.wikipedia.org/wiki/Google_Translate. Accessed on: 29 May 2014.