

## **Post-editing wildlife documentary films: A new possible scenario?** **Carla Ortiz-Boix and Anna Matamala, Universitat Autònoma de Barcelona**

### **ABSTRACT**

Several studies have proven that, when machine translation followed by post-editing is used to translate general and specialised texts, there is an increase in the productivity, as the post-editing effort is lower than translating *ex novo*. Although the use of machine translation and post-editing has been investigated in Audiovisual Translation, this has never been researched in non-fictional audiovisual genres in which voice-over and off-screen dubbing are applied. Using an English wildlife documentary film as the source text, and Spanish as the target language, this study intends to research whether post-editing involves more or less effort than translating a documentary. Conclusions on the experiment described in this article, in which 12 Audiovisual Translation MA students took part, seem to indicate that post-editing involves less effort than translating.

### **KEYWORDS**

Audiovisual translation, machine translation, post-editing, voice-over and off-screen dubbing.

### **1. Introduction**

In the last two decades, the use of Machine Translation (MT) followed by post-editing when applied to general and specialised translation has been expanding. Such growth has affected not only the market (TAUS, 2009), but also research on post-editing. However, the market of audiovisual translation has barely been affected. Research studies that intend to include MT and post-editing into the process of translating audiovisual products only started a few years ago thanks to European projects such as eTITLE (Melero *et al*, 2006) or, more recently, SUMAT (Del Pozo *et al*, 2013), both focusing on subtitling. The promising results presented by the latter led us to believe that applying MT and post-editing to other audiovisual translation modalities might be feasible and worth researching. This has been precisely the aim of the ALST project (Matamala *et al*, 2012): to investigate the possible application of MT and post-editing into two oral audiovisual transfer modes, namely audio description and voice-over.

The research presented in this article is part of the aforementioned ALST project (FFI-2012-31024), which is financed by the Spanish “Ministerio de Economía y Competitividad”, and focuses exclusively on wildlife documentary films which are translated by means of voice-over and off-screen dubbing. Voice-over is the revoicing of an audiovisual text in another language in which a translating voice is superimposed on the original voice (Franco *et al*, 2010). It is frequently used in non-fictional audiovisual genres, especially when speakers appear on-screen, but also in fictional TV programmes in Eastern Europe. On the other hand, off-screen dubbing generally refers to the audiovisual transfer mode used to

revoice off-screen narrations in which the original voice is substituted by a target language version (Franco *et al*, 2010). Wildlife documentary films have been selected because, according to a preliminary study by Ortiz-Boix (forthcoming) on a corpus of documentaries, many elements (such as the promising results of the analysed free online MT engines, and the types of errors these engines produce) seem to indicate that it would be feasible to apply MT to this specific genre. However, testing this new scenario in comparison with existing practices with users is yet to be carried out. This is precisely the aim of the research described in this paper: to compare the effort when post-editing a machine translated wildlife documentary and when translating it. Our hypothesis is that post-editing will require less effort than translating.

The article is structured as follows: Section 2 discusses the theoretical approach taken in this paper. In section 3, the methodology used is explained, describing in detail the experiments carried out in June 2014, as well as the methods used to analyse the data. Section 4 discusses the results, taking into account the different types of efforts analysed (temporal, technical, cognitive), and section 5 presents the conclusions and avenues for further research.

## **2. Theoretical approach: post-editing effort in audiovisual translation**

This section defines post-editing and how the effort involved has been measured in previous experiments. It also highlights the specificities of the audiovisual transfer modes under analysis.

Post-editing is the "term used for the correction of MT output by human linguists/editors" (Veale and Way, 1997, cited in O'Brien, 2010:1) and, therefore, "the task of the post-editor is to edit, modify and/or correct pre-translated text" (Allen, 2003:297). Post-editing can basically be carried out on two different levels: minimal or light, and full (Allen, 2003:304-306) and, depending on the level of post-editing used, the required effort will vary.

During the last decade, defining and measuring effort within post-editing research has been in the spotlight, thanks to works carried out by Krings (2001), O'Brien (2004, 2005 and 2006) or Martínez (2003), to name just a few. Krings (2001) led the way by determining how to calculate such effort and setting the standard for the majority of the other works on this topic. According to Krings (2001), post-editing effort can be divided into three types: temporal, technical and cognitive. Temporal effort is understood as the time taken to post-edit a document. Technical effort refers to the number of keystrokes, mouse movements and clicks. And cognitive effort applies to "the extent and type of cognitive processes that must be activated to remedy a deficiency in the MT output" (Krings 2001:179).

While temporal and technical efforts can be directly observed thanks to keylogging software, as can be seen in Allen (2001), Martínez (2003) or Tatsumi and Roturier (2010), cognitive effort cannot be directly observed. Hence, several methods have been used to observe it: Krings (2001) used Think-Aloud Protocols, although he later realised that verbalising all the movements slowed down the process. O'Brien (2004) observed cognitive effort using Translog, a keylogging software. Although Translog did not permit the direct observation of cognitive effort, it did succeed in measuring the number, location and duration of pauses, which were all considered good indicators of cognitive load (O'Brien 2006; Shreve *et al* 2011). Eye-tracking, a non-intrusive equipment that records eye movements and fixations, is another tool used to measure cognitive effort (O'Brien, 2011). To determine the cognitive load of post-editing effort, processing speed, average fixation time and count are generally taken into account. More recently, Lacruz *et al* (2014a; 2014b) have claimed that there are two formulae that correlate well with cognitive effort: average pause ratio (APR) and pause to word ratio (PWR). According to them, a low APR (the least possible amount of time spent pausing) combined with a high PWR (the most possible time spent pausing per word) are associated with high levels of cognitive effort. To allow for a lower level of applied cognitive effort, a combination of high APR and low PWR, would be beneficial. Both data can be obtained using keylogging software.

Although an increasing number of researchers study post-editing effort and compare it to translation to determine which one is more productive (Almeida and O'Brien 2010; Guerberof 2009), only a few have analysed post-editing effort as applied to audiovisual translation (de Sousa *et al* 2011; Läubli *et al* 2013), and specifically to subtitling. Other investigations linking audiovisual translation with post-editing have mostly focussed on the quality assessment of machine translated or post-edited subtitles (Armstrong *et al* 2006; Melero *et al* 2006; Volk, 2008; Del Pozo *et al* 2013 or Bywood *et al* 2013).

In order to apply MT and post-editing into the current audiovisual translation workflow, some specificities linked to the genre (wildlife documentary films) and audiovisual transfer modes under analysis (voice-over and off-screen dubbing) need to be taken into account. Voice-over is, together with off-screen dubbing, a modality generally used to translate non-fictional genres in Western Europe (Franco *et al* 2010). Among these non-fictional genres, one can find wildlife documentaries, which form the focus of this research. The main characteristics of documentaries are the presence of both a narrator with a generally planned discourse and experts who tend to use a more spontaneous language (Matamala 2009). Narrators are usually off-screen and dubbed in the target language version, meaning the original narrator cannot be heard and is substituted by a translating voice, whilst on-screen speakers are voiced-over, meaning the translating voice is heard on top of the original, whose sound is lowered down. In both modalities there are synchronisation

requirements: translations must take into account the movements and actions on screen (action and kinetic synchronies), and the length of the utterance (isochrony) (Orero 2006). As far as working conditions are concerned, translators sometimes work without a script or with a script riddled with errors due to the possible lack of post-production scripts (Franco *et al*, 2010). All these features may be additional challenges when implementing MT in this specific field, as pointed out in a preliminary study by Ortiz-Boix (forthcoming), which suggested pre-editing, as a necessary step for a more successful implementation of MT. Pre-editing (Pym 1990) is understood as the revision of the format and content of a text before machine translating it. This allows for a higher quality MT output.

### **3. The experiment: methodological aspects**

As stated above, the aim of this experiment was to compare the effort involved in translating and post-editing wildlife documentaries. Following the theoretical approach in section 2, effort was measured in terms of temporal (seconds spent to perform the task), technical (keyboard and mouse usage) and cognitive features (pauses). It was therefore decided that data would be gathered using keylogging software.

#### **3.1. Participants**

12 Master students specialising in audiovisual translation participated in this study. They had all taken a specific course on voice-over, in which they were taught to translate documentaries. Tests were carried out in June, when all participants had successfully finished their courses and were working on their MA thesis. Half of the participants were males and the other half were females, ages ranged between 22 and 27 years old, and all of them had completed a BA in Translation and Interpreting. They had minimal or no previous experience as professional audiovisual translators and no experience as post-editors. All participants had Spanish as their first language and were highly proficient in English language.

#### **3.2. Materials**

Two excerpts of the 7-minute wildlife documentary *Must Watch: A Lioness Adopts a Baby Antelope* were used. They are available on Youtube as an independent documentary (<https://www.youtube.com/watch?v=mZw-1BfHFKM>) although it is part of the episode *Odd Couples* from the series *Unlikely Animal Friends* by National Geographic (2009). Both excerpts are comparable in terms of length and content, as shown in Table 1.

		FIRST EXCERPT	SECOND EXCERPT
DURATION		1:41 minutes (101 seconds)	1:52 minutes (112 seconds)
WORDS	TOTAL	283	287
	NARRATOR	50	58
	EXPERTS	222	229
INTERVENTIONS	TOTAL	8	9
	NARRATOR	3	3
	EXPERTS	5	6

**Table 1. Comparison of excerpts**

Both excerpts were machine translated from English into Spanish by *Google Translate* as, according to previous research by Ortiz-Boix (forthcoming), this is the best free online MT engine to translate wildlife documentary films in this language pair. Automatic measures were calculated with the translations and the post-editings produced by the participants (see Table 2 in 5.3.): BLEUs (Papineni 2002), h-BLEU<sup>1</sup>s (Snover *et al* 2006:224), TERs (Snover *et al* 2006) and h-TER<sup>2</sup>s (Snover *et al* 2006:224).

### 3.3. Data gathering tools

Inputlog (Leijten *et al* 2013), a research tool for logging and analysing writing processes developed at the University of Antwerp, was used to record the data. The following measures were obtained: total time, time spent while performing the task and while searching, keylogging, number of mouse movements and clicks, pause thresholds, type of visited internet webpages and type of used software. Although other post-editing tools were considered, they were discarded because they did not integrate audiovisuals (Ortiz-Boix, *forthcoming*). Inputlog was prioritised over other keylogging software because it allowed for a better simulation of the current workflow of audiovisual translators. It also means that audiovisual materials could be watched without interfering with the tool.

### 3.4. Test development

Participants volunteered to take part in the experiment, which was carried out in a lab environment simulating real-life working conditions. They were instructed about the nature of the experiment and signed informed

consent forms, following the procedures approved by the Ethical Committee at Universitat Autònoma de Barcelona (UAB). They were instructed that the experiment would develop as follows: they would have to translate an excerpt of a wildlife documentary, and post-edit the machine translated output of another excerpt. They were required to use a Microsoft Word template for both tasks, as this was the software used in the MA course they had all taken, but they were free to use any resources available to them online (search engines, video software, etc.). The specific instructions that were given to them were to translate or post-edit, being aware that they had to produce a final document ready to be recorded at a sound studio. They were required to include timecodes in (not out), and they were provided with pre-established timecodes which they could modify if necessary. In the specific case of post-editing, they were instructed to post-edit only when there was a semantic or grammatical error, when some information was omitted or added, and when there were spelling and punctuation mistakes. They were told not to post-edit merely stylistic problems but were asked to rephrase the sentences if, despite being correct, they did not meet the standard conventions of voice-over and off-screen dubbing (this refers to synchronisation features and presentation layouts). After finishing the tasks, they were given a questionnaire on subjective data, the analysis for which is beyond the scope of this paper. Participants were randomly assigned to four different groups in which the two conditions (post-editing/translation) and excerpts (1 and 2) were randomised to avoid any bias regarding the order of presentation.

### **3.5. Data and methods**

20 valid Inputlog files were collected due to technical problems with four files. Data was obtained from the General Analysis Documents file and exported into Microsoft Excel files. They were analysed using the statistical system R-3.1.2, developed at Bell Laboratories by John Chambers and colleagues.

The following data was obtained for all excerpts and tasks:

- a) Analysis of temporal effort: average time spent translating and post-editing, average time spent while working on the Word document, on search engines and using video software.
- b) Analysis of technical effort: average number of keyboard and mouse usage, average number of mouse movements and scrolls, average number of mouse clicks and average number of keystrokes. Average number of mouse movements and scrolls, mouse clicks, and keystrokes while working on the Word document, on search engines and on video software were also analysed.
- c) Analysis of cognitive effort: average number of pauses and average number of pauses while working on the Word document. To

determine PWR and APR, the number of words of each final document and the average time per pause were also assessed.

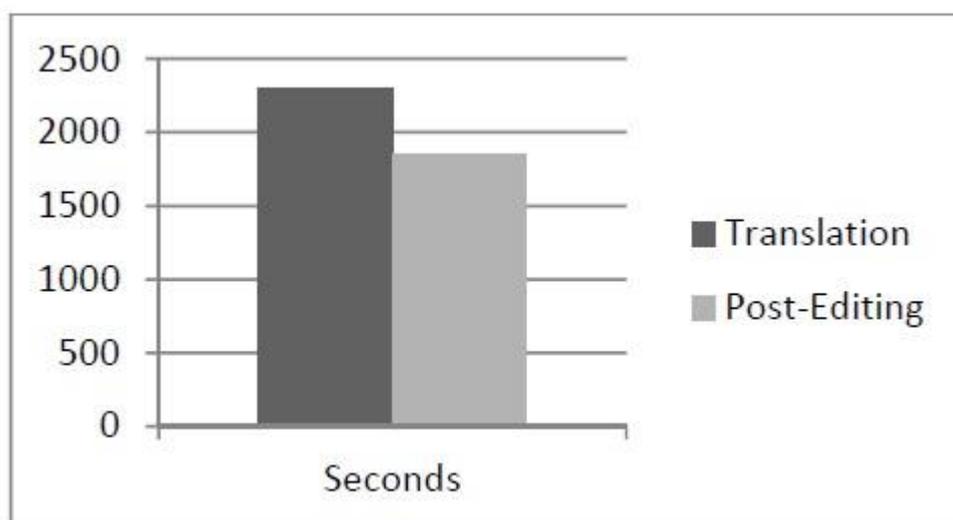
An ANOVA variance test was used to determine the significance of the results. According to the test, the null-hypothesis can be rejected when the probability value (p-value) is equal or lower than 0.05 ( $p < 0.05$ ). The general null-hypothesis of this research states that "there is a significant difference between post-editing effort and translating effort when working with wildlife documentary films scripts."

#### **4. Results**

The global analysis indicates that the post-editing effort is significantly lower than the translating effort in the case of technical effort ( $F=4.417$ ,  $p=0.050$ ) and cognitive effort ( $F=5.979$ ,  $p=0.025$ ). However, temporal effort is not ( $F=1.297$ ;  $p=0.270$ ). This may be due to the time one participant spent post-editing, as he spent nearly double the time the others did. When this participant is not taken into account, the post-editing temporal effort is also lower than the translation temporal effort ( $F=6.756$ ,  $p=0.019$ ). Although these results validate our hypothesis, when data from the two different excerpts are analysed in more detail, it can be observed that the difference between post-editing effort and translation effort is not always significant. In the following subsections, and according to the three types of effort identified above, an in-depth analysis is presented.

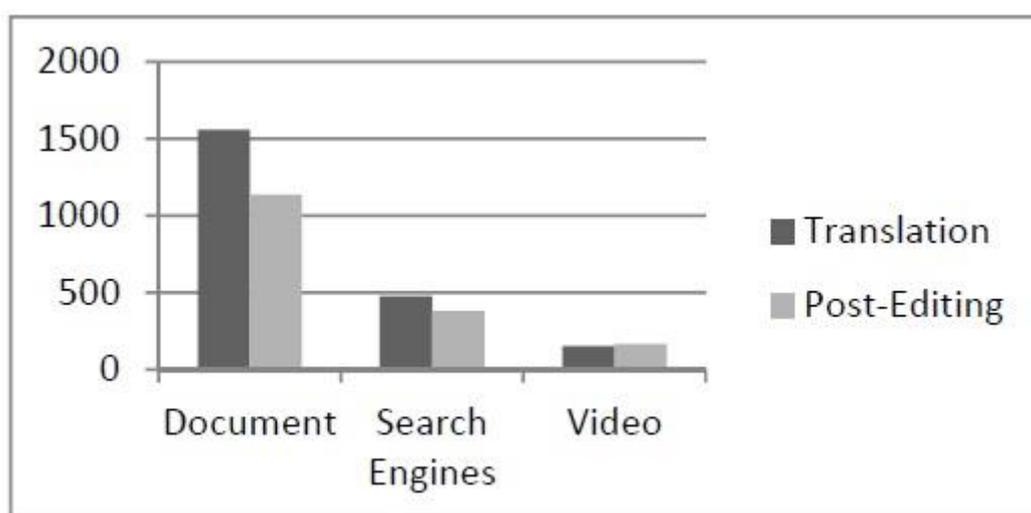
##### **4.1. Temporal effort**

The analysis of temporal effort indicates that, in the first excerpt, participants spent less time post-editing than translating (see Figure 1): the average time spent translating was 2301.833 seconds (38.36 minutes) and 1853.8 seconds (30.9 minutes) for post-editing. The difference between both tasks being 448.033 seconds (7.47 minutes). ANOVA significance test shows that the temporal effort is significantly lower when post-editing ( $F=12.940$ ;  $p=0.006$ ), confirming the results of the general analysis.



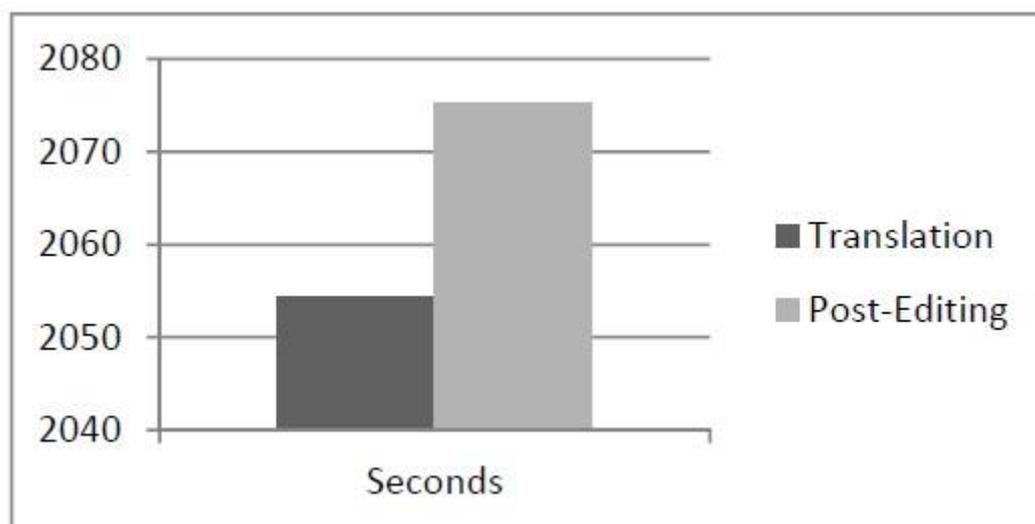
**Figure 1. Comparison of Temporal Effort. Excerpt 1.**

If the timings are explored in more detail, it can be observed (see Figure 2) that, from all the time dedicated to the performance of the translation, participants spent, in excerpt 1, an average of 1556.1438 seconds (25.94 minutes) on the document (67.605% of the time), 477.0633 seconds (7.95 minutes) on search engines (20.725% of the time) and 152.3562 seconds (2.54 minutes) using the video software (6.619% of the time). When post-editing, the difference between the time performing the task on the document (1137.7662 seconds (18.96 minutes), 61.375% of the time) and on the Internet (378.4386 seconds (6.31 minutes), 20.414% of the time) is smaller. Furthermore, post-editors spent more time using video software (165.263 seconds (2.75 minutes), 8.915% of the time). According to the results, there is evidence leading to the belief that post-editors and translators devote approximately the same time to research ( $F=1.345$ ;  $p=0.276$ ) and to the video ( $F=0.034$ ;  $p=0.612$ ). However, the time spent on each task within the document is significantly different ( $F=9.918$ ;  $p=0.012$ ).



**Figure 2. Division of Temporal Effort. Excerpt 1.**

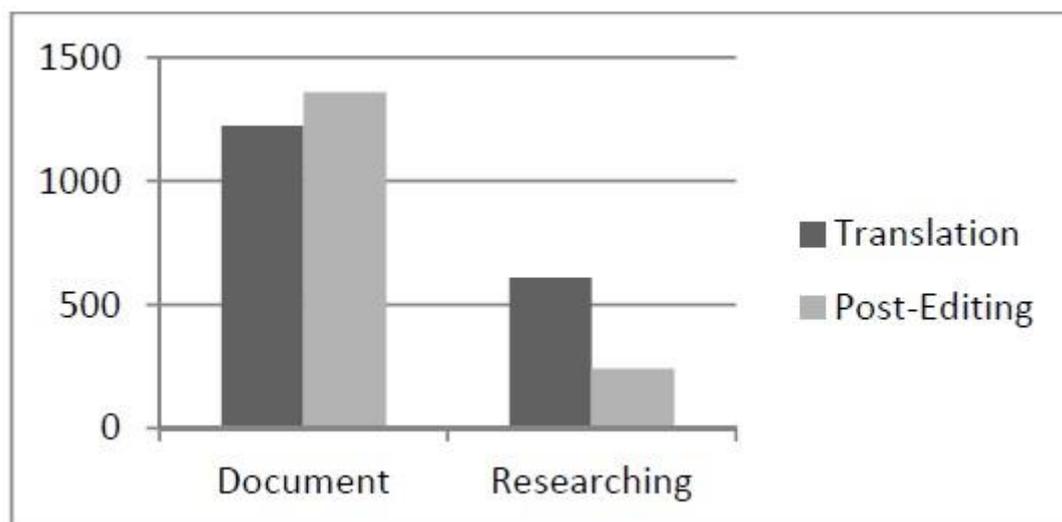
In the second excerpt, however, the results of the general analysis are not ratified. In this case, the differences between both tasks are minimal (see Figure 3) and the tendency of greater temporal effort when translating does not continue. The average time for translating is 2054.4 seconds (34.24 minutes) and, for post-editing, 2075.25 (34.59 minutes). This means that it took 20.85 more seconds to post-edit this excerpt. Such a change of tendency, as indicated above, is due to the amount of time one of the participants spent post-editing the excerpt. If this participant is considered an outlier and his data is not taken into account for the analysis, the differences are more similar to those of the first excerpt (see Figure 4): 2,054.4 seconds translating (34.24 minutes) and 1,674.6667 seconds post-editing (27.91 minutes), reversing the difference to 379.7333 seconds in favour of post-editing. In this case, ANOVA significance test ( $F= 0.002$ ;  $p=0.965$ ) shows that the difference between post-editing and translation in terms of time is not significant. The difference is closer to be significant when the participant who doubled the time is not included in the data ( $F= 1.265$ ;  $p=0.304$ ). As this participant's behaviour differed considerably from the others, this participant's results were excluded in the analysis of all the other parameters, which are presented below.



**Figure 3. Comparison of Temporal Effort. Excerpt 2.**

When the temporal effort for the second excerpt is divided into time spent performing the task within the document, on the search engines or on the audiovisual display, the results are slightly different from the ones obtained in excerpt 1 (see Figure 4). Post-editors spent more time working on the document (1357.577 seconds (22.63 minutes), 81.066% of the time) than translators (1222.78696 seconds (20.38 minutes), 59.520% of the time). Post-editors, however, spent less time on the Internet and using the video software (118.9193 seconds (1.98 minutes), 7.101% of the time, and 122.8303 seconds (2.05 minutes), 7.335% of the time, respectively). Translators spent 328.2352 seconds (5.47

minutes, 15.977% of the time) on search engines and 280.6964 seconds (4.68 minutes, 13.663% of the time) on the audiovisual display. The ANOVA significance test shows that there is no significant difference between translation and post-editing in either the Word document ( $F=0.355$ ;  $p=0.573$ ), the search process ( $F=3.480$ ;  $p=0.111$ ) or when working with the audiovisuals ( $F=0.562$ ;  $p=0.482$ ).

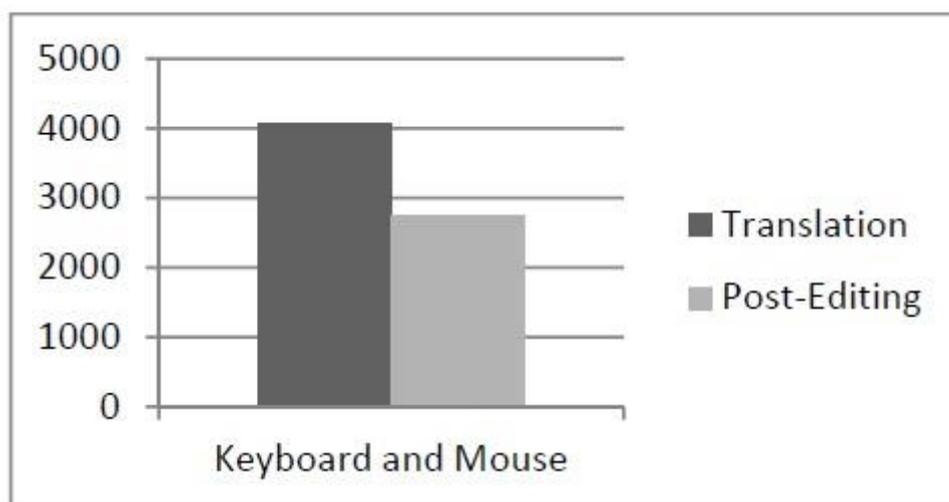


**Figure 4. Division of Temporal Effort. Excerpt 2**

To sum up, although the general analysis indicates that the post-editing temporal effort is lower than the translation temporal effort, a separate analysis of the two excerpts shows inconsistencies. While in the first excerpt the temporal effort is greater in translation than in post-editing, in the second excerpt there are no significant differences between post-editing and translating in terms of temporal effort. In both, no difference can be seen when considering the time spent when performing the task on the document. However, there is also no significant difference in any of the excerpts when considering the time spent both researching and working with the video.

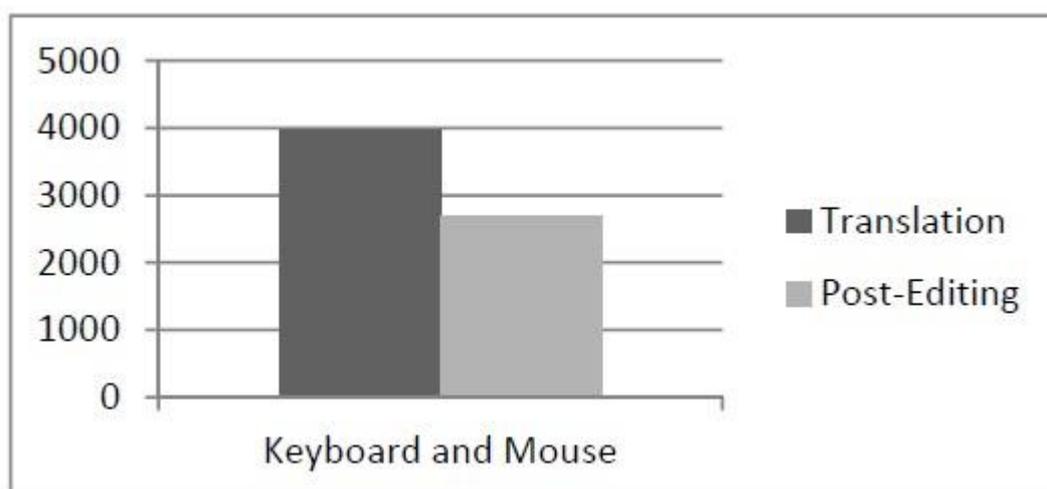
#### **4.2. Technical effort**

The analysis shows that technical effort is higher when translating in both excerpts (see Figures 5 and 6). Translators used the keyboard and the mouse an average of 4079.167 times for the first excerpt and 3972.4 for the second, whilst post-editors used them an average of 2733.8 times for the first excerpt and 2679.333, for the second.



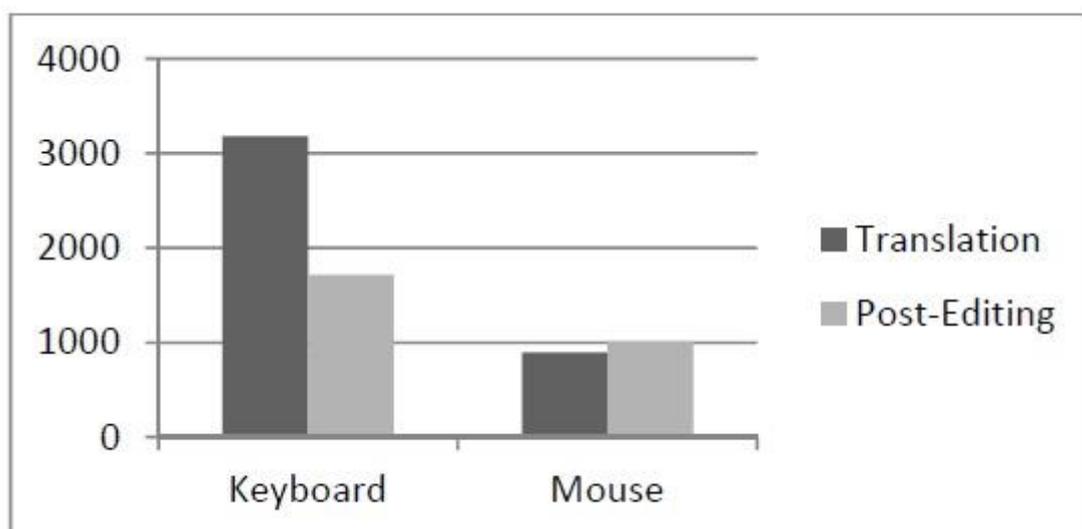
**Figure 5. Comparison of Technical Effort. Excerpt 1**

In the case of the first excerpt, the difference between the use of technical features when translating and post-editing is of 1345.367 keystrokes and mouse movements and clicks (see Figure 5). For the second excerpt, the difference is a little bit lower (see Figure 6): 1293.067.



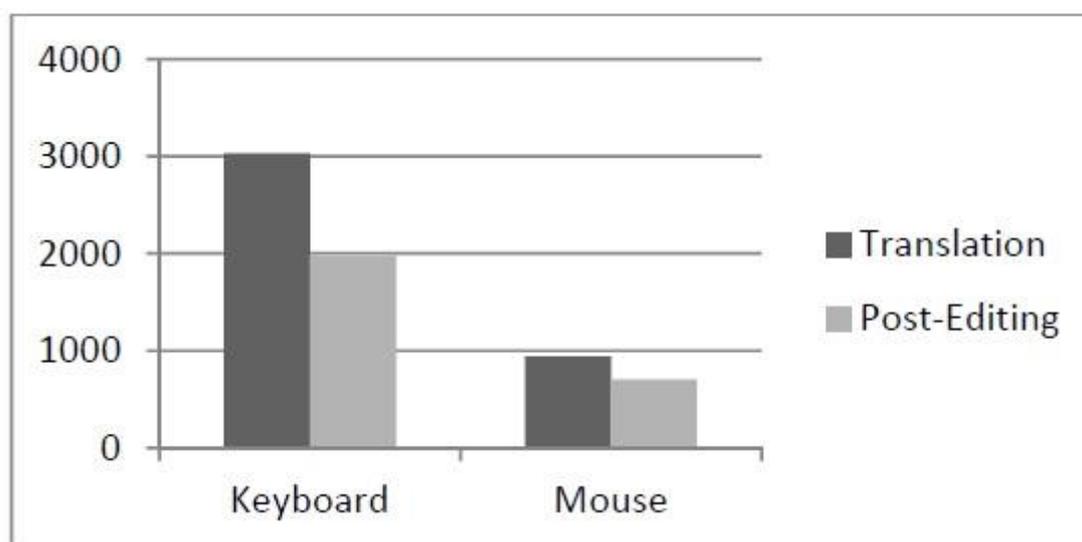
**Figure 6. Comparison of Technical Effort. Excerpt 2**

According to the results there is evidence to suggest that technical effort is higher when translating than when post-editing. However, the difference is only statistically significant in the first excerpt ( $F=6.365$ ,  $p=0.033$ ; excerpt 2:  $F=3.529$ ,  $p=0.109$ ). When technical effort is divided into keyboard strokes and mouse usage, these results show that the difference between post-editing and translating technical efforts is due to keyboard use ( $F=9.943$ ,  $p=0.012$ ). While the participants who translated the first excerpt used the keyboard an average of 3183 times and the mouse 896.167 times, the ones who post-edited the same excerpt only used the keyboard 1719 times but moved or clicked the mouse more: 1014.8 times (see Figure 7).



**Figure 7. Division of Technical Effort 1. Excerpt 1**

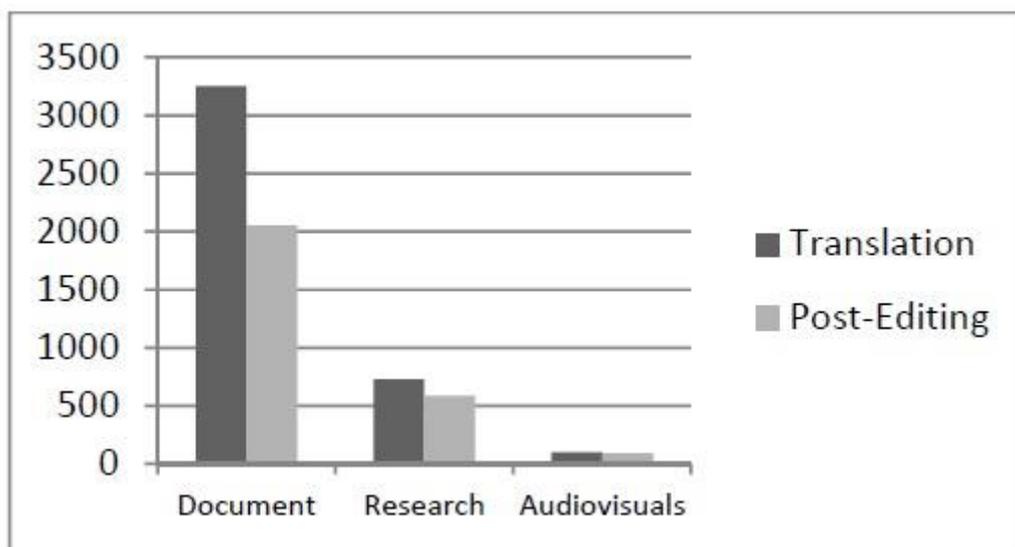
The tendency to use the mouse more in post-editing is not followed in the second excerpt (Figure 8). Instead, the participants who translated the second excerpt did so. Translators used the keyboard 3029.2 times and the mouse 943.2 times on average; post-editors made an average of 1974.334 keystrokes and 705 mouse clicks or movements (see Figure 8). Despite the translators making 1,000 keystrokes more than the post-editors, the difference in this case is not significant ( $F = 4.644$ ,  $p = 0.075$ ).



**Figure 8. Division of Technical Effort 1. Excerpt 2**

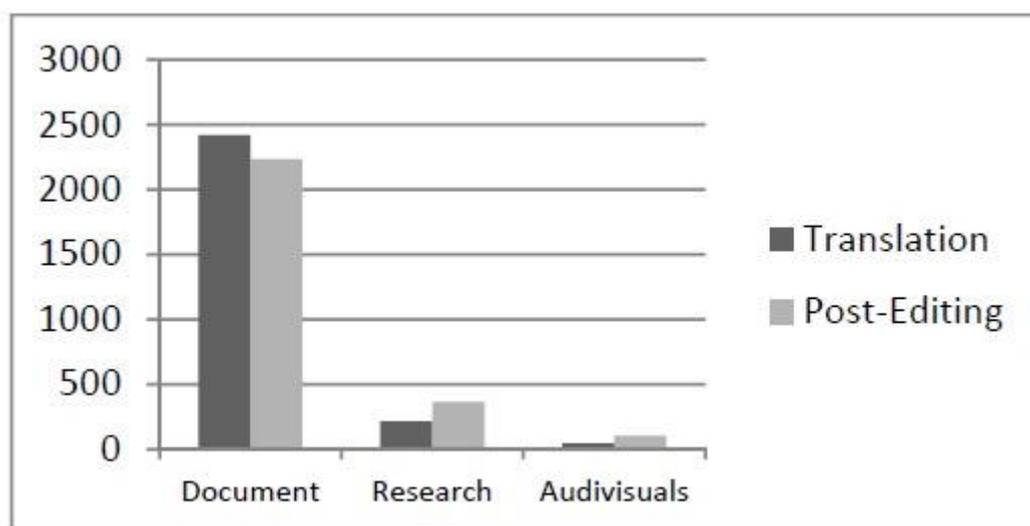
When analysing the technical effort distribution in the main document, the search engine and the audiovisual display, one can observe that 79.779% of the technical effort (3254.333 keystrokes and mouse movements and clicks) made by the translators of the first excerpt is concentrated on the main document, 17.802% (726.167 keystrokes and mouse movements and clicks) on search engines and only 2.419% of the effort (98.667 keystrokes and mouse movements and clicks) while using the video

software. The post-editors who dealt with the same excerpt dedicated almost the same effort to the audiovisual display (3.382%, 92.4 keystrokes and mouse movements and clicks). Their effort on the main document, 4.679 points lower than the translators' (2051.6 keystrokes and mouse movements and clicks), affected the technical effort while searching on the Internet, which reached 21.517% (587.8 keystrokes and mouse movements and clicks). According to these results, it can be stated that a great majority of the technical effort is concentrated in the main document regardless of the task (see Figure 9).



**Figure 9. Division of Technical Effort 2. Excerpt 1**

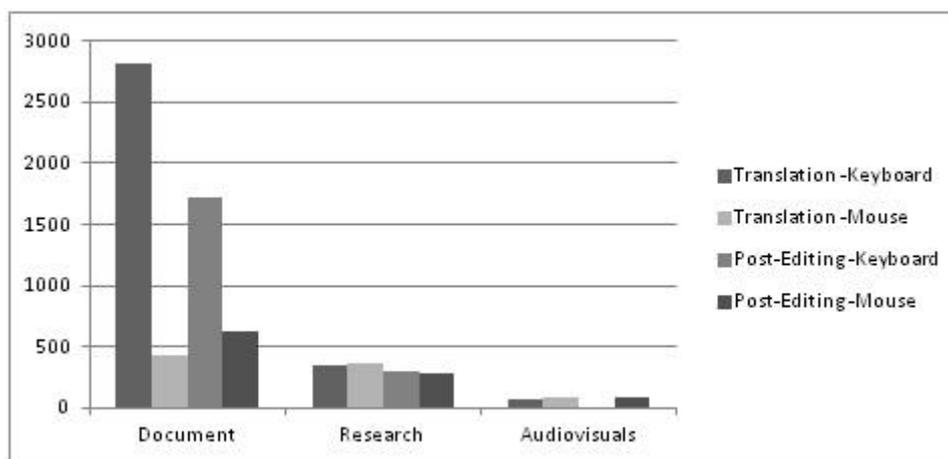
The results of the second excerpt follow a similar pattern; technical effort is more concentrated in the document and therefore less technical effort is required where research and audiovisual effort is concerned (see Figure 10): when translating, 81.432% of the technical effort (2420.333 keystrokes and mouse movements and clicks) is concentrated in the main document, while 15.935% (214.667 keystrokes and mouse movements and clicks) is dedicated to the search engines and 2.633% (44.333 keystrokes and mouse movements and clicks) to the audiovisual display. In the case of post-editing, 90.333% of the effort (2234.8 keystrokes and mouse movements and clicks) is made on the document, 8.012% (363 keystrokes and mouse movements and clicks) on the Internet and 1.655% (104.6 keystrokes and mouse movements and clicks) while using the video software.



**Figure 10. Division of Technical Effort 2. Excerpt 2**

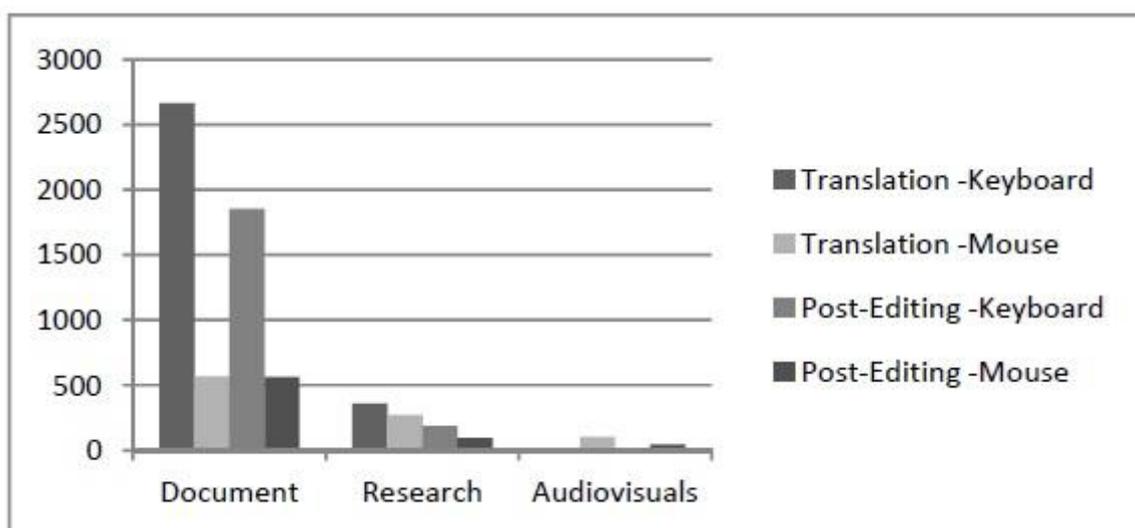
Apart from showing that technical effort is basically focused on the main document, the in-depth analysis also shows that when translating and post-editing, the use of the keyboard or the mouse varies: keyboard usage is more intensive when working on the document, while it is almost non-existent when working with the video. When doing online searches, the difference between using the keyboard or the mouse is minimal.

When working within the document, the participants who translated the first excerpt (see Figure 11) used the keyboard an average of 2819.334 times (86.633%) and the mouse, 435 times (13.367%). Translators made an average of 355.833 keystrokes (49.002%) and 370.333 mouse movements and clicks (50.998%) while searching on the Internet; and 78.33 keystrokes (7.939%) and 90.833 mouse clicks and movements (92.061%) while using the video software. The ones who post-edited the same excerpt (see Figure 11) made fewer keystrokes (1419 keystrokes, 69.166%) and used the mouse more extensively (632.6 mouse movements and clicks, 30.834%) while working within the document. In the case of using the search engines and the video software, the difference compared with the results of the translators is minimal. They made an average of 294.6 keystrokes (50.119%) and 293.2 mouse movements and clicks (49.881%), and an average of 3.4 keystrokes (3.679%) and 90.833 mouse clicks and movements (92.061%), respectively.



**Figure 11. Division of Technical Effort 3. Excerpt 1**

Regarding the second excerpt (see Figure 12), the results indicate that the trend continues in the case of working within the document and the video software, but the difference between post-editing and translating with regards to technical efforts while searching on the Internet is a bit higher. On the one hand, the translators used the keyboard an average of 2665.8 times (82.410%) and the mouse 569 times (17.590%), when working within the document. In the case of using search engines, they did 361.2 keystrokes (57.062%) and 271.8 mouse movements and clicks (42.938%). Regarding the technical effort while using the audiovisual display, they used the keyboard an average of 2.2 times (2.103%) and the mouse, 102.4 (97.897%). On the other hand, post-editors made 1,855.333 keystrokes (76.656%) and 565 mouse movements and clicks (23.344%) on the document; and used the keyboard 188.667 times (55.279%) and the mouse 96 times (44.721%) on search engines. In the case of the video software, post-editors used the keyboard an average of 0.334 times (0.752%) and the mouse 44 times (99.248%).



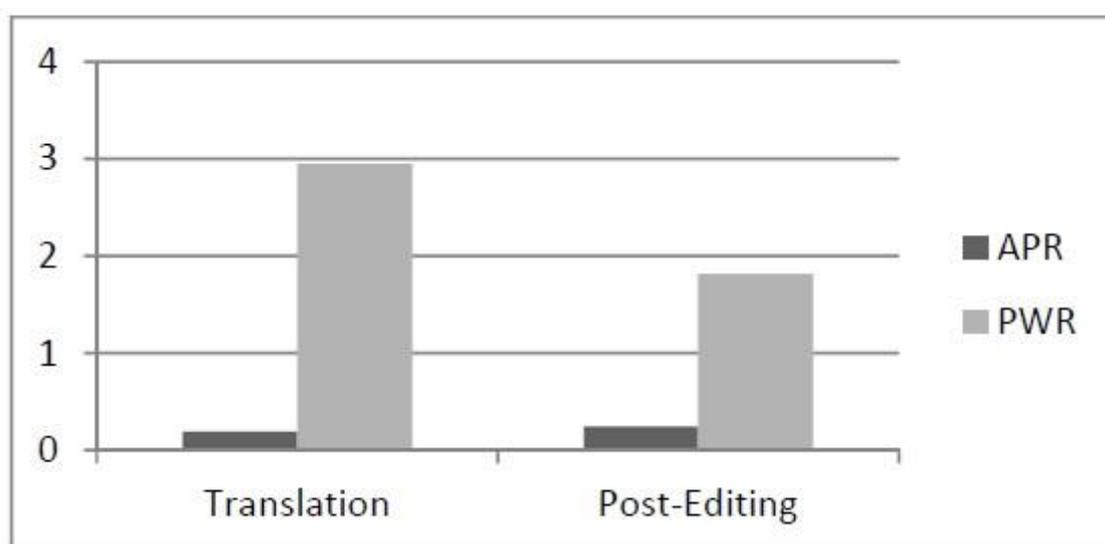
**Figure 12. Division of Technical Effort 3. Excerpt 2**

To summarise, as in the temporal effort, only the first excerpt follows the trend set by the general analysis, which includes both excerpts. The results show that the improvement of the technical effort is due to the decrease in keyboard usage, which is significantly lower only for the first excerpt. Most of the technical effort is concentrated in the main document, where keyboard usage is more intensive.

### 4.3. Cognitive Effort

Cognitive effort was assessed using the Lacruz *et al* (2014a) proposal, which states that the higher the difference between APR and PWR, the more cognitive effort is involved. In order to calculate the APR and the PWR for each task and excerpt, two measures gathered by Inputlog were used: total number of pauses and number of pauses while working on the document.

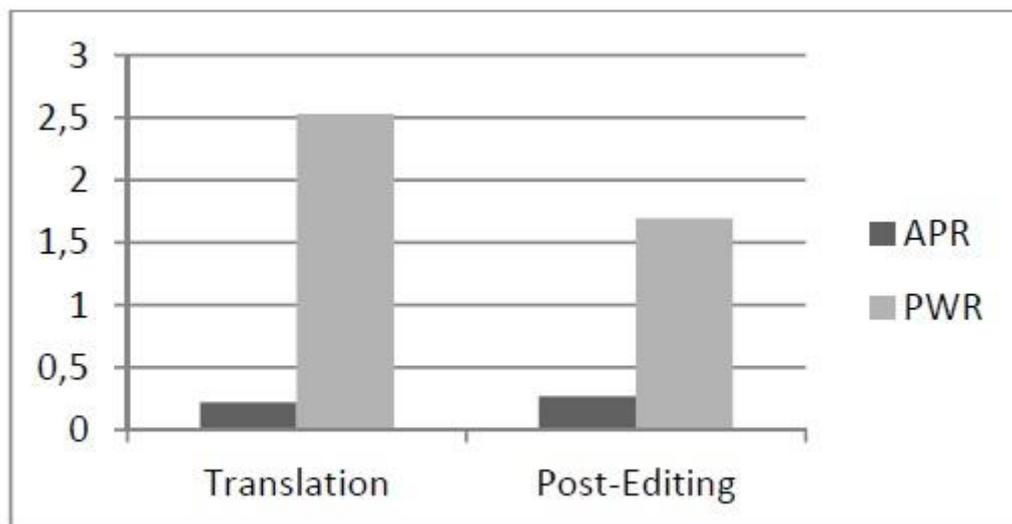
The results obtained for the first excerpt (see Figure 13) showed that the average APR is 0.191301 in the case of translation and 0.244064 for post-editing. The PWR of the same excerpt is 2.947685 for translation and 1.827491 for post-editing. As discussed in section 2, the lower the APR and the higher the PWR, the more cognitive effort is required during the task. Thus, the bigger the difference between APR and PWR, the greater the cognitive effort. The difference between APR and PWR, aka cognitive effort, is significantly higher when translating<sup>3</sup> (total: 2.756384; only document: 2.123383) than when post-editing (total: 1.583427; only document: 1.134013) if the total number of pauses are taken into account ( $F=11.959$ ;  $p=0.007$ ) or if only the pauses within the document are considered ( $F=11.332$ ,  $p=0.008$ ).



**Figure 13. Comparison of Cognitive Effort. Excerpt 1**

In the case of the second excerpt (see Figure 14), however, the difference between the translation cognitive effort (total: 1.261884; only document: 1.891389) and the post-editing cognitive effort (total: 1.920086; only

document: 2.310353) is not significant even when the total number of pauses are taken into account ( $F=2.712$ ,  $p=0.151$ ), or when only the pauses while working within the document are chosen ( $F=4.155$ ,  $p=0.088$ ).



**Figure 14. Comparison of Cognitive Effort. Excerpt 2**

To sum up, the translation cognitive effort is only significant in the case of the first excerpt. However, although the results of the second excerpt are not significant, the translation cognitive effort is also higher.

#### **4.4. Discussion of results**

The results generally confirm the hypothesis that the post-editing effort is lower than the translation effort. Both the general analysis and the analysis of the first excerpt validate the hypothesis, as the temporal, the technical and the cognitive efforts are significantly lower where post-editing is concerned. Nevertheless, the analysis of the second excerpt presents non-significant results. This was unexpected since a previous analysis was carried out to find two comparable excerpts. However, the non-significant results for the second excerpt might be due to three factors:

(1) Features of chosen documentary: although comparable in terms of number of words and interventions, the excerpts were not terminologically and syntactically identical. Furthermore, the MT of the second excerpt was worst, as indicated by the BLEU and TER scores presented (see Table 2).

	FIRST EXCERPT	SECOND EXCERPT
BLEU	44.97	33.75
H-BLEU	51.18	39.5
TER	69.17	59.68
H-TER	74.46	65.34

**Table 2. Automatic Measures**

(2) Technical skills of the participants: although all participants had the same training background and were assigned randomly to one of the groups, the analysis shows that the participants who post-edited the second excerpt were probably less skilled with the keyboard than the participants who translated it. This caused an increase in the amount of mouse usage and an increase on the time spent post-editing. Furthermore, the difference was high enough to presume that this may be the main reason why non-significant differences were observed.

(3) Amount of data: the limited number of participants may have had an impact on the significance tests. Therefore, we decided to simulate a situation in which the number of participants who post-edited was hypothetically duplicated. When doubling the number of participants, results are statistically significant only for cognitive effort ( $F=7.968$ ,  $p=0.011$ ). Temporal ( $F=1.249$ ,  $p=0.296$ ) and technical ( $F=4.207$ ,  $p=0.74$ ) efforts, although improving their results in the ANOVA significance test, are still not significant.

## 5. Conclusions and further research

Departing from previous research on post-editing effort, this study built upon the hypothesis that the post-editing effort is lower than the translating effort when working with wildlife documentary films. Global results proved the null-hypothesis of the study. However, results for the second excerpt do not. The excerpt specificities, the uneven technical skills of the participants, and the low number of participants may account for the diverging results.

The data analysis has taken into account the three types of effort specified by Krings (2001), and the following results have been obtained:

(1) Temporal effort: the global analysis shows that post-editing is faster. However, results are only statistically significant in the first excerpt.

(2) Technical effort: post-editing requires globally less keyboard and mouse usage. Again, the differences are statistically different in the first excerpt but not in the second one.

(3) Cognitive effort: post-editing has been proven to be less cognitively demanding although results are not statistically significant in the second excerpt.

Our data also suggests that the effort is concentrated in the main document and it is precisely there where the effort is reduced. In fact, the effort devoted to the search engines or to the audiovisual display does not vary significantly from one task to the other.

In conclusion, the results seem to indicate that it may be possible to use MT followed by post-editing in specific audiovisual genres such as wildlife documentaries which are voiced-over. However, further research should be carried out to confirm the trends shown in this study, which is limited in scope because it only focuses on one language pair (English into Spanish) and has included a small number of participants. Future research could encompass other types of text and include additional language pairs, with their own specificities. It could also take into account other relevant elements such as the subjective opinions and perceived effort of participants. Other aspects worth researching would be the output quality and audience acceptance of post-edited content in comparison with translated products, along with investigations carried out in other translation modalities (Fiederer *et al* 2009). It would also be highly relevant to measure the professional performance efforts of audiovisual translators. All in all, there are many aspects to be researched but this article has hopefully been a first step towards future studies on the implementation of translation technologies in the field of audiovisual translation and media accessibility, an area that is still under-researched especially when oral modalities such as voice-over, dubbing or even audio description are concerned.

## Bibliography

- **Allen, Jeff** (2001). "Postediting: an integrated part of a translation software program." *Language International* 13(2), 26-29.
- — (2003). "Post-editing." Harold Somers (ed.) *Computers and Translation: A translator's guide*. Benjamins Translation Library 35, 297-318.
- **De Almeida, Gisela and Sharon O'Brien** (2010). "Analysing post-editing performance: corrections with years of translation experience." *Proceedings of the 14<sup>th</sup> annual conference of the European association for machine translation, St. Raphaël, France*.
- **Armstrong, Stephen, Colm Caffrey, Marian Flanagan, Minako O'Hagan, Dorothy Kenny and Andy Way** (2006). "Improving the Quality of DVD Subtitles via Example-Based Machine Translation." *Proceedings of the Translating and the Computer 28 Conference*. London, England.
- **Bywood, Lindsay; Martin Volk; Mark Fisheland Panayota Georgakopoulou** (2013). "Parallel subtitle corpora and their applications in machine translation and translatology." *Perspectives: Studies in Translatology* 21(4), 595-610.
- **Del Pozo, Arantza, Gerard van Loenhout, Anthony Walker, Panayota Georgakopoulou and Thierry Etchegoyhen** (2013). *SUMAT: An Online Service for Subtitling by Machine Translation. Annual Public Report*.
- **Fiederer, Rebecca and Sharon O'Brien** (2009). "Quality and machine translation: A realistic objective." *JoSTrans, The Journal of Specialised Translation* 11, 52-74.
- **Franco, Eliana, Anna Matamala and Pilar Orero** (2010). *Voice-over translation: An overview*. Bern: Peter Lang.
- **Guerberof, Ana** (2009). "Productivity and quality in MT post-editing." *MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT*.
- **Krings, Hans P.** (2001). *Repairing texts: empirical investigations of machine translation post-editing processes* (Vol. 5). Kent: Kent State University Press.
- **Läbuli, Samuel, Martin Fishel, Martin Volk and Manuela Weibel** (2013). "Combining Statistical Machine Translation and Translation Memories with Domain Adaptation." *Proceedings of the 19<sup>th</sup> Nordic Conference of Computational Linguistics (NODALIDA 2013)*, May 22-24, 2013, Oslo University, Norway. NEALT Proceedings Series 16.
- **Lacruz, Isabel, Michael Denkowski and Alon Lavie** (2014a). "Cognitive Demand and Cognitive Effort in Post-Editing." *Third Workshop on Post-Editing Technology and Practice*.
- — (2014b). "Real Time Adaptive Machine Translation for Post-Editing with cdec and TransCenter." *EACL 2014*.
- **Leijten, Marielle and Luuk Van Waes** (2013). "Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes." *Written Communication* 30(3), 358-392.
- **Martínez, Lorena G.** (2003). *Human translation versus machine translation and full post-editing of raw machine translation output*. MA Diss. Dublin City University,

<http://sceuromix.com/enlaces/MASTER%20IN%20TRANSLATION%20STUDIES%20BY%20LORENA%20GUERRA-2003.pdf> (consulted 10.05.2016).

- **Matamala, Anna** (2009). "Main Challenges in the Translation of Documentaries." In Jorge Díaz Cintas (ed.). *New Trends in Audiovisual Translation*. Bristol: Multilingual Matters, 109-120.
- **Matamala, Anna, Anna Fernández-Torné and Carla Ortiz-Boix** (2012). "Technology and AD: The TECNACC Project." *Languages and the Media 2012*, Berlin. < [http://ddd.uab.cat/pub/presentacions/2012/117159/fernandez\\_matamala\\_ortiz\\_berlin\\_2012.pdf](http://ddd.uab.cat/pub/presentacions/2012/117159/fernandez_matamala_ortiz_berlin_2012.pdf) > (last accessed: 16th May 2016)
- **Melero, Maite and Antoni Oliverand Toni Badia** (2006). "Automatic multilingual subtitling in the eTITLE project." *Proc. of the 28<sup>th</sup> International Conference on Translating and the Computer*, 28 16-17 November 2006 in London. London: ASLIB.
- **O'Brien, Sharon** (2004). "Machine translatability and post-editing effort: How do they relate." *Proc. of the 26<sup>th</sup> International Conference on Translating and the Computer*, 18-19 November 2004 in London. London: ASLIB.
- — (2005). "Methodologies for measuring the correlations between post-editing effort and machine translatability." *Machine Translation*, 19(1), 37-58.
- — (2006). "Pauses as indicators of cognitive effort in post-editing machine translating output." *Across Languages and Cultures*, 7(1), 1-21.
- — (2010). "Introduction to Post-Editing: Who, What, How and Where to Next." *The Ninth Conference of the Association for Machine Translation in the Americas*, Denver, Colorado.
- — (2011). "Towards predicting post-editing productivity." *Machine Translation*, 25(3), 197-215.
- **Orero, Pilar** (2006). "Synchronisation in Voice-over." J.M. Bravo (ed.) *New Spectrum in Translation Studies*. Valladolid: University of Valladolid, 255-264.
- **Ortiz-Boix, Carla** (*Forthcoming*). "Post-Editing Wildlife Documentaries: Challenges and Possible Solutions."
- **Papineni, Kishore; Salim Roukos; Todd Ward and Wei-Jing Zhu** (2002). "BLEU: a method for automatic evaluation of machine translation." *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics (ACL), 311-318.
- **Pym, Peter J.** (1990). "Pre-editing and the use of simplified writing for MT: an engineer's experience of operating an MT system." Pamela Mayorcas (ed.) *Translating and the Computer* 10. London: Aslib, 80-96.
- **Shreve, Gregory M., Isabel Lacruz and Erik Angelone** (2011). "Sight Translation and Speech Disfluency: Performance Analysis as Window to Cognitive Translation Processes." Cecilia Alvstad, Adelina Hild and Elisabet Tiselius (eds) *Methods and Strategies of Process Research. Integrative approaches in Translation Studies*. Amsterdam/Philadelphia: John Benjamins, 121-146.
- **Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla and John Makhoul** (2006). "A Study of Translation Edit Rate with Targeted Human Annotation." *Proceedings of the 7<sup>th</sup> Conference of the Association for Machine Translation of the*

Americas, 8-12 August 2006 in Cambridge, Massachusetts, USA. Massachusetts: AMTA, 223-231.

- **De Sousa, Sheila C., Wilker Aziz and Lucia Specia** (2011). "Assessing the Post-Editing Effort for Automatic and Semi-Automatic Translations of DVD Subtitles." *RANLP*, 97-103.
- **Tatsumi, Midori and Johann Roturier** (2010). "Source Text Characteristics and Technical and Temporal Post-Editing Effort: What is Their Relationship." *Proceedings of the Second Joint EM+/CNGL Workshop Bringing MT to the User: Research on Integrating MT in the Translation Industry (JEC10)*, 43-51.
- **TAUS** (2009). *LSPs in the MT loop: current practices, further requirements*. <<https://www.taus.net/think-tank/reports/translate-reports/lsp-in-the-mt-loop-current-practices-future-requirements>> (last accessed: 18<sup>th</sup> May 2016)
- **Veale, Tony and Andy Way** (1997). "Gaijin: A bootstrapping, template-driven approach to example-based MT." *Proceedings of the New Methods in Natural Language Processing (NeMNL97)*. Sofia: 239-244.
- **Volk, Martin** (2008). "The Automatic Translation of Film Subtitles. A Machine Translation Success Story?" *Journal for Language Technology and Computational Linguistics*, 23(2), 113-125.

### Filmography

- National Geographic (ed.) (2009). "Must Watch: A lioness adopts a baby antelope." *Unlikely Animal Friends*. Episode: "Odd Couples."

### Acknowledgements

This article is part of the *ALST project* (reference FFI-2012-31024), directed by Anna Matamala and funded by the Spanish "Ministerio de Economía y Competitividad," and also of research group TransMedia Catalonia (2014SGR27). This article is also part of the research carried out by Carla Ortiz-Boix under the supervision of Dr Anna Matamala within the PhD in Translation and Intercultural Studies (Department of Translation, Interpreting, and East Asian Studies) at Universitat Autònoma de Barcelona.

The data analysis was carried out during a research stay at Dublin City University. We would also like to thank Dr Sharon O'Brien for welcoming Carla Ortiz-Boix on a research stay at Dublin City University.

## Biographies

**Carla Ortiz-Boix**, BA in Translation and Interpreting (UAB) and MA in Translation, Interpreting and Intercultural Studies (UAB), is a PhD student at the Translation and Interpreting Department at the Universitat Autònoma de Barcelona. Member of the CAIAC research centre and Transmedia Catalana, she was awarded the FI-DGR 2013 pre-doctoral scholarship by AGAUR, an Agency of the Catalan Government.



E-mail: [Carla.Ortiz@uab.cat](mailto:Carla.Ortiz@uab.cat)

**Anna Matamala**, BA in Translation and Interpreting (UAB) and PhD in Applied Linguistics (UPF). Tenured Lecturer at the Department of Translation and Interpreting at the Universitat Autònoma de Barcelona. Member of the CAIAC research centre and of TransMedia. Audiovisual translator 1996-2007. Anna Matamala has participated in many funded projects on audiovisual translation and accessibility and has published in international journals such as *Meta*, *The Translator*, *Perspectives*, *Babel*, *Linguistica Antverpiensia* and *Jostrans*. She has published a book on interjections and lexicography (2005), and co-authored one on voice – over (with Eliana Franco and Pilar Orero) of a book. She also co-edited three books on audiovisual translation (*Listening to Subtitles*, with Pilar Orero; *Audiovisual Translation in Close-Up*, with Adriana Serban and Jean-Marc Lavaur, and *New Insights in Audiovisual Translation*, with Jorge Díaz-Cintas and Josélia Neves).



E-mail: [Anna.Matamala@uab.cat](mailto:Anna.Matamala@uab.cat).

---

**Notes**

<sup>1</sup> BLEU (Bilingual Evaluation Understudy) and h-BLEU (human targeted Bilingual Evaluation Understudy) are standard automatic measures used to evaluate MT output. The result of these measures arises by comparing MT output with a reference text that can be either its post-editing (BLEU) or a human translation (h-BLEU).

<sup>2</sup> TER (Translation Edit Rate) and h-TER (human targeted Translation Edit Rate) are two other automatic measures used to evaluate MT output. These metrics highlight errors and calculate the edits required in the MT output, in order for the text being edited to resemble a reference text that can be either its post-editing (TER) or a human translation (h-TER).

<sup>3</sup> APR and PWR have been calculated using the total number of pauses and with those pauses being made only in the main document. These two conditions have been selected because the first determines the total cognitive effort and the second specifies cognitive effort within the document where technical effort is the focus.