



What does Translation Memory do to translation? The effect of Translation Memory output on specific aspects of the translation process

Benjamin Alun Screen
Cardiff University
screenb@cardiff.ac.uk

DOI: 10.12807/ti.108201.2016.a01

Abstract: This article reports on a key-logging experiment carried out in order to investigate the effect that Translation Memory matches in the 70%-95% range have on particular aspects of the translation process. Operationalising the translation process as text (re)production following Englund-Dimitrova (2005), *Translog-II* is used to investigate whether the use of fuzzy matches in this range can reduce cognitive effort based on Working Memory Capacity and recorded pauses, to study the effect that adapting and correcting fuzzy matches in this range has on linear and non-linear writing processes, and to examine variables related to revision, time and productivity. Results show that initial reading time and self-revision is longer in the case of fuzzy match correction compared to manual translation. Data also show however that cognitive load as measured by pauses is reduced and that productivity is also increased. Significant differences are also observed in terms of text production strategies between the translators who edited the fuzzy matches and those who translated without them.

Keywords: Translation Memory; Cognitive Effort; Text Production; Minority Languages; Productivity

1. Introduction

Translation Memory (TM) software first became available commercially in the 1990s, with the concept having first been debated in the 1970s (Somers, 2003, p. 31). Since then, the use of TM tools has grown steadily. This growth in uptake has coincided with a change in pricing, and a three-tier pricing structure has emerged within which different ‘matches’ provided by the TM systems are often remunerated differently (O’Brien, 2007, p. 80). A general decrease in the amount translation commissioners and language service users are prepared to pay for translation has also been observed (DePalma, Stewart, & Whittaker, 2010). This article reports on a keylogging experiment using *Translog-II* carried out to investigate the effect that the use of fuzzy matches from a TM system within the 70%-95% range has on translation, concentrating on the effect that the adaptation and correction process has on the translation process compared to manual translation. This article therefore posits the question as to whether the use of TM material to form a translation could be considered a revision process which is easier than manual translation, thus deserving of lower remuneration. This is done within a hypothetico-deductive framework, in which the research question is broken down into eight related deductive hypotheses which have been tested experimentally. In

order to answer this research question, participants were required to translate from English to Welsh, a language pair that has undergone considerable professionalization over the last few years, with its professional organization (Welsh Association of Translators and Interpreters) having been established in 1976.¹ The experimental procedure measures dependent variables which are directly linked to cognitive and text production effort, as well as those related to time and productivity.

2. Theoretical Framework

In order to select the appropriate dependent variables, a description of the translation process must be provided. Following Englund-Dimitrova's (2005) process study on explication in translation, the translation process can usefully be modelled by considering translation first and foremost as a text production process (or text reproduction process), albeit an expert one for which excellent knowledge of two languages and their cultural backgrounds are the most basic requirements. Englund-Dimitrova (2005), based on the original research of Hayes (1996), identifies within translation three broad processes, which each contain a number of sub-tasks. Translation and its three broad processes will be described below.

2.1 Planning

Limiting the discussion to professional written translation, the purpose of translation is to produce a (target) text (TT) in Language B of a (source) text (ST) which already exists in Language A. In contrast to monolingual writing therefore, the translator always bases the TT on another text; the text created by the translator is neither independent of the original text upon which it is based, nor is it necessarily a text in its own right as it will always be compared for fidelity against another extant text. In order to create the new target text, the translator is first required to acquire degrees and shades of meaning contained within the ST by a process of reading and comprehending. This process is essential, for without it no new TT, which in professional translation contexts must communicate the meaning of the ST correctly and accurately, could exist. A translator could create a new text without reference to another text, but in such a scenario the resultant text could not be considered a translation. Relevant elements of the ST must first be read and understood therefore, and within the context of TM systems, this is likely to be the whole source segment displayed, or sub-sentential elements within it. This initial period of reading and understanding for meaning, or the process of finding a 'Meaning Hypothesis' (Gile, 1995, p. 102), has been called by Translation Process researchers 'Orientation Time' (Jakobsen, 2002, 2005) and 'Translation Onset Time' (Vandepitte, Hartsuiker, & Van Assche, 2015). The strategies that translators apply in this initial stage of producing a translation tend to vary; some translators may prefer to read the whole segment (or even the whole text) before beginning to form a mental translation, whilst others prefer to read sections of a segment and translate each individually (Englund-Dimitrova, 2005, p. 24). It is this initial reading period that can be referred to as 'planning', as the translator first has to analyze and deconstruct the ST in order to be able to transfer meaning accurately to the TT.

¹ A Welsh medium MA in Translation Studies has also been established recently as a result of increased demand in Wales for professional, high quality translations in Welsh of English material.

2.2 Producing Mental Target Language Material and Text Production

Having read the ST in an initial process of gathering ST meaning, the translator is then required to formulate Target Language (TL) material, before following a process of text production where the translation is recorded. It follows that just as initial reading time to capture ST meaning is essential, so is forming TL material and producing this material on screen. Within professional translation contexts, this will most likely be done by interacting with computer hardware such as the keyboard and mouse. Such is the position of computers in professional translation that one researcher has suggested that translation could also be viewed as a form a Human-Computer Interaction (O'Brien, 2012). This process of forming an acceptable mental target translation before writing it on screen can take one of two courses (Englund-Dimitrova, 2005, p. 26);

- The translator, having captured ST meaning, will form appropriate TL meaning before producing it on screen through text production processes, and this overall process will be simultaneous and not be characterized by frequent pausing between mental TL material forming and TT production;
- The second course is slower, filled with pauses between the forming of mental TL material and TT production even after initial reading time (which would also be recorded as pauses by *Translog-II* software), and may also contain false starts where the translator tries and tests more than one translation. This course of translation is littered with 'internal' pauses that reflect cognitive processing related to problem solving.

The second course is likely to be familiar to many a translator, as this tendency to pause between forming TL material and TT production in response to SL meaning is necessary due to working memory capacity (WMC) (Englund-Dimitrova, 2005, p. 27). As working memory within the information processing paradigm is considered to be limited (Baddeley, 1999, p. 46), it follows that translation, as a linguistic task, is also subject to these constraints on working memory. Within segments that contain SL material that require deeper processing on the part of the translator (as opposed to SL material for which the translator can find acceptable TL material quickly based on Long Term Memory and experience), the translator will be forced to concentrate on individual problem elements of the segment at a sub-sentential level. The translator is required to do this whilst keeping in mind the whole due to WMC (Dragsted, 2005, p. 50), which is broadly similar to the definition offered by Tyler (1979).^{2 3} Disregarding momentarily the time spent pausing at the beginning of a segment, as well as time spent revising, it would follow that internal pauses recorded between initial reading and segment revision are highly likely to be correlated with cognitive problem solving within that segment. If the number of internal pauses in a segment is high, then one could postulate that due to WMC the translator divided attention at different times between different translation problems, and so the cognitive effort related to this particular segment (or whole text) can be considered high. A similar point is made by Shreve and Diamond (1997, p. 243), "Frontal systems may employ the central executive to initiate more effortful processing

² A similar point is made by Newell and Simon (1972), who argue that due to WMC, the human brain cannot focus on different demanding tasks at one time and carry them out successfully.

³ The definition offered by Tyler (1979) is based entirely on WMC and is as follows "[Cognitive Effort is] the amount of available processing capacity of the limited-capacity central processor utilized in performing an information-processing task".

and allocate attentional and other resources when we become aware of what we might loosely call ‘*information processing problems*’.⁴ This will then lead to ‘reductions in the efficiency with which a particular task is performed’ (Shreve and Diamond, 1997, p. 243). Based on the theories of Butterworth (1980) and Schilperoord (1996) regarding the correlation of pauses and cognitively effortful processing, a number of Translation Process researchers have also utilized pause metrics as a way to gauge cognitive effort (cf. Dragsted, 2005; Jakobsen, 2002; Vandepitte, et al., 2015), as well as researchers interested in the relationship between cognitive effort and the post-editing of Machine Translation (Krings, 2001; Lacruz & Shreve, 2014; O’Brien, 2006). When investigating cognitive effort, pauses associated with *effortful* processing in text translation are arguably more likely to be captured by excluding pauses thought to be associated with orientation and revision time, i.e. by measuring pause activity between first and final keystrokes. Recent research has shown that more effort is invested by translators in transfer processes than in the orientation and revision phases (Jakobsen & Jensen, 2008; Sharmin et al., 2008; Pavlović & Jensen, 2009), and so it follows that an attempt to isolate the pauses linked to these different processes would be desirable. As noted by Jensen (2011, p. 49), “Assuming that processing effort is identical across all three production stages entails the risk of basing an analysis on data that reflect several tasks”.

The second stage of producing a translation then consists of forming mental TL material, having gleaned SL meaning, before transferring the formed translation to its written form on screen. The terms used to define this second phase include the ‘Drafting Phase’ (Jakobsen, 2002) and ‘Drafting’ (Carl, Kay, & Jensen, 2010). The process of forming mental TL material will manifest itself as a pause before TT production, and the number and duration of these pauses can be linked to cognitive effort (Kumpulainen, 2015), as operationalized by WMC.

2.3 Revising and Evaluating Produced TL material On Screen

Typed translations will also undergo a third and final process of ‘self-revision’ whereby the original translation produced by the translator will be read over to ensure grammaticality and fidelity. This final process may lead to further changes being made, but this is not guaranteed. Mossop (2014) recognizes two main forms of ‘self-revision’. The first is revision performed whilst a translation is being drafted, and has been called a ‘Monitoring Pause’ (Dragsted, 2010) and ‘Online Revision’ (Jakobsen, 2003). The second type refers to revision carried out by the same translator having completed a first draft. A third category could also be added to this in order to capture the tendency to revise whilst translating *as well as* revising a complete draft, a double revision strategy that the translators who took part in this study also used. These different revision strategies have also been discovered by Alves and Vale (2011) and Dragsted and Carl (2013), who combined Eye Tracking and Keylogging to glean valuable insights into the Translation Process.

The Translation Process then, following the insight of Englund-Dimitrova (2005) who modelled translation on Hayes’ (1996) description of the monolingual writing process, can be broken down into three main constituent parts which broadly include reading for meaning, forming a mental translation and typing it, before an evaluation exercise called ‘self-revision’. Now that a useful model has been outlined, it is possible to describe the dependent variables measured in the controlled experiment carried out in *Translog-II*.

⁴ Frontal systems as used by Shreve and Diamond (1997) would relate to the frontal lobe of the human brain.

3. How does the use of TM matches affect aspects of the Translation Process?

Given the outline provided above of the translation process in three parts, how does the adaption and correction of fuzzy matches between 70%-95% disrupt, change or affect this process? Does the adaption and correction of these matches increase or decrease cognitive effort for example, and what is the relationship between these matches and efficiency? Table 1 below posits 8 hypotheses related to possible ways in which the use of TMs may alter this process, as well as noting the dependent variables chosen to test them. These variables could also be linked to Krings (2001) triadic structure of effort, namely Cognitive Effort, Technical Effort and Temporal Effort, usually found within Machine Translation research. Recent examples of Krings' (2001) triadic classification of effort being applied in comparative studies are Koglin (2015) who considered the comparative cognitive effort required to post-edit metaphors as opposed to translating them, as well as considering technical and temporal effort, and O'Brien (2006; 2007), who compared the cognitive, technical and temporal effort required to post-edit segments that had been subjected to controlled language rules as opposed to post-editing segments that had not.

4. Data Collection

An experiment was conducted at Cardiff University, UK, designed to investigate the effect that TM fuzzy matches have on certain aspects of the translation process defined above. A secondary aim was to contribute to the Translation Process literature from the perspective of an under-researched language pair, and to contribute evidence from a controlled experiment. Data was gleaned from the Linear Representation provided by *Translog-II*, as well as its Replay Function. All statistical analysis was done using IBM SPSS, and the confidence threshold used was 95%.

4.1 Participants

Nine professional translators were recruited, all of whom were members of the Welsh Association of Translators and Interpreters. Membership of this professional translation organization, and working as a full-time Welsh translator, were prerequisites for taking part in the study. All participants were familiar with Translation Memory tools, and all participants had experience of editing fuzzy matches of various ranges as well as experience of translation.

4.2 Experimental Design

Utilizing a between-group design methodology, the nine translators were randomly split into two groups; the Control Group (n=4) translated 11 sentences of text (240 words) about Web2.0 into Welsh manually (Appendix A), while an Experimental Group (n=4) adapted 11 matches within the 70%-95% fuzzy match range (Appendix B). Participant 6's data had to be discarded as the translator spoiled their data by criticizing the quality of the ST in the TT. Participants 1, 2, 3 & 4 were asked to adapt and correct the fuzzy matches, while Participants 5, 6, 7, 8 & 9 were asked to translate without the aid of any translation technology. Using a TMX file freely available to Welsh translators from the now dismantled Welsh Language Board, this TMX file was imported into a TM system (Déjà Vu X3 Professional), with 11 segments extracted to form a ST for the Control Group.

Table 1: Hypotheses and Dependent Variables Measured

Hypotheses and Dependent Variables Measured			
Sub-task	Hypothesis	Related Dependent Variable	Operationalisation
Planning	(1) TM material will need to be checked for fidelity against another text. This requirement to read two texts will extend the initial phase of translation for the Experimental Group	(1) Total Duration of Initial Pauses recorded before any text production	All pauses recorded after the appearance of the segment and striking the first key in that segment
Finding TL material and Text Production	(2) The revising of TM material makes the process of finding TT material cognitively easier for the Experimental Group, as this material will prime the translator, rather than the translator having to form original material	(2) Total Duration of Internal Pauses recorded	All pauses recorded after the first key stroke and before the last key stroke in that segment
	(3) As the TM material will contain text that will be recycled, the number of alphanumeric keystrokes struck by the Experimental Group in the process of text production will be less than those struck by the Control Group	(3) Any alphanumeric keys struck in producing the TT	Any letter, number or punctuation struck in typing the TT
	(4) Given that the revision of TM material will require the translator to delete and edit text rather than type it, the number of keys struck that are related to editing extant text will be higher than in translation	(4) Any key struck to manipulate, seek and adapt text, as well as mouse clicks	All BACKSPACE, DELETE, Space Bar, arrows, CTRL combinations and mouse clicks
	(5) There will be no difference between the Control and Experimental Groups in terms of reading typed text during the drafting phase	(5) Duration of Final Pauses	Any pause recorded after the last key stroke and appearance of next segment
Self-Revision	(6) There will be no difference between the Control and Experimental Groups in terms of final revision of completed TT	(6) Time spent after completing first draft of whole text	Time recorded after the first draft of the last segment has been completed
	(7) The Experimental Group will complete first drafts of all segments quicker than the Control Group	(7) Processing Time	Time taken to produce a first draft of all segments
Productivity	(8) As a result of Hypothesis (7), the Experimental Group's productivity, measured in Words per Minute, will be higher than that of the Control Group	(8) Words per Minute	WPM is calculated by dividing the Total Processing Time by the number of words in the ST

The segments that were to be given to the Control Group were then adapted by adding and deleting elements of these 11 segments. This same file was then reloaded into the TM, and 11 fuzzy matches were created using the 'Pretranslate' process. No time limit was set for the completion of either task.

4.3 Apparatus

The software used to collect data was *Translog-II* and as this research software is unfamiliar to most translators, all participants were asked to type a short paragraph in English in the software in order to gain familiarity with how *Translog-II* looks, how to open projects and how to save files. It also allowed participants to become accustomed to a new keyboard and a different workstation. In terms of the Control Group, the English ST was shown on the left and the TT window on the right within the *Translog-II* interface, using its parallel screen option. Participants were asked to click 'ESC' in order to be able to see the next segment. Participants were asked not to proceed to the next segment until they had finished the previous one. In terms of the Experimental Group, all 11 source segments were displayed on a parallel screen, but in order to see the next match the participants were required to strike 'ESC' also. The participants in the Experimental Group were asked not to press 'ESC' until they had finished processing the previous fuzzy match. The parallel layout chosen for the *Translog-II* GUI therefore was kept constant for both groups.

5. Results

5.1 Reading Time

The time usually associated with reading the ST, defined here as the Initial Pauses recorded by *Translog-II* between the striking of 'ESC' – causing the appearance of the active segment – and the first translation keystroke (Variable 1), was found to be longer in the case of the Experimental Group. The differences observed were also found to be statistically significant according to a Mann-Whitney test; $U=1,508.5$, $p=.000$, Control Group Median = 3.75, Experimental Group Median= 10.5. It is likely that the translators who processed the fuzzy matches read the ST as well as the content of the fuzzy match in order to check it for fidelity, and more than likely also for grammaticality. Therefore Hypothesis (1) can be accepted. Figure 1 below shows the median for both groups.

5.2 TT Production and Cognitive Effort

The adapting and correcting of fuzzy matches within the chosen range also had a tangible effect on a number of other variables selected. Starting first of all with the finding of TL material after the ST has been processed before text production, as measured by pauses located between the first and last key stroke in the production of a first draft of the segment (Variable 2), participants in the Control Group on average tended to pause more before TT production than their counterparts in the Experimental Group. This is shown in Figure 2 below. The difference in average pause length between first and last key stroke in a segment between the two groups was also found to be statistically significant according to a Mann-Whitney test; $U=450.500$, $p=.000$, Control Group Median= 28.5, Experimental Group Median= 11.25. Given this statistically significant result, Hypothesis (2) must be accepted.

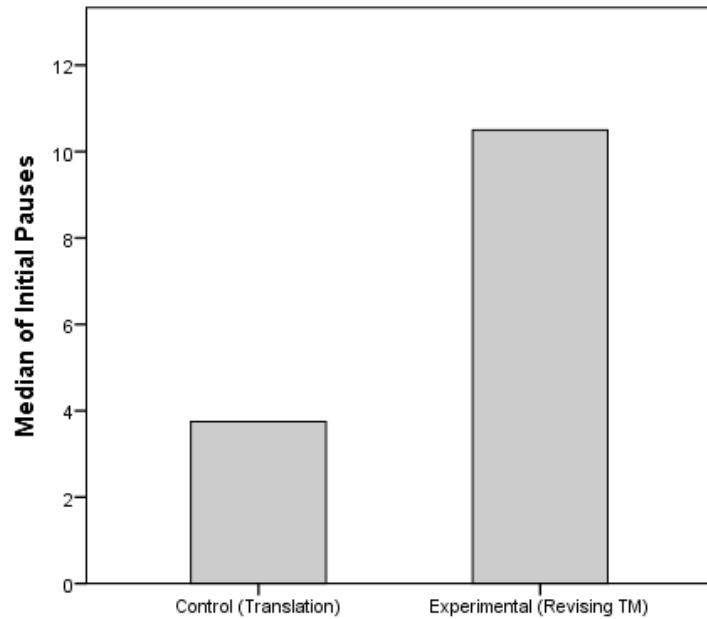


Figure 1: Median of Initial Pauses for both groups

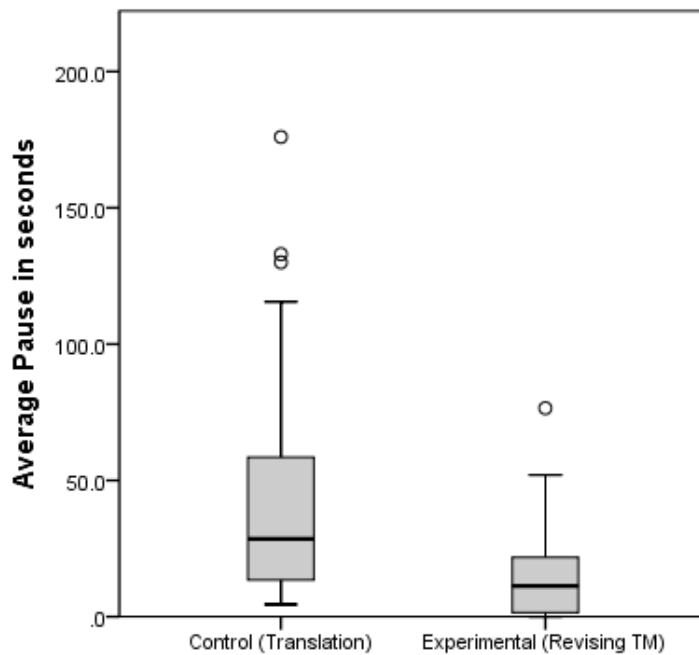


Figure 2: Average Length of Pauses between First and Final Key Strokes

Once TL material has been found, this mental translation then needs to be typed. In terms of TM systems and the adapting and correcting of fuzzy matches, this will also include the analyzing of matches and planning a sufficient and appropriate correction strategy. In terms of text production, these different cognitive process also lead to different text production processes. Whilst translation usually leads to linear text production, adapting and correcting fuzzy matches will likely lead to non-linear writing processes, whereby text is not typed in a linear fashion without any barriers, but manipulated, deleted, moved around and adapted. As it can be expected that correcting the output of the TM system will require more non-linear writing strategies, this will manifest itself by the use of certain keys and an increased

reliance on the mouse. As such, Variable (4) was measured as a way to gauge this difference and to test Hypothesis (4). Because much of the required text is already on the screen for the translator, it could also be expected that the number of alphanumeric characters and therefore words that are typed will be significantly less than what it would be in translation, and so Variable (3) was measured also to test this hypothesis (namely Hypothesis (3)).

In terms of linear text production and Hypothesis (3), there was a statistically significant difference between the average number of alphanumeric characters produced by the Control and Experimental Group according to the Mann-Whitney Test ($U=1.000$, $p=.000$, Control Group Median= 115.5, Experimental Group Median= 9). Hypothesis (3) can therefore be accepted. Figure 3 shows the difference observed between the two groups.

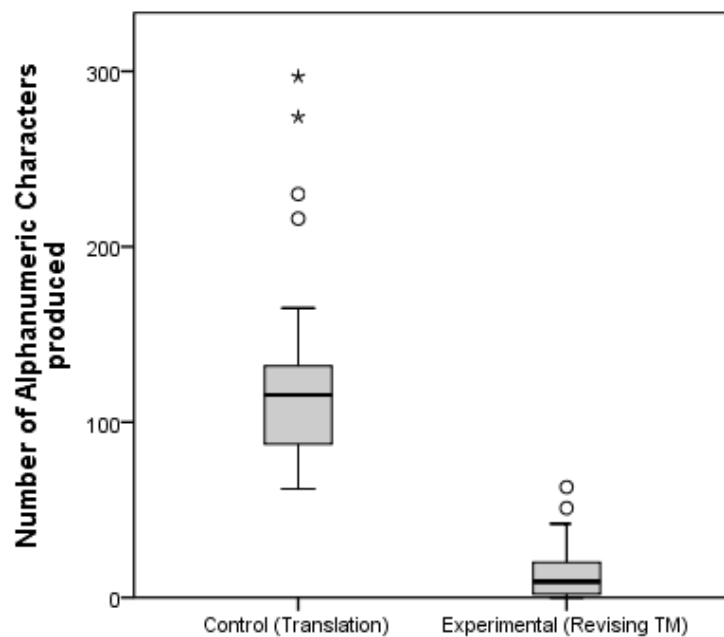


Figure 3: Number of alphanumeric characters produced by participants in both groups

Non-linear writing processes manifest themselves as copy and paste procedures, deletions, addition to text, CTRL combinations, mouse clicks and the use of the arrows. The Linear Representation did not suggest that any copy and paste actions occurred in any of the Control Group and Experimental Group translators' text production strategies, and when all sessions were watched back using the Replay Function of *Translog-II*, this was confirmed. There was a significant difference between the number of mouse clicks recorded on average between the Control and Experimental Groups according to a Mann-Whitney test ($U=1,228$, $p=.025$, Control Group Median = 1, Experimental Group Median= 1.5). The average number of keystrokes used to move through and around text, namely arrows, CTRL+Right/Left and the Space Bar were also found to be higher for the Experimental Group than in the Control Group, and this difference was also statistically significant according to a Mann-Whitney test ($U=1,245$, $p=.014$, Control Group Median= 0, Experimental Group Median= 6.5). The average number of deletions recorded (use of BACKSPACE and DELETE) was found to be higher however for the Control Group and this difference was also statistically significant according a Mann-Whitney test ($U=.685$, $p=.018$, Control Group Median= 9.5, Experimental Group Median= 4). One possible explanation for this is that the

translators in the Control Group tended to change their first renditions having read over them, rather than type a suitable translation before moving on. Using *Translog-II's* log replay functionality to watch back each Control Group translator's session, this explanation is likely as this group of translators did type a rendition before deleting it and starting again in a number of cases. Despite this result then it seems more likely that, all in all, the adaption and correction of fuzzy matches will often lead to more non-linear writing strategies and as such Hypothesis (4) can tentatively be accepted.

Table 2 below notes the total number of keystrokes recorded according to segment. Data from every participant was added in order to obtain each segment score. 'Other' refers to CTRL combinations, Space Bar and arrows.

Table 2: Total Keystrokes Recorded for each segment

Seg	Control Group					Experimental Group				
	Alphanumeric Characters	Deletions	Clicks	Other	Total Keys Strokes	Alphanumeric Characters	Deletions	Clicks	Other	Total Keys Strokes
1	536	62	12	206	816	19	35	5	114	173
2	386	70	10	5	471	6	19	5	61	91
3	516	41	5	2	564	35	23	8	17	83
4	1017	67	8	77	1169	101	127	17	148	393
5	331	12	4	0	347	50	35	3	22	110
6	373	73	9	87	542	62	34	3	49	148
7	510	77	1	40	628	91	43	6	125	265
8	363	16	3	0	382	48	38	8	8	102
9	482	53	8	119	662	68	38	8	156	270
10	306	22	2	0	330	46	46	13	48	153
11	551	101	5	97	754	50	60	13	88	211
Total	5371	594	67	633	6665	576	498	89	836	1999
Mean	488	54	6	58	606	52	45	8	76	182
Median	482	62	5	40	564	50	38	8	61	153
Standard Deviation	187	27	3	64	234	27	28	4	51	91

5.3 Self-revision

All three self-revision strategies outlined above were to be found in the data, but the tendency to self-review after completing the first draft of a segment, as opposed to leaving most revision until the end, was preferred by the Experimental Group. As a result, Participants 2 and 3 in this group did not then revise the text in its entirety after completing the translation (denoted by 'NFR' ('No Final Revision') in Table 3), and the final revision times (Variable 6) for Participants 1 and 4 were lower than the times for those in the Control Group (apart from the revision time of P8 whose revision time was lower than the time recorded for P4 but higher than that recorded for P1). The translators in the Control Group tended to leave revision until after they had completed a first draft. The total of Final Pauses (Variable 5) recorded for both groups is shown in Figure 4 below. The differences between the two groups in terms of total duration of pauses after the final key stroke was recorded were found to be statistically significant according to a Mann-Whitney test; $U=1,275$, $p=.000$, Control Group Median = 1.5, Experimental Group Median=3. As a result, Hypothesis (5) must be rejected.

The fact that the translators in the Experimental Group tended to spend longer reading over the corrected TM segments however is an interesting finding. It is possible that the translators in the Experimental Group felt that the final rendering needed a last check as the TM material was written by translators other than themselves, and because they did not know the origin of the TMX at the time. However, the fact that P2 and P3 carried out no final revision undermines this tentative conclusion. This then warrants further investigation, possibly through a mixed methods design where qualitative data is collected regarding translators thoughts about the use of TMs and quality, as

well as quantitative process data. Final revision times (Variable 6), i.e. time taken to perform revision activity after the final segment had been completed, is shown in Table 3.

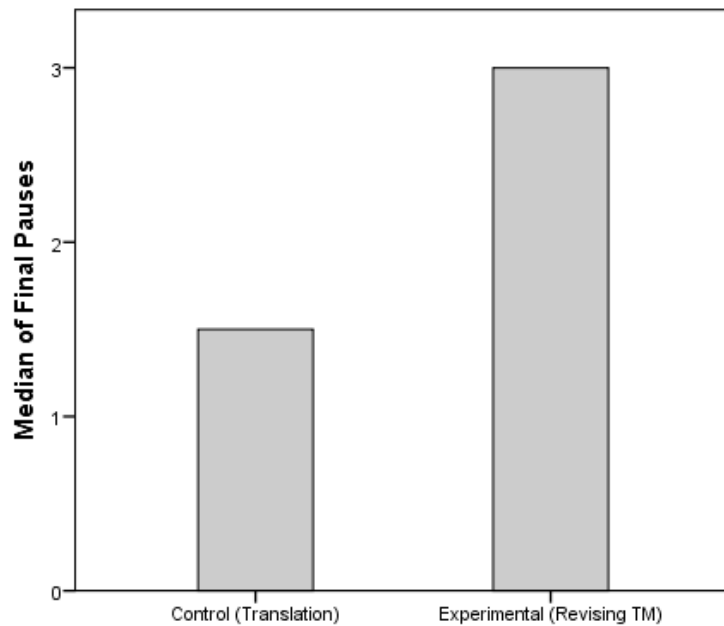


Figure 4: Median of Final Pauses for both groups

Table 3 shows the total time spent revising the text after a first draft of all 11 segments had been completed. The final revision times for the Control Group are much higher than those recorded for Participants 1 and 4 (apart from the case of P8 whose revision time was less than P4 but higher than that of P1). As a result, Hypothesis (6) must be rejected.

Table 3: Final Revision Times for Both Groups

Participant	Control Group				Experimental Group			
	P5	P7	P8	P9	P1	P2	P3	P4
Time	00:11:24	00:15:22	00:02:41	00:05:31	00:01:44	NFR	NFR	00:03:12

5.4 Time and Productivity

The final hypotheses to be considered are those related to time and productivity. The hypothesis related to time posited that the adaption and correction of fuzzy matches within the 70%-95% range as an independent variable would speed up the translation process (Hypothesis 7), and that this in turn would then improve productivity as measured by Words per Minute (WPM) (Hypothesis 8). When the times taken to complete a first *draft* of all segments are considered for both groups, as opposed to the time taken to complete a first draft of the text in its entirety *including* final revision, the adaption and correction of the fuzzy matches in the chosen range by the Experimental Group (Figure 5) was found to be quicker than translation by the Control Group (Figure 5). This difference was also found to be statistically significant according to the non-parametric Mann-Whitney test ($U=369.00$, $p=.000$, Control Group Median= 73.5, Experimental Group Median= 36.5). As a result, Hypothesis (7) may be accepted. Figure 5 below displays the time taken then to complete a first draft of every segment, before the translator went through the text again during the final self-revision process:

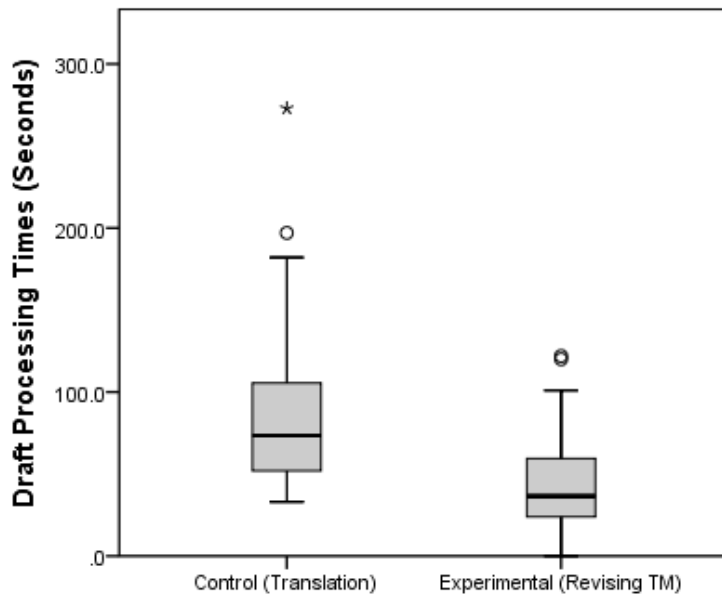
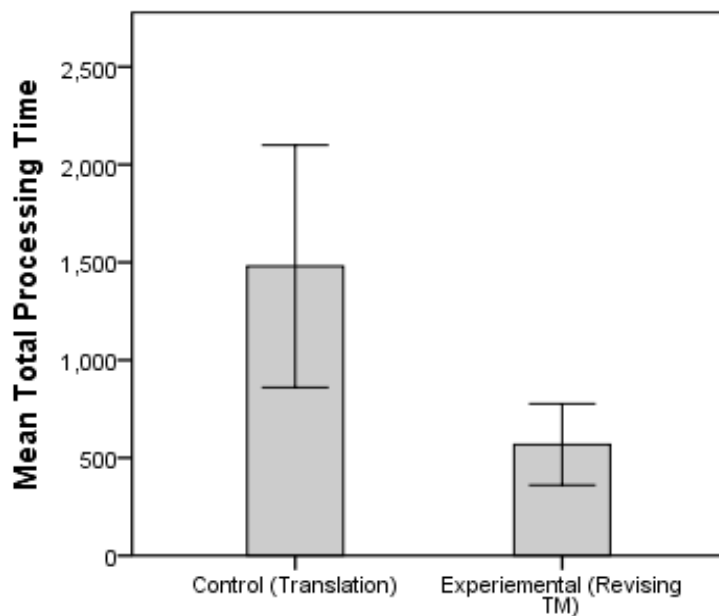


Figure 5: Draft Processing Times in Seconds for both groups

When Total Processing Time is considered, namely the time taken to complete the translated text in its entirety *including* final revision, the same trend can be observed; the Experimental Group completed the final translation quicker than the Control Group who translated without any form of output from a TM system of any match value. This is plotted below in Figure 6:



Error Bars: 95% CI

Figure 6: Total Processing Time

Productivity is also an important consideration in the translation industry, and one metric used is Words per Minute (WPM). This is calculated by dividing the number of ST words by the Total Processing Time. The WPM values (Variable 8) are plotted in Figure 7 below; the differences in productivity between the Control and Experimental Group are considerable

(Control Group Mean = 10.25, Standard Deviation = 3.3, Experimental Group Mean = 27.5, Standard Deviation = 7.3). The productivity data set was the only data set collected that did not violate assumptions of normality and homogeneity of variance, and so in order to discover whether this difference was significant, a parametric Independent Samples t-Test (two-tailed) was used. The test showed that the difference in WPM (Variable 8) was indeed statistically significant ($t(6) = -4.3, p = .005$). Figure 7 displays this difference graphically:

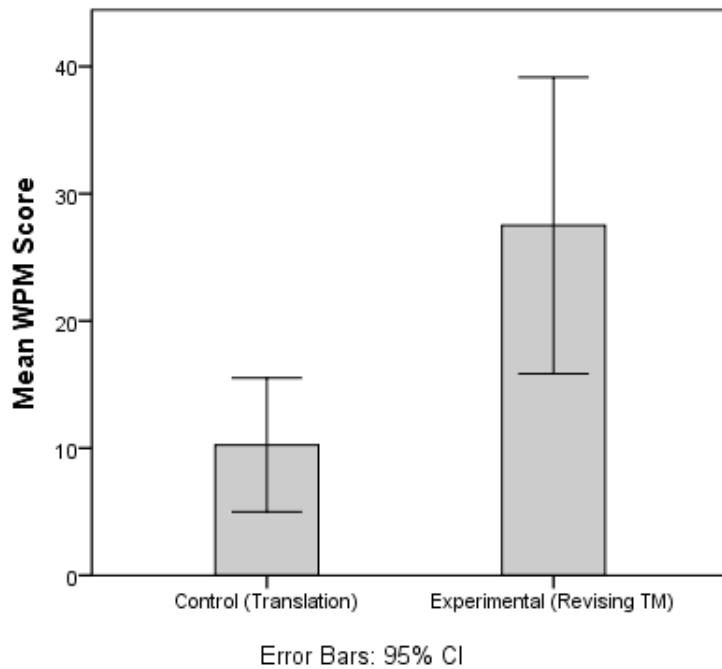


Figure 7: WPM calculated for each participant

Accordingly, the use of fuzzy matches within the 70%-95% range speeded up the translation process in the experiment reported on here. WPM, as a popular metric for productivity, was also higher in the case of those who had access to the TM output.

6. Discussion and Conclusion

This article proposed to investigate the effect that TM matches in the 70%-95% range have on specific aspects of the translation process, and in so doing *Translog-II* was used for the first time in a Welsh language context. Through the medium of the article's main research question, cognitive effort was considered, as were variables related to text production effort and translator productivity. This could be linked to Krings' (2001) three-pronged classification of effort, a classification usually utilized when investigating effort in the post-editing of Machine Translation output. A further aim was to contribute to wider research in the field of translation technology and the effect it has on translators and translation from the point of view of a language which in its own country is translated into and from as an official language on a daily basis, but which up until now has received very little attention.

In terms of main findings, it was found that the initial period of reading for meaning, referred to as 'Planning' by Englund-Dimitrova (2005), was longer in the case of those who were expected to adapt and correct the fuzzy matches compared to the translators who translated from scratch. It was suggested that the most likely reason for this increase in reading time was due

to the tendency of the participants in the Experimental Group to read over the fuzzy match after reading the ST. This is to be expected however, as even exact matches will need to be checked for fidelity (even if not for grammaticality depending on the quality of the memory being used). Cognitive effort was operationalized here as Internal Pauses, which are considered to be a manifestation of the limitations of Storage Capacity of the working memory. Using the pauses in this middle position between reading time and revision time as an indicator of cognitive effort, it was found that the adaption and correction of fuzzy matches in the 70%-95% range was cognitively easier than manual translation. The differences between the Control and Experimental Groups in terms of average Internal Pause Length was also statistically significant. In keeping with Mossop's (2014) two-tier classification of self-revision types, as well as the double revision strategy favoured by some participants, it was found that participants in the Experimental Group preferred to revise as they went rather than revising the whole text in its entirety, whereas the opposite was seen in the case of the Control Group. There were also significant differences to be found between a number of variables related to text production effort. Those in the Experimental Group produced less alphanumeric characters than the Control Group; as most of these characters will include letters, it follows that the number of words that needed to be typed was also less in the case of adapting and correcting fuzzy matches. This difference was also found to be statistically significant. As one would expect, manifestations of non-linear writing processes were more common in the case of the Experimental Group (apart from deletions, which were more common in the Control Group); statistical differences were found between the number of mouse clicks, arrow keys, CTRL combinations and space-bar keys struck. This can be seen to be a result of the need to adapt text that is already extant, as opposed to typing it anew. Interestingly, no translator opted to recycle text by copying and pasting it. The translators in the Experimental Group who adapted and corrected the fuzzy matches in the 70%-95% range were also faster on average than their Control Group counterparts, and their WPM was also calculated to be much higher as a result.

Although O'Brien (2007) used eye tracking and pupillometrics with a questionnaire to operationalize cognitive effort, the results presented here also show that a reduction in cognitive effort is possible when users interact with TM matches as opposed to translating without them. The study reported on here also provides support for Christensen's (2011, p. 155) statement that TMs affect the translator's mental processing. Further work therefore could include the use of eye tracking in combination with *Translog-II* to measure effort, as reported recently by Carl, Gutermuth & Hansen-Schirra (2015), Koglin (2015) and Sjørup (2011), as well as scaling up the study with a larger ST and greater number of translators. Another avenue of research would be to carry out empirical process research on the post-editing of Machine Translation and the use of TM systems; the outcomes of the variables described above in Table 1 could be compared and contrasted across these conditions in order to gauge the differences and possible similarities between these two particular revision processes.

Acknowledgements

The author would like to thank the translators for giving up their very little professional time to take part in this study, as well as the *Coleg Cymraeg Cenedlaethol* for funding the PhD research upon which this article is based.

References

- Alves, F., & Vale, D. C. (2011). On drafting and revision in translation: A corpus linguistic oriented analysis of translation process data. *Translation: Computation, Corpora, Cognition*, 1(1), 105-122.
- Baddeley, A.D. (1999). *Essentials of human memory*. Hove: Psychology Press.
- Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Ed.), *Language production. Volume 1- Speech and talk* (pp. 155-176). London: Academic Press.
- Carl, M., Gutermuth, S., & Hansen-Schirra, S. (2015). Post-editing machine translation: Efficiency, strategies and revision processes in professional translation settings. In A. Ferreira & J. W. Schwieter (Eds.), *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting* (pp. 145-174). Amsterdam/Philadelphia: John Benjamins.
- Christensen, T.P. (2011). Studies on the mental processes in translation memory-assisted translation - The state of the art. *Trans-Kom*, 4(2), 137-160.
- DePalma, D., Stewart, R., & Whittaker, B. (2010). *Translation and localization pricing: A comprehensive study of what language services cost*. Lowell, Massachusetts: Common Sense Advisory.
- Dragsted, B. (2005). Segmentation in translation: Differences across levels of experience and difficulty. *Target*, 17(1), 49-70.
- Dragsted, B. (2010). Co-ordination of reading and writing processes in translation: An eye on uncharted territory. In G. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 41-63). Amsterdam/Philadelphia: John Benjamins.
- Dragsted, B., & Carl, M. (2013). Towards a classification of translation styles based on eye-tracking and key-logging data. *Journal of Writing Research*, 5(1), 133-158.
- Englund-Dimitrova, B. (2005). *Expertise and explication in the translation process*. Amsterdam/Philadelphia: John Benjamins.
- Gile, D. (1995). *Basic models and concepts for interpreter and translator training*. Amsterdam/Philadelphia: John Benjamins.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing. Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Jakobsen, A. (2002). Translation drafting by professional translators and by translation students. In G. Hansen (Ed.), *Empirical translation studies* (pp. 191-204). Copenhagen: Samfundslitteratur.
- Jakobsen, A. (2003). Effects of Think Aloud on translation speed, revision and segmentation. In F. Alves (Ed.), *Triangulating translation* (pp. 69-97). Amsterdam/Philadelphia: John Benjamins.
- Jakobsen, A. (2005). Investigating expert translators' processing knowledge. In H. V. Dam, J. Engberg & H. Gerzymisch-Arbogast (Eds.), *Knowledge systems and translation* (pp. 173-193). Berlin/New York: Mouton de Gruyter.
- Jakobsen, A., & Jensen, K.T.H. (2008). Eye movement behaviour across four different types of reading task. In S. Göpferich, A. L. Jakobsen & I. M. Mees (Eds.), *Looking at eyes: Eye-tracking studies of reading and translation processing* (pp. 103-124). Copenhagen: Samfundslitteratur.
- Jensen, K.T.H. (2011). *Allocation of cognitive resources in translation: An eye-tracking and key-logging study*. Doctoral dissertation, Copenhagen Business School.
- Koglin, A. (2015). An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors. *Translation and Interpreting*, 7(1), 126-141.
- Krings, H. P. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes*. Kent, Ohio: Kent State University Press.
- Kumpulainen, M. (2015). On the operationalisation of 'pauses' in translation process research. *Translation and Interpreting*, 7(1), 47-58.
- Lacruz, I., & Shreve, G. (2014). Pauses and cognitive effort in post-editing. In S. O'Brien, L. W. Balling, M. Carl, M. Simard, & L. Specia (Eds.), *Post-editing of*

- machine translation: Processes and applications* (pp. 246-274). Newcastle: Cambridge Scholars.
- Newell, A. & Simon, H. (1972). *Human problem solving*. New Jersey: Englewood Cliffs.
- Mossop, B. (2014). *Revising and editing for translators*. London: Routledge.
- O'Brien, S. (2006). *Machine-translatability and post-editing effort: An empirical study using Translog and Choice Network Analysis*. Doctoral dissertation, Dublin City University.
- O'Brien, S. (2006). Pauses as indicators of cognitive effort in post-editing machine translation output. *Accross Languages and Cultures*, 7(1), 1-21.
- O'Brien, S. (2007). An empirical investigation of temporal and technical post-editing effort.
- O'Brien, S. (2007). Eye-tracking and translation memory matches. *Perspectives: Studies in Translatology*, 14(3), 185-205.
- O'Brien, S. (2012). Translation as human-computer interaction. *Translation Spaces*, 1(1), 101-122.
- Pavlović, N. & Jensen, K. T. H. (2009). Eye tracking translation directionality. In A. Pym & A. Perekrestenko (Eds.), *Translation research projects 2* (pp. 101-191). Tarragona: Universitat Rovira i Virgili.
- Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production*. Amsterdam: Rodopi.
- Sharmin, S., Spakov, S., Raiha, K.J., & Jakobsen, A.L. (2008). Where and for how long do translation students look at the screen while translating? In S. Göpferich, A.L. Jakobsen & I. Mees. *Looking at eyes: Eye-tracking studies of reading and translation processing* (pp. 31-51). Copenhagen: Samsfundslitteratur.
- Shreve, G., & Diamond, B. (1997). Cognitive processes in translation and interpreting: Critical issues. In J. Danks, G. Shreve, S. Fountain, & M. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 233-252). Thousand Oaks: Sage Publications.
- Sjørup, A. (2011). Cognitive effort in metaphor translation: An eye-tracking study. In S. O'Brien (Ed.), *Cognitive explorations of translation* (pp. 197-213). London: Continuum.
- Somers, H. (2003). Translation memory systems. In H. Somers (Ed.), *Computers and translation: A translators' guide* (pp. 31-49). Amsterdam/Philadelphia: John Benjamins.
- Tyler, S. e. a. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory*, 5(6), 607-617.
- Vandepitte, S., Hartsuiker, R. J., & Van Assche, E. (2015). Process and text studies of a translation problem. In A. Ferreira & J. W. Schwieter (Eds.), *Psycholinguistic and cognitive inquiries into translation and interpreting* (pp. 127-145). Amsterdam/Philadelphia: John Benjamins.

Appendix A: The English Source Text

1. This issue is not explored here as it is a web2.0 issue and organisation will have to decide where this boundary lies.
2. While web2.0 technologies take different forms, we can conceptualise their operation thus;
3. While the present report only refer to Welsh and English, there may be contexts where other languages must be included
4. Messages are more substantial in terms of content (text, images, and video) which are much more likely to be subjected to quality Control and are less responsive (less formal, likely to be perceived of as being from the organisation instead of a member of staff).
5. Therefore there is more of an opportunity to take these through a translation Process or produce a bilingual version.
6. Messages will usually be discussion starters instead of a response.
7. The originating individual may be identified, but a message could well be presented as a collective output, e.g. from ‘‘the team’’.
8. One or more comments in response to the ‘‘message’’ for example may be posted by members of the public.
9. Any content that is created by the public (user-generated content) is referred to in this report as a comment.
10. One or more responses in response to the comments or messages may be posted by staff.
11. It is not likely that these will be subjected to any form of quality control and present more the personality of the individual member of staff.

Appendix B: The 11 Fuzzy Matches

1. Ni chaiff y mater hwn ei drafod yma gan ei fod yn fater ehangach i we2.0 a bydd angen i sefydliadau benderfynu drostynt eu hunain ble mae'r ffiniau i fod.
2. er bod sawl ffurf i dechnoleg gwe2.0, gellir sôn am eu gweithredoedd cyffredin fel a ganlyn;
3. Er mai cyfeirio'n unig at y Gymraeg a'r Saesneg y mae'r adroddiad hwn, efallai bod cyd-destunau lle y mae angen cynnwys ieithoedd ychwanegol.
4. Mae negeseuon yn ddarnau mwy sylweddol o gynnwys (testun, delweddau, sain, fideo) sy'n fwy tebygol o fod yn destun rhyw fath o reolaeth ansawdd ac felly'n llai ymatebol (llai anffurfiol, mwy tebygol o gael eu hystyried yn rhywbeth gan y sefydliad yn hytrach nag oddi wrth aelod unigol o staff, llai o bersonoliaeth).
5. Felly mae mwy o gyfle i dywys y rhain drwy broses gyfieithu neu ddarparu fersiwn gyfochrog mewn iaith arall.
6. Fel arfer, bydd negeseuon yn agor rhyw drafodaeth yn hytrach nag yn sylwadau.
7. Gallai'r unigolyn sy'n ei hanfon fod wedi'i enwi, ond gellid hefyd gyflwyno neges fel petai'n allbwn ar y cyd, e.e. oddi wrth y "tîm" ehangach.
8. Gall un neu ragor o'r sylwadau mewn ateb i'r "neges" gael eu hanfon gan y cyhoedd.
9. Yn yr adroddiad hwn, cyfeirir at o'r cynnwys sydd wedi ei greu gan y cyhoedd (cynnwys a gynhyrchir gan y cyhoedd) fel "sylwadau".
10. Gall un neu ragor o'r ymatebion sy'n ateb y sylwadau, negeseuon neu ddiwyddiadau eraill, gael eu hanfon gan y staff.
11. Mae'n annhebygol y bydd y rhain yn gallu dod o dan unrhyw drefn rheoli ansawdd. Maent yn cyflwyno personoliaeth aelod o'r staff.