Helena 29/04/2009

Looking at Eyes

Eye-Tracking studies of Reading and Translation Processing

Edited by

Susanne Göpferich
Arnt Lykke Jakobsen
Inger M. Mees

Copenhagen Studies in Language 36
Samfundslitteratur

Processing fuzzy matches in Translation Memory tools: an eye-tracking analysis[1]

Sharon O'Brien

Abstract

Little research has been carried out to date on how translators interact with Translation Memory (TM) tools. In particular, there has been very little research on the effort required to process "Fuzzy Matches". Previous research (O'Brien 2006) suggested that the relationship between Fuzzy Match value and cognitive effort was perhaps not so linear. The aim of the current paper is to investigate that relationship in more detail. Five participants translated a technical text from German into English using a translation memory. The TM was seeded with Fuzzy Matches across different categories of match value, ranging from 52% to 99% similarity. The participants' eye movements and pupil dilations were recorded using the Tobii 1750 eye tracker. Cognitive effort was measured using processing speed (words per second) and pupil dilation. A qualitative approach was also taken by asking participants to rate their perceived editing effort for each segment once the task was completed. The results suggest that when processing speed is used to measure cognitive effort, there appears to be a linear relationship between effort and fuzzy match value. However, when cognitive effort is measured via pupil dilations, no linear relationship is detected. We suggest that the results for pupil dilations demonstrate a "capacity-constrained response". The qualitative survey results suggest that segments between 80 and 99% match value are rated by participants as requiring little effort while anything between 50 and 79% is rated as requiring more editing effort.

[1] I wish to thank my colleague Dr. Minako O'Hagan for her helpful comments on a draft of this paper.

1. Introduction

The motivation for the research described in this paper stems from a general interest in how translators interact with TM tools. An additional motivation is an interest in the economics of the translation industry. Since the introduction of Translation Memory (TM) tools in the mid- to late nineties, there has been an increasing downward pressure on the rates paid for translating words (Van der Meer 2008; Bierman 2008; Beninatto & de Palma 2004; Sargent 2008).

TM tools have facilitated this downward pressure by allowing translation clients to argue that the rates paid for 'exact matches' (i.e. source-text segments contained in a translation memory that are identical to a source-text segment requiring translation in a new text) should be low, or indeed should be zero, because exact matches should not be 'touched' since they have already been translated. Evidence of this expectation, along with counter-arguments from translators, can be found in online translation discussion forums such as ProZ.com. In addition, rates for fuzzy matches (i.e. source-text segments contained in a translation memory that are similar to a source-text segment requiring translation in a new text) are also the subject of debate.

Our motivation here is to investigate the cognitive load involved in translating fuzzy matches in an empirical way to find out whether reduced rates for their translation are justified. The research is further motivated by the results of a pilot experiment using eye-tracking methodology (O'Brien 2006), which hinted at the possibility that the relationship between cognitive effort and fuzzy match value was not exactly a straightforward one.

A decision was therefore made to design an experiment that would enable the comparison of cognitive load across fuzzy matches ranging from 50% to 99% similarity. Our hypothesis was that the relationship between cognitive load and fuzzy match value was not a straightforward linear one.

The approach was to measure cognitive effort in both a quantitative way, using processing speed and pupil dilations recorded by an eye-tracker, and in a qualitative way, through a survey. The methodology is discussed in more detail below, but we should first point out some limitations of the study.

2. Limitations of the study

While the analysis of translation processes can yield some very interesting results in general (for an overview see Jääskeläinen 2002), one of the limitations associated with this mode of research is the number of participants that can be recruited. In order to make any valid claims about trends, one ideally has to have a large number of participants. In addition, unless one is interested in investigating *differences* in skills sets, equal competence and experience on the part of participants is desirable. In this study, we aimed for equal competence in terms of translation and typing skills as well as competence in using the specific translation memory tool selected (SDL Translator's Workbench). This equal competence was important in order to eliminate any effect from parameters other than the level of the fuzzy match. However, in recruiting participants we were limited to a small number of potential participants for a number of reasons: (1) a certain level of translation competence was required; (2) a specific language pair/direction was sought (German-English); (3) the ability to touch-type was necessary in order to reduce loss of eye-tracking data;[2] (4) participants had to be competent in the use of the Translator's Workbench and, (5) in order to secure ethics approval for research involving humans, participants had to volunteer. Using these criteria, the number of qualifying participants out of a group of 20 final year and MA-level translation students was eight, but this was then reduced to five following elimination of participants due to poor data results.[3] We discuss participant training, selection and elimination below under Methodology. We acknowledge that five participants is a small number, and it prevents us from making any general claims regarding the mental effort involved in processing fuzzy matches. In addition, it is a limitation of translation process studies in general, and eye-tracking studies specifically, that large amounts of data are generated whose interpretation is very time-consuming. This, along with the need for equal competence, reduces the number of participants such that we can generate and refine hypotheses here, but we cannot make general claims without further validation with similar participants.

2 Eye-tracking data are lost when a participant looks away from the eye tracker.

3 Here we refer to the fact that eye-tracking studies inevitably have to eliminate some participants due to technical or participant-specific problems when tracking eye movements.

3. Methodology

The source language in this study was German and the target language English. Source and parallel target files were located on the *eCoLoRe* resources web site.[4] The originator of these files was the company SAP and the content was technical and related to a computer application. It was necessary to create a translation memory by aligning the source and target files using SDL Trados WinAlign. The motivation for selecting the SDL Trados suite of tools was twofold: firstly, SDL Trados Translator's Workbench is one of the most commonly used TM tools (Lagoudaki 2006: 24); secondly, it is also the tool that the author and study participants were most familiar with.

Each participant was either in his/her final year of an undergraduate degree in Translation Studies or in his/her first year of an MA in Translation Studies at Dublin City University. Each had taken a module in technical translation from German to English at either final-year BA level (participant A) or post-graduate diploma level (participants C, D, G, H).[5] The participants had also all successfully completed a module on Translation Technology, a significant component of which involves the use of the TM environment used in this research. One of the assessments for this module involves detailed interaction with the tool. The participants were recruited on the basis of having achieved more than a minimum mark in the technical translation and technology modules (set at 60% for the study). This was used as an indicator of relatively good translation and translation technology competence, but their competence was not tested beyond this. Following this selection procedure, only eight participants were available. Checks were performed following a pilot run of the experiment (using different text) to ensure that the participants were indeed suitable for eye tracking. Since everything appeared to be working fine, the eight participants proceeded with the main task. However, it was later noted that the eye movements for three participants were not adequately tracked and the data for three participants had to be removed from the study, leaving us with five participants (A, C, D, G and H). It is known that

4 http://ecolore.leeds.ac.uk/ [Last accessed: 05/07/08.] eCoLoRe provides shareable resources for eContent and computer-aided translation training.

5 Participant identifiers are non-sequential due to the elimination of three participants.

a level of participant attrition exists for eye-tracking studies. Problematic data can be caused by various factors such as very dark irises, poor study environment, relatively small eyes, dirt on the lenses of glasses, or just participant behaviour on the day. Data problems do not always present themselves during pilot studies and are sometimes only apparent when it comes to analysing the data. All participants were female. However, significantly, it has been observed that there are no gender differences in pupil response to mental workload (Beatty 1982).

The source text that was used during alignment was edited so that matches with different fuzzy values would be encountered by the participants while translating. The participants had 25 segments to translate (348 words in total – see Appendix A for detailed information on the source text and the fuzzy matches). The source text is admittedly quite short. However, given the amount of data generated by eye-tracking studies and the size of the video files that can also be generated, we did not wish to have an overly long source text. In addition, we did not want fatigue to act as a factor in the measurement of cognitive load. Table 1 shows the fuzzy match values and the number of segments for each:

Table 1. Number of fuzzy match types contained in the source text

Fuzzy Match Value	*Number of Segments*
99%	1
94%	1
93%	1
91%	1
90%	1
89%	1
88%	3
81%	2
76%	1
75%	2
73%	1
69%	1
65%	2
62%	2
59%	1
58%	1
54%	1
52%	2

The fuzzy match values were deliberately distributed between the low 50% value and 99% so that we could investigate the potential differences in cognitive load between high, medium and low fuzzy values. The actual fuzzy match values assigned were beyond the control of the researcher as information on the algorithm that calculates the match value is not publicly available. In summary, there were five segments between the values of 90 and 99%, six between 81 and 89%, four between 73 and 76%, five between 62 and 69% and five between 52 and 58%. It was felt that this represented a reasonable spread across match values.

Pupil dilation is acknowledged as a good measurement of cognitive load (see, for example, Hess and Polt 1964; Marshall *et al.* 2002). However, since pupils are sensitive to various stimuli (light, audio, emotion, etc.), care must be taken to conduct studies in a controlled environment. The recording of eye movements for this study was carried out on an individual basis in a quiet research lab using the Tobii 1750 eye tracker and the ClearView analysis software (version 2.6.3). The lab is equipped with black-out blinds, and an overhead fluorescent light was used in an attempt to keep the light levels in the room consistent so that pupils were not dilating or constricting due to sudden changes in light or audio stimuli. These efforts do not represent a very high level of control on ambient influences, but they do aim to take into account and reduce possible interferences in some of the measures we were interested in.

As mentioned previously, the participants were deemed to be reasonably competent in both technical translation and the TM tool. However, prior to the task being set, they were given some time to translate a different portion of the same text in the TM environment in order to make sure that they were comfortable with the task being set. This was presented to them as a 'warm-up task'.

Participants were specifically asked not to research terminology on the Web or to use hard-copy terminological resources as this would have impacted on the processing speed for any match. In addition, no glossary was provided. Although terminology research is an integral part of the translation process, we felt it necessary to eliminate this step so we could perform a more direct comparison of processing effort between fuzzy match categories. It is true to say that translators feel uncomfortable when this integral step in the translation process is removed. In Immonen's (2006) study of translation as a writing process, the majority of participants commented that they would have used a dictionary or other sources of

information, if they had been permitted. For them it was not a matter of knowing the meaning of a word but rather of "getting more depth of understanding and finding synonyms from the dictionaries" (2006: 319). Nonetheless, we also felt that the elimination of terminology look-up during the current study would have only limited impact because many of the TM matches already provided sound parallel terms for the participants since the TM had been created from a ST and TT that were approved by the originator, SAP.[6]

The importance of providing participants with a realistic translation brief during translation process research has been acknowledged by Krings (2001: 75). Participants were given the brief of translating the text as if it were a commercial translation task. They were told not to talk to the researcher during the task, unless they encountered a serious technical difficulty which they could not resolve by themselves. The researcher sat away from the participant with her back turned in order to put the participant at her ease. All participants completed the task without having to speak to the researcher.

Once the task was completed, the participants were presented with a paper-based survey. The survey presented the same source text segments and fuzzy matches that they had just seen on screen. The main differences were that they did not see the fuzzy match value that was assigned by the TM system to the match, nor were any differences between the 'new' ST and the ST in the TM highlighted. Participants were asked to rate their "perceived editing effort" for each match on the basis of 1-5 where "1" indicated "Read-only, no edits" and 5 indicated "a lot of editing effort". The survey is, of course, subjective, but the aim was to see if there were any correlations between the quantitative data on cognitive effort accumulated in the study and qualitative data on participants' perceived effort for match types. The results of the survey are discussed in more detail below.

Following the translation task, quantitative data on (1) processing speed per match value and (2) pupil dilation were analysed. It is assumed here that processing speed is a good measurement for cognitive effort on the basis that difficult tasks generally take longer than easier tasks. And, as

[6] We assume implicit approval here since the ST and TT were placed on a public web site and labelled as resources for training translators.

already mentioned, it has long been established that pupil dilation is also a significant indicator of cognitive effort. The processing speed was calculated with the help of the play-back feature in ClearView which shows precisely when a participant opens a new segment and moves on to the next segment by using the appropriate buttons on the Translator's Workbench Toolbar. ClearView also allows for the export of pupil dilation data, captured on a millisecond basis, along with precise timestamp data. This in turn allows for median pupil dilations to be calculated across all participants for each match type. The results for both of these measures are discussed below.

4. Results

4.1 Processing speed

The processing speed (expressed in words per second) for each participant and each segment was calculated by dividing the number of words in the source segment by the number of seconds the participant spent on that segment. The start was calculated from the moment the participant clicked on the "OpenGet" button in the TM toolbar, which searches in the TM for a match, to the moment they clicked the "SetCloseOpenGet" button, which moves on to the next segment. The median processing speed was then calculated for all participants.[7] Our data provide clear evidence that the processing speed decreases as the fuzzy match value decreases, i.e. the lower match values required more time than the higher match values (see Figure 1). This suggests that the lower match values require more cognitive effort than the higher match values because, in theory, more work is required to 'fix' fuzzy matches of 50% similarity than those of 90% similarity, for example. We note, however, that the difference in processing speeds between the 60-69% match class and the 50-59% class is small (0.20 vs. 0.24 respectively). This is significant when we come to discuss the median pupil dilation.

[7] We use median instead of mean because we want to know the exact midpoint of distributions and because individual differences could lead to some extreme scores that will distort the mean.

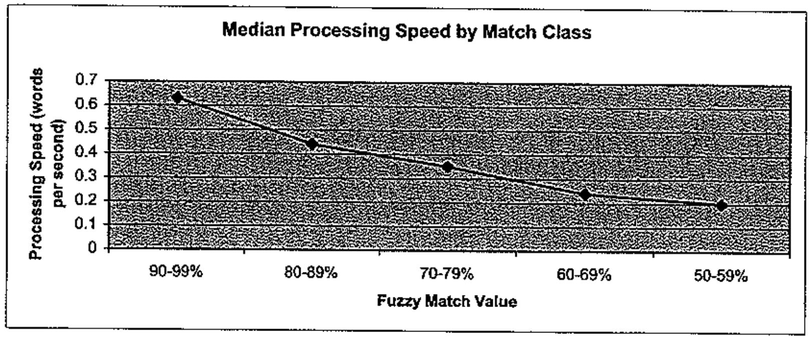

Figure 1. Median processing speed by match class

4.2 Pupil dilation

The pupil dilation measurement was captured by ClearView for all segments and the median dilation for each segment was calculated for all participants. A somewhat surprising finding for pupil dilation measurement was that, when seen as a median measurement across all participants, dilations increased as match value decreased until the 60-69% match class was reached and then a decrease in median pupil dilation was noted, as illustrated in Figure 2.

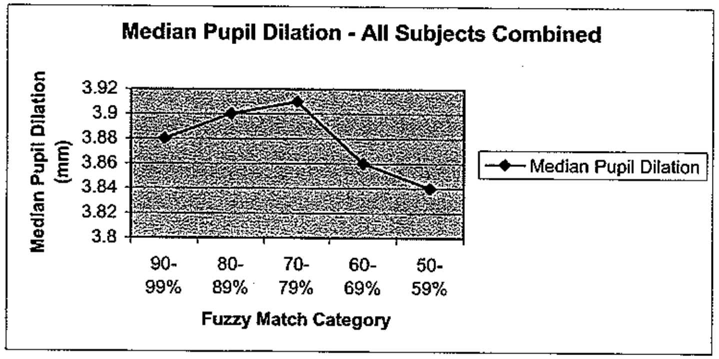

Figure 2. Median pupil dilation

While there was some variation between participants, this finding is similar to the phenomenon observed in O'Brien (2006) whereby cognitive effort, as expressed by pupil dilations, appeared to decrease around the mid-70% fuzzy match category. When we view the results of pupil dilation and processing speed together, we get a very clear picture of the anomalies – Figure 3.

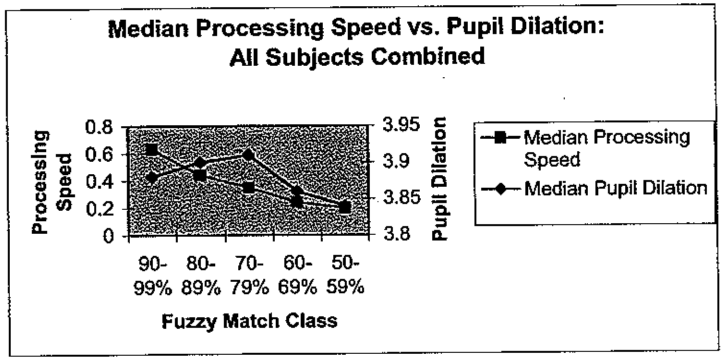

Figure 3. Processing speed and pupil dilation combined

While processing speed clearly decreases as fuzzy match value decreases, pupil dilation increases and then decreases once the match value dips below 70%. In other words, pupil dilations follow an upward trend until the 60-69% match value and then the two measurements appear to converge.

We can offer a potential explanation for this but, due to the relatively small sample size, we can only hypothesize about what might cause this result. We mentioned previously that the difference in processing speed between the 60-69% and 50-59% match values was small. Can we assume that once a translator is presented with a match value in or below the 60-69% range, a baseline processing speed has been reached? In other words, as match values decrease we cannot expect processing speed to decrease exponentially once we have exceeded a specific match boundary. There is anecdotal evidence to suggest that when translators are faced with a 'low' fuzzy match, they intuitively know that it would be 'faster' to translate the ST from scratch and, therefore, they do not waste time trying to figure out what piece of the ST matches with the proposed TT. The plateauing of the processing speed graph observed here might provide evidence of that.

Pupil dilations grow as the match value decreases up to the match class of 60-69%. The initial growth trend is in keeping with the link established between pupil dilations and mental workload by, for example, Hess and Polt (1964). However, the reduction in pupil dilation for the 60-69% and 50-59% match classes is surprising. We can again hypothesize

that the translating participants have perhaps reached a baseline cognitive effort when they reach the 60-69% match class. We can pose the question: is it the case that translators are not concerned with piecing fuzzy segments together for low fuzzy matches but would rather translate from scratch?

We can of course look deeper into this phenomenon by examining exactly what participants *do* when faced with fuzzy matches in the lower values. This is facilitated by using ClearView to play back the video file generated during task processing. A full analysis of all sub-processes is beyond the scope of this paper. However, we can comment in general on our findings regarding translation behaviour for lower fuzzy match values. Our data show that all participants process lower fuzzy matches in apparently the same manner as higher scoring fuzzy matches. There is no evidence of a decision by participants that the fuzzy match value is "too low" to warrant comparison with the ST. If such a decision were made, one could expect that the translation offered by the Translation Memory might be completely discarded in the form of a blanket deletion from the TT window. Instead, there is evidence that the participants read the ST and compare it with the TT. They also compare with the "Old Source Text" displayed in the TM Window. They rarely look at the fuzzy match value in the top left-hand corner of the TM Window to inform themselves about the actual value of the fuzzy match (see below for further discussion of this). Having compared the ST and TT, they then delete unwanted text, add any new text that might be required, and move on to the next match. Therefore, we believe that the reduction in pupil dilation for the 60-69% and 50-59% match classes cannot be attributed to an explicit change in translation strategy by participants. We therefore have to look elsewhere for a possible explanation for the plateauing of pupil dilation.

The bottoming out of cognitive load reflects the observation in O'Brien (2006) that cognitive load does not necessarily continue to increase as fuzzy match value decreases. This finding echoes findings in neuroimaging studies where brain activity was monitored during the performance of tasks with varying task difficulty (Gould *et al.* 2003; Callicott *et al.* 1999). In Gould *et al.* (2003), an "inverted-U" or "capacity-constrained" response was recorded, i.e. once a certain level of cognitive load is attained in the brain, a plateauing or even diminishing response to task difficulty is recorded. Similarly, in their comparison of cognitive load in the online reading of easy and difficult texts, Schultheis and Jameson (2004) found that there were no significant differences in mean pupil diameter for easy and difficult texts whereas other tests of cognitive load conducted simultaneously (reading speed, subjective ratings, and P300[8] amplitude) showed positive correlations with task difficulty. Callicott *et al.* (1999) discuss the physiological reactions of different parts of the brain to increasing mental workload. While some parts of the brain seem to become more engaged with increasing workload, other parts appear to disengage after certain levels of mental workload are reached and this disengagement could be attributed to participants being overwhelmed by the increasing task difficulty. While the theory of "disengagement" could be used to explain the decrease in pupil dilation for lower level fuzzy matches, the evidence from the playback in ClearView, as already discussed, strongly suggests that participants did not disengage from the task at hand. Iqbal *et al.* (2005) used pupil size to demonstrate that different subtasks in a task hierarchy demand different levels of mental workload. Their conclusions were that mental workload decreases more at task boundaries higher up in the task model and less at task boundaries lower in the model. They also demonstrated that there are changes in mental workload in subtasks located at the same level in a task hierarchy, leading them to suggest that "effective understanding of why changes in mental workload occur requires that the measure be tightly coupled to a validated task model" (2005: 318). The current study did not involve the description of a detailed task model. Rather, tasks were defined according to the level of fuzzy match being processed. A more detailed analysis of mental workload according to a detailed task model would, of course, present an interesting research challenge in the future.

4.3 Survey results

As mentioned, immediately following the translation task, participants were asked to fill out a survey in order to measure their *perceptions* of editing effort for each match. They were presented with the matches on paper in the order they appeared in the task with the source text first and

[8] The P300 is a specific wave emitted by the brain when it recognises an object, sound etc.

the target text underneath that. In order to aid recall, the source text and target text were presented with the same formatting used during the translation task. Before completion of the survey, it was explained that they had to select one option only out of five that best represented their perceptions of the level of effort required for each segment. The effort was rated on a scale of 1-5 where 1 represented "read only, no edits", 2 "little editing effort", 3 "some editing effort", 4 "significant editing effort" and 5 "a lot of editing effort". It is acknowledged here that perceptions of the difference between "little" and "some" or "significant" and "a lot" will vary from one individual to the next. Rather than specify how each one should be interpreted, the interpretation was left up to each participant since the results of the survey were to be analysed in conjunction with the more quantitative data of processing speed and pupil dilation.

The survey results were collated for each participant and the average score for each segment was calculated. Segments were then classified into match categories, as was the case for processing speed and pupil dilation. Figure 4 shows the results:

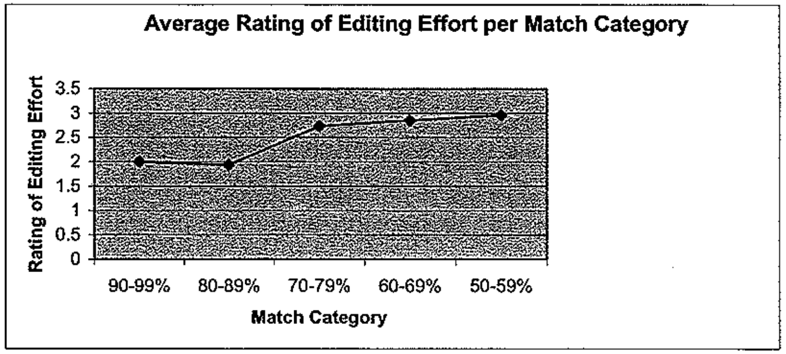

Figure 4. Average rating of editing effort

The results suggest that, on average, participants rated matches between 80 and 99% as requiring "little editing effort". For the match category of 70-79%, the perceived editing effort then moved closer to 3 (2.73) "some editing effort" and then rose to 2.85 for the category 60-69%. Finally, the match category of 50-59% was rated as closest to 3 at 2.96. Although some individual segments scored 4 (e.g. Participant A rated segments 20, 21 and 25 as a 4), no segment was given a rating of 5 and, when averaged, no segment went above a value of 2.96.

These results show that segments of between 80 and 99% are rated as requiring little effort while anything between 50 and 79% is seen to require more editing effort, but, even with a fuzzy match value of 52%, segments were not rated as requiring "a lot of editing effort". Interestingly, we again see the plateauing effect of the graph here for the 60-69% and 50-59% match categories which was also visible in our Processing Speed graph (Figure 1).

4.4 Influence of text type?

It is important to acknowledge that the nature of the ST domain and of the edits made to the new ST may have had an influence on these results. The ST, as mentioned under Methodology is from the IT domain and the text type is that of a user manual. Such texts are frequently updated when new versions of the software are released, but chunks of text also remain unchanged. Often, the changes made from one version to another involve substitutions of version numbers and product names. In our test data, Segment 6 is a typical example of the types of changes made:

Old source text:
UNIX®, X/Open®, OSF/1® und Motif® sind eingetragene Marken der Open Group.

New source text:
UNIX® und Motif® sind eingetragene Marken der Open Group.

As can be seen, the brand names "X/Open" and "OSF/1" have both been deleted from the new source text. Segment 6 is given the match value of 65% by the Translator's Workbench. Despite the relatively low value, the translator simply has to select and delete these brand names from the target text proposed by the TM. Perhaps this explains why cognitive load, as expressed in median pupil dilation, actually plateaus at these values: the text is quite different from a fuzzy match value point of view, but the edit is not very taxing on the translator. At the same time, while this is a relatively simple edit, we must not underestimate the cognitive effort involved in first of all comparing the old and new source text and secondly in perceiving what exactly the differences are and how they should be dealt with (we discuss this in more detail later). TM tools in general highlight such

differences to the translator using coloured coding, and Translator's Workbench is no exception to this (see Figure 5 below). However, the usefulness of such colour coding in TM tools has not yet been tested empirically, to the best of our knowledge.

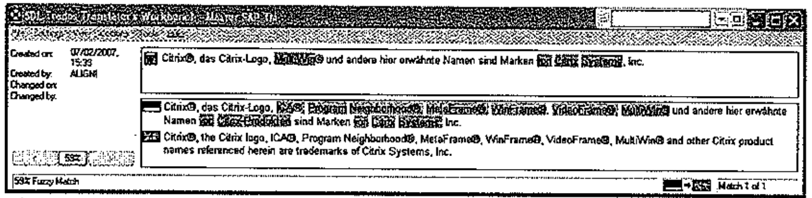
Figure 5. Example of highlighting for fuzzy matching in Translator's Workbench

It is interesting to note that the participants rated Segment 6, on average, as requiring "little post-editing effort" (2.2 average, with one participant rating it as "3" and all others as a "2"), despite its 65% fuzzy match value. In comparison, Segment 12, rated as 94% match value by Translator's Workbench, scored an average of 2.6 (three participants rated it as requiring "little editing effort", one as "some editing effort" and another as "significant editing effort"). As can be seen below, only the text "SAP Logo" had to be deleted from the new TT:

Old source text:
*SAP, **SAP Logo**, R/2, R/3, mySAP, mySAP.com und weitere im Text erwähnte SAP-Produkte und Dienstleistungen so wie die entsprenchenden Logos sind Marken oder eingetragene Marken der SAP AG in Deutschland und anderen Ländern weltweit.*

New source text:
SAP, R/2, R/3, mySAP, mySAP.com und weitere im Text erwähnte SAP-Produkte und Dienstleistungen so wie die entsprenchenden Logos sind Marken oder eingetragene Marken der SAP AG in Deutschland und anderen Ländern weltweit.

Target text proposed by TM:
*SAP, **SAP Logo**, R/2, R/3, mySAP, mySAP.com and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries all over the world.*

Thus, while the average ratings of perceived editing effort by participants provide us with unsurprising results, sometimes the match values accorded by the TM tool do not correlate well with the translators' perceived editing effort. This adds further to the suggestion that a more complex task hierarchy is perhaps involved in processing these matches and we cannot measure effort simply by using the fuzzy match value. This leads us to a question that is somewhat separate from, but related to, the question at the core of this study. What cues do translators use when interacting with the TM tool to determine the level of cognitive effort required?

4.5 *Relevance of the fuzzy match value?*

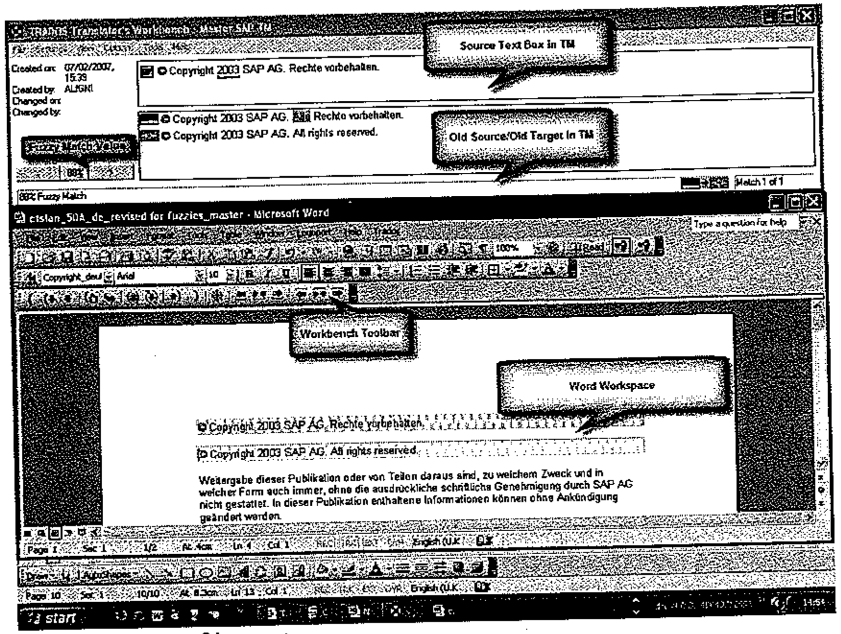

Figure 6. Areas of interest

In Figure 6, a screenshot of the Translator's Workbench representation of the New ST, Old ST and TT is presented and the value of the fuzzy match is given in the left-hand corner of this screenshot (88% in the example given). We saw from the survey results that sometimes the perceived editing effort does not always correlate with the fuzzy match value allocated by the translation memory system. Therefore, one of the

questions of interest in this research was: to what extent does the translator take the allocated match value into account? One way of estimating the extent to which the participants took the match value into account is by looking at the number of fixations (Fixation Count) and the length of time spent fixating on the fuzzy match value (Fixation Duration).[9] In order to calculate this, different parts of the screen were first defined as "Areas of Interest" or "AOIs" in ClearView. These were Source Text Box in TM, Old Source/Old Target in TM, Fuzzy Match Value, Workbench Toolbar and Word Workspace (see Figure 6).

Table 2 shows the average Fixation Counts and Fixation Duration for all AOIs:

Table 2. Fixation Counts/Duration for AOIs

AOI	*Average Fixation Count*	*Average Fixation Duration (in milliseconds)*
Source Text Box in TM	410.60	220.92
Old Source/Old Target in TM	395.60	215.78
Fuzzy Match Value	25.60	254.43
Workbench Toolbar	79.20	255.33
Word Workspace	1354.40	228.15

We can see that the total number of fixations was lowest for the Fuzzy Match Value AOI. This was followed in second place by the Workbench Toolbar which suggests that participants do not actually pay much attention to the value attributed by the TM system to the match. Given that the fixation count for the Old Source/Old Target AOI is much higher than the fuzzy match value, we can hypothesize that translators prefer to look at the match in the TM to process *differences* between the new source text and the old source text. This is borne out when replaying the eye movement videos in ClearView which demonstrate that the participants' eyes moved frequently between the "Word Workspace" and the "Source Text Box in TM" and "Old Source/Old Target in TM" AOIs. The participants seem to use the cues in these AOIs to process fuzzy matches and, in fact, one might suggest that the presentation of effectively the same data on different parts of the screen acts as "visual noise" to some extent, distracting the translator from the core task. Although participants do not look at the fuzzy match

9 Fixation count is the number of fixations on each defined Area of Interest (AOI) and fixation duration is the total length of time (in ms) spent fixating on that AOI.

value AOI often, we should note that when they do look at it, they fixate on it for a similar amount of time as on the other AOIs (mean=234.92).

5. Conclusions and future directions

We have seen that the relationship between cognitive load and fuzzy match value is not straightforward. If we rely on processing speed alone, we may well concur with translators' opinions that decreasing fuzzy matches mean increasing effort and, therefore, payment should be scaled accordingly. Even so, there is some evidence of a non-linear relationship between processing speed and fuzzy match value and this clouds the economic argument somewhat. We have also seen that when using pupil dilation as a measure of cognitive load, the picture is not so clear. We could simply dismiss pupillometry as a method, but that would be unwise given that it has been shown to be an accurate measure of cognitive load in many other disciplines and over relatively long periods of time. While pupillometry may not be an exact methodology, it still has some merit. The data we have recorded here may represent further evidence of a capacity-constrained response to task difficulty. It may also reflect the findings from other studies that while some measures of cognitive load demonstrate linear relationships between increasing task difficulty and the measure itself, the measurement of pupil dilation appears to hit a ceiling at a certain point during task execution. Also, it appears that cognitive load is linked in a complex way to task hierarchy and more extensive studies would be required in order to investigate this further.

In this study, translators did not treat lower value fuzzy matches differently from higher value matches, though admittedly this behaviour might change in industrial settings according to economic motivation. A more fine-grained combination of a complex task hierarchy with measures of cognitive load would help to investigate the strategies for higher and lower fuzzy matches and is one possible direction for future research.

The data also demonstrate that translators do not make extensive use of the Fuzzy Match value presented by the TM system in this study, but are drawn to looking at and comparing text as a means of establishing differences between old and new segments. This raises questions about how much text a translator needs to see on the screen and whether or not

colour coding of differences is useful or simply represents visual "noise"? This is an area that is quite worthy of further investigation since, as Lagoudaki (2006) points out, few, if any, usability studies are carried out on TM tools prior to their launch on the market. Of course, the possibility also exists that more experienced translators might behave in a different way and this is another potential avenue of investigation.

We have also pointed to the disconnection between the fuzzy match value and the translators' perceived editing effort in some instances. Text type, domain and/or length of segment could play a role here. A comparison of cognitive load and behaviour in different domains/text types and with different segment lengths might also be worthy of further investigation.

Finally, we must acknowledge here the possibility that the results might change if the number of participants were increased tenfold. However, given the time-consuming nature of this type of research and the difficulties in acquiring appropriate and equally competent participants who can touch-type, we will have to make do with a small sample size for the present. It is our intention, however, to test our hypotheses further by expanding the number of participants in the study.

References

Beatty, J. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91 (2): 276-292.

Beninatto, R. & DePalma, D. 2004. *What Translation Costs*. Common Sense Advisory. Available online to members only at http://www.commonsenseadvisory.com/research/reports.php (Last accessed 30/07/08.)

Bierman, B. 2008. Everything's comin' up roses (with apologies to Stephen Sondheim), *Translation Journal* 12 (3), available online at http://accurapid.com/journal/45compensation.htm (Last accessed: 30/07/08.)

Callicott, J., Mattay, V., Bertolino, A., Finn, K., Coppola, R., Frank, J., Goldberg, T. & Weinberger, W. 1999. Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex* 9: 20-26.

De Palma, D., Sargent, B., Bassetti, T. & Beninatto, R. 2008. *The Price of Translation*. Common Sense Advisory. Available online to members only at http://www.commonsenseadvisory.com/research/reports.php (Last accessed 30/07/08.)

Gould, R., Brown, R., Owen, A., Ffytche, D. & Howard, R. 2003. fMRI BOLD response to task difficulty during successful paired associates learning. *NeuroImage* 20: 1006-1019.

Hess, E. & Polt, J. 1964. Pupil size in relation to mental activity in simple problem solving. *Science* 143: 1190-1192.

Immonen, S. 2006. Translation as a writing process: pauses in translation versus monolingual text production. *Target* 18 (2): 313-335.

Iqbal, S., Adamczyk, P., Zheng, X. & Bailey, P. 2005. Towards an index of opportunity: understanding changes in mental workload during task execution. *Human Factors in Computing Systems: Proceedings of CHI'05*. New York: ACM Press. 311-320.

Jääskeläinen, R. 2002. Think-aloud protocol studies into translation – An annotated bibliography. *Target* 14 (1): 107-136.

Krings, H.P. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. Kent, Ohio: The Kent State University Press, edited/translated by G.S. Koby.

Lagoudaki, E. 2006. *Translation Memories Survey 2006*. Available online at: http://www3.imperial.ac.uk/pls/portallive/docs/1/7307707.PDF (Last accessed 30/07/08.)

Marshall, S., Pleydell-Pearce, C. & Dickson, B.T. 2002. Integrating psycho-physiological measures of cognitive workload and eye movements to detect strategy shifts. *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*: 130b-135b.

O'Brien, S. 2006. Eye-tracking and Translation Memory matches. *Perspectives: Studies in Translatology* 14 (3): 185-205.

Sargent, B. 2008. *Through the Looking Glass of Translation Pricing*. Common Sense Advisory. Available online to members only at http://www.-commonsenseadvisory.com/research/reports.php (Last accessed 30/07/08.)

Schultheis, H. & Jameson, A. 2004. Assessing cognitive load in adaptive Hypermedia Systems: physiological and behavioural methods. In *Adaptive Hypermedia and Adaptive Web-Based Systems*, Lecture Notes in Computer Science Vol. 3137, Berlin-Heidelberg: Springer: 225-234.

Van der Meer, J. 2008. *Localization Business Innovation – Translation in Transition: From Cost to Opportunity, to Strategy, and Now to Utility.* TAUS White Paper,. June 2008. Available online (subscribers only): http://www.translationautomation.com/joomla/ (Last Accessed: 30/07/08.)

Appendix A: Source and target segments

Note: The match value is given in between the ST and TT segments

Segment 1

{0>© Copyright 2003 SAP AG. Rechte vorbehalten.

<}88{>

© Copyright 2003 SAP AG. All rights reserved.

<0}

Segment 2

{0>Weitergabe dieser Publikation oder von Teilen daraus sind, zu welchem Zweck und in welcher Form auch immer, ohne die ausdrückliche schriftliche Genehmigung durch SAP AG nicht gestattet.

<}93{>

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG.

<0}

Segment 3

{0>In dieser Publikation enthaltene Informationen können ohne Ankündigung geändert werden.

<}90{>

The information contained herein may be changed without prior notice.

<0}

Segment 4

{0>Die von SAP AG angebotenen Softwareprodukte können Softwarekomponenten auch anderer Softwarehersteller enthalten.

<}76{>

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors.

<0}

Segment 5

{0>Microsoft®, WINDOWS®, NT®, EXCEL®, sind eingetragene Marken der Microsoft Corporation.

<}73{>

Microsoft®, WINDOWS®, NT®, EXCEL®, Word®, PowerPoint® and SQL Server® are registered trademarks of Microsoft Corporation.

<0}

Segment 6

{0>UNIX® und Motif® sind eingetragene Marken der Open Group.

<}65{>

UNIX®, X/Open®, OSF/1®, and Motif® are registered trademarks of the Open Group.

<0}

Segment 7

{0>Citrix®, das Citrix-Logo, MultiWin® und andere hier erwähnte Namen sind Marken von Citrix Systems, Inc.

<}59{>

Citrix®, the Citrix logo, ICA®, Program Neighborhood®, MetaFrame®, WinFrame®, VideoFrame®, MultiWin® and other Citrix product names referenced herein are trademarks of Citrix Systems, Inc.

<0}

Segment 8

{0>HTML, XML sind Marken oder eingetragene Marken des W3C®, World Wide Web Consortium, Massachusetts Institute of Technology.

<}88{>

HTML, DHTML, XML, XHTML are trademarks or registered trademarks of W3C®, World Wide Web Consortium, Massachusetts Institute of Technology.

<0}

Segment 9

{0>JAVA® ist eine eingetragene Marke der Sun Microsystems.

<}88{>

JAVA® is a registered trademark of Sun Microsystems, Inc.

<0}

Segment 10

{0>JAVASCRIPT® ist eine eingetragene Marke der Sun Microsystems, Inc., verwendet unter der Lizenz der von Netscape entwickelten Technologie.

<}91{>

JAVASCRIPT® is a registered trademark of Sun Microsystems, Inc., used under license for technology invented and implemented by Netscape.

<0}

Segment 11

{0>Enterprise Buyer ist eine gemeinsame Marke von SAP AG und Commerce One.

<}65{>

MarketSet and Enterprise Buyer are jointly owned trademarks of SAP AG and Commerce One.

<0}

Segment 12

{0>SAP, R/2, R/3, mySAP, mySAP.com und weitere im Text erwähnte SAP-Produkte und -Dienst-leistungen sowie die entsprechenden Logos sind Marken oder eingetragene Marken der SAP AG in Deutschland und anderen Ländern weltweit.

<}94{>

SAP, SAP Logo, R/2, R/3, mySAP, mySAP.com and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries all over the world.

<0}

Segment 13

{0>Alle anderen Namen von Produkten sind Marken der jeweiligen Firmen.

<}81{>

All other product and service names mentioned are trademarks of their respective companies.

<0}

Segment 14

{0>Sprachentransport (BC-LAN)

<}81{>

Language Transport (BC-CTS-LAN)

<0}

Segment 15

{0>Diese Dokumentation beschreibt wie Sie Sprachen in ein SAP-System importieren.

<}99{> (missing comma)

This documentation tells you how to import languages into an SAP System.

<0}

Segment 16

{0>Beachten Sie, dass diese Dokumentation nur in den Sprachen Deutsch und Französisch zur Verfügung steht.

<}89{>

This documentation is only available in English and German.

<0}

Segment 17

{0>

In dieses Release sind die Funktionen zum Importieren von Sprachenanteilen von Support Packages in die Transaktion SMLT integriert.

<}75{>

In Release 6.10, functions for importing the language content of Support Packages, and for client maintenance, have been integrated into transaction SMLT.

<0}

Segment 18

{0>Importvorgänge werden jetzt in der SMLT Komponent angezeigt.

<}58{>

Transaction SMLT now displays exports.

<0}

Segment 19

{0>In der SMLT Komponent finden Sie unter *Tools* alles für den Sprachentransport.

<}62{>

All language transport tools are now available in transaction SMLT by choosing *Goto □Other Tools*.

<0}

Segment 20

{0>Von hier aus können Sie die Werkzeuge starten - bisher mußten Sie diese Werkzeuge separat aufrufen.

<}54{>

You can start the tools from here.

<0}

Segment 21

{0>Weitere Entwicklungen wurden vorgenommen, um die SMLT besser zu machen.

<}62{>

Other improvements have been made to make transaction SMLT easier to use.

<0}

Segment 22

{0>Ab dieses Release können Sie alles parallel starten.

<}52{>

As of Basis Release 4.6D, you can start transport processes in parallel.

<0}

Segment 23

{0>Das gilt sowohl für den Import als auch für den Export und damit kann der Import einfacher ausgeführt werden.

<}69{>
This applies to both the language import and the language export, and makes the import or export faster.

<0}

Segment 24

{0>Es wurden Änderungen an der Transporttechnologie vorgenommen mit dem Ziel, die Performance und die Durchführbarkeit des Sprachentransportes weiter zu verbessern.

<}75{>
Changes have been made to the transport technology and the user interface that improve both the performance and usability of the language transport tools.

<0}

Segment 25

{0>Zu Basis-Release 4.6C wurde der Sprachentransport komplett umgestellt, weil Sprachen nun auch mit dem SAP-Transportprogramm R3trans transportiert werden.

<}52{>
The language transport function has been changed completely in Basis Release 4.6C.

Eye movement behaviour across four different types of reading task

Arnt Lykke Jakobsen and Kristian T.H. Jensen

Abstract

A group of six professional translators and a group of six translation students read four similar texts on the same news topic while their eye movements were tracked. The first two texts were read with different reading purposes, (a) for comprehension and (b) with the intention of translating the text afterwards. Texts three and four were read while being simultaneously (c) translated orally and (d) translated in writing. It was found that professionals were faster than students. For both groups, task time, fixation frequency, gaze time and average fixation duration showed a consistent, linear progression from task to task. In the final task it was shown that the distribution of visual attention to the source text for students was higher than that for the target text, whereas professional translators prioritised visual attention to their own target text.

1. Background

Eye movements in reading have been studied intensively for decades (see e.g. Just & Carpenter 1980; Rayner & Pollatsek 1989; Rayner 1998; Hyönä *et al.* 2003; Radach *et al.* 2004) and several basic facts about eye movements in reading have been convincingly documented. We now know the typical duration and length of saccades and the typical duration of fixations, and we know that such factors as word familiarity (Williams & Morris 2004), word predictability (Frisson *et al.* 1999), word length and complexity (Kliegl *et al.* 2004; Bertram & Hyönä 2003; Rayner & Duffy 1986), lexical and/or syntactic ambiguity (Juhasz & Rayner 2003) all affect fixation duration. What we still know little about is how reading varies