Translation Methods and Experience: A Comparative Analysis of Human Translation and Post-editing with Students and Professional Translators

JOKE DAEMS
Ghent University, Ghent, Belgium
joke.daems@ugent.be

SONIA VANDEPITTE
Ghent University, Ghent, Belgium
sonia.vandepitte@ugent.be

ROBERT J. HARTSUIKER
Ghent University, Ghent, Belgium
robert.hartsuiker@ugent.be

LIEVE MACKEN
Ghent University, Ghent, Belgium
lieve.macken@ugent.be

RÉSUMÉ

Bien que l'on ait une certaine connaissance des avantages de la post-édition pour les textes techniques, il reste difficile de savoir si celle-ci est une solution de remplacement viable à la traduction humaine pour les textes généraux. Il faut par ailleurs comprendre davantage les deux méthodes de traduction, et la façon dont elles sont appliquées par les étudiants et les traducteurs professionnels, pour en déterminer les écueils et adapter la formation des traducteurs en conséquence. Dans cet article, nous cherchons à mieux cerner les différences entre la traduction humaine et la post-édition pour les articles de journaux. L'oculométrie et l'enregistrement de frappes ont été utilisés pour étudier le processus de traduction et dégager les gains en productivité, la charge cognitive et l'utilisation de ressources externes. Nous avons également examiné la qualité de la traduction et l'attitude des traducteurs à l'égard des deux méthodes de traduction. L'étude de ces différentes perspectives met en évidence que les deux groupes de traducteurs comme les deux méthodes de traduction sont plus proches que ce que l'on pouvait présumer.

ABSTRACT

While the benefits of using post-editing for technical texts have been more or less acknowledged, it remains unclear whether post-editing is a viable alternative to human translation for more general text types. In addition, we need a better understanding of both translation methods and how they are performed by students as well as professionals, so that pitfalls can be determined and translator training can be adapted accordingly. In this article, we aim to get a better understanding of the differences between human translation and post-editing for newspaper articles. Processes are registered by means of eye tracking and keystroke logging, which allows us to study translation speed, cognitive load, and the use of external resources. We also look at the final quality of the product as well as translators' attitude towards both methods of translation. Studying these different aspects shows that both methods and groups are more similar than anticipated.

RESUMEN

Mientras que se reconocen cada vez más los beneficios del uso de la post-edición para la traducción de textos técnicos, aún no existen pruebas de que la post-edición sea una alternativa viable a la traducción humana para los textos más generales. Además, necesitamos una mejor comprensión de ambos métodos de traducción, así como de la forma en que son realizados por estudiantes y por profesionales, para que se puedan definir los posibles obstáculos y adaptar la formación de los traductores en función de ellos. En este artículo, nuestro objetivo es lograr una mejor comprensión de las diferencias entre la traducción humana y la post-edición de artículos periodísticos. Los procesos de traducción se registran mediante el seguimiento ocular y el registro de pulsaciones de teclas, lo que nos permite estudiar la velocidad de traducción, la carga cognitiva y el uso de recursos externos. También analizamos la calidad final del producto, así como la actitud de los traductores hacia ambos métodos de traducción. El estudio de estos diferentes aspectos demuestra que tanto los métodos como los grupos de traductores son más similares de lo previsto**.**

MOTS-CLÉS/KEYWORDS/PALABRAS CLAVE

traduction, post-édition, expérience, processus de traduction, qualité de traduction
translation, post-editing, experience, translation process, translation quality
traducción, post-edición, experiencia, proceso de traducción, calidad de traducción

1. Introduction

Because of the current level of globalization, traditional human translation cannot keep up with translation needs (Doherty 2016; Gambier 2014). Machine translation (MT) is therefore often suggested as a necessary addition to the translation workflow. Though MT output in itself does not always yield the quality desired by the customer – depending on the type of system, the languages involved and the level of specialization – post-editing of the output by a translator has yielded promising results (Bowker and Buitrago Ciro 2015). Companies and researchers report an increase in productivity when using post-editing compared to regular human translation (Aranberri, Labaka, *et al.* 2014; Plitt and Masselot 2010), while at the same time still delivering products of comparable quality (Garcia 2010; O'Curran 2014).

Post-editing will presumably become an integral part of the translation process under certain circumstances – text types, language pairs – and better understanding of all features of the post-editing process and its impact may lead to improved translation processes. How do translators handle MT output? Is there a difference in the way they carry out the post-editing process compared to how they translate from scratch? Better insight into translators' activities can help improve translation tools as well as translator training. Depending on the elements in MT output that post-editors struggle with the most, either the MT system or the translation tool can be improved to better support the post-editing process, or translators can be trained to deal with these particular types of problems. Additionally, if we wish to make some suggestions towards translator training, we need to know whether students differ from professional translators, and if so, how.

In this paper, we report on translators' translation speed, gaze behavior, the use of external resources, the final quality of the product and also translators' attitudes towards human translation and post-editing. In the remainder of this introduction, we first discuss translation competence and the related notions of experience, profes-

sionalism and expertise. In the second part, we compare human translation with post-editing with regard to process, product, and attitude. In the final section, we bring both aspects together, and discuss implications and expectations for the present study.

1.1. Translation competence and experience

Translation process research is necessary to learn about the qualities good translators possess (Hansen 2010). This knowledge can then be integrated into translation training. Longitudinal studies like the work conducted by the PACTE group or within the framework of TransComp sought to create models for translation competence as well as its acquisition. The translation competence model developed by PACTE (2003) is a model of characteristics that define professional translators' competence and consists of several interacting subcompetences. Göpferich suggested another, though comparable, translation competence model in 2009. She assumed that the interaction between and coordination of the different subcompetences will improve with increased translation competence, and that beginning translators focus more on the surface level of a text, whereas more advanced translators use more global and diversified strategies. Both models implicitly contain the assumption that professional translators are the more competent translators.

Differences have indeed been found between professional and non-professional translators. Tirkkonen-Condit (1990), for example, found that participants with a higher degree of professionalism made more translation decisions (i.e., choices made among alternatives to carry on the translation process), but required less time overall to translate. In addition, non-professionals seemed to treat translation as a linguistic task, depending heavily on dictionaries during translation, whereas the professionals monitored the task at a higher level, taking aspects such as coherence and structure into account. It must be noted, that even though Tirkkonen-Condit (1990) uses the word 'professional,' she was working with translation students in their first and fifth years. Professionalism in this context has to be seen as 'level of experience' rather than actual professional translation experience. Tirkkonen-Condit's findings, though, were confirmed by Séguinot (1991), who claimed that the better students monitored at different levels, ranging from meaning to structure, cohesion, and register. Jensen's findings (1999) supported Tirkkonen-Condit's finding that more experienced translators use fewer dictionaries, but differed from Tirkkonen-Condit with regards to problem solving: experienced translators were found to perform fewer editing events, and fewer problem-solving activities, not more. Alves and Campos (2009) only looked at professional translators' translation processes and found that, despite the fact that all translators consulted external resources, most support came from internal resources, i.e., their own problem-solving strategies.

Other studies, however, have identified likenesses between professional and non-professional translators. Kiraly (1995), for example, established that there was no clear difference in the final quality of a target text produced by professional and non-professional subjects or even in their processes. He suggested that 'translator confidence' could be a more important factor than actual translation experience, which had previously been proposed by Laukkanen (1993) in an unpublished study referred to by Jääskeläinen (1996). Jääskeläinen (1996) compared two studies, the first conducted by

Gerloff in 1988, the second by herself in 1990, and came to conclusions similar to Kiraly's (1995): she found that professional translators did not necessarily perform better than novice translators, nor did they necessarily translate faster. Language proficiency was discarded as an obvious predictor of successful translation. What did seem to lead to successful translations, was spending more time on the translation, the intensity of research activity (dictionary consultations) and an increased number of processing activities (reading the text out loud or producing the translation). In an earlier study, Jääskeläinen and Tirkkonen-Condit (1991) looked for differences between more and less successful translators – irrespective of their actual translational experience – and found that the more successful translators paid more attention to factual contents of the source text as well as the needs of potential readers, whereas the weaker translators approached the task at a linguistic level.

An elaborate discussion of the issues related to experience and competence can be found in Jääskeläinen (2010). She addressed a few potential explanations for the seemingly incongruent findings listed above: professional translators might underperform in an experimental setup because they are not performing routine tasks, not all professionals can be expected to be experts – i.e., exhibit consistently superior performance – and specialization might play an important role as well. Jääskeläinen (2010) concluded the chapter by stressing that future research needs to include clear definitions of expertise and professionalism, as well as relevant background information on subjects.

This paper will consider professionalism as 'having experience working as a professional translator.' We will compare this level of experience with that of student translators, who do not have any experience beyond their studies, for two aspects of translation: the translation process and the translation product. With regard to the process and in light of the above-mentioned research, we expect professional translators to work faster than students (Tirkkonen-Condit 1990), and process texts on a higher level than students (Séguinot 1991). In practice, we expect them to consult dictionaries less frequently (Jensen 1999). With regard to the final product, we expect professionals to make fewer content and coherence errors. Overall, we do not expect the quality of the students' translations to be necessarily worse (Kiraly 1995), but we do expect the translators who specialize in general text types – the domain under scrutiny in the present paper – to perform better (Jääskeläinen 2010).

1.2. Post-editing vs. human translation

Since post-editing was proposed as a faster and thus cheaper alternative to regular human translation, early research into post-editing was mainly interested with identifying whether or not post-editing was indeed faster than regular translation. Especially for technical texts, this seemed to hold true (Plitt and Masselot 2010). Findings for more general text types, however, were not always as convincing. There were indications of post-editing being faster, though not always significantly so (Carl, Dragsted, Elming, *et al.* 2011; Garcia 2011).

In addition to speed, it is also important to study the cognitive aspects of both processes. Even if post-editing is found to be faster than human translation, if it is more cognitively demanding as well, perhaps translators will become exhausted sooner when post-editing compared to when translating from scratch, which will

lead to decreased productivity in the long run. Cognitive processing and effort can be studied via gaze data, building on the eye-mind hypothesis by Just and Carpenter (1980). An increased number of fixations (Doherty, O'Brien, *et al.* 2010) or higher average fixation durations (Carl, Dragsted, Elming, *et al.* 2011) have been used as indicators of increased cognitive processing. O'Brien (2007) compared human translation with post-editing and translation memory (TM) matches and found post-editing to be less cognitively demanding than human translation. Doherty, O'Brien, *et al.* (2010) used eye-tracking to identify MT comprehensibility and discovered that the total gaze time and number of fixations correlate well with MT quality. When considering the difference between source text processing and target text processing, findings become more complicated. Table 1 compares four studies: Carl, Dragsted, Elming, *et al.* (2011), Koglin (2015), Nitzke and Oster (2016), Sharmin, Šparkov, *et al.* (2008).

TABLE 1

Four-study comparison of findings related to source and target text fixation behavior

	Sharmin, *et al.* (2008)	Carl, Dragsted, Elming, *et al.* (2011)	Koglin (2015)	Nitzke & Oster (2016)
Translation method(s)	HT	HT & PE	HT & PE	HT & PE
Participants	students	professionals & bilingual	students	professionals & students
most visual attention during HT	TT	TT	ST	TT
most visual attention during PE	n/a	TT	TT	TT
visual attention on ST	n/a	HT: approaching TT attention PE: much lower than TT attention	HT: approaching TT attention PE: much lower than TT attention	HT: approaching TT attention PE: much lower than TT attention

HT = human translation, PE = post-editing, ST = source text, TT = target text, n/a = non-applicable

Table 1 shows that the target text receives the most visual attention for both methods of translation, with the exception of human translation in the Koglin (2015) study. When comparing the difference in attention between source text and target text, the difference is found to be smaller for human translation than for post-editing. Koglin (2015) suggested that the differences in experimental design could account for some of these different results, as participants in the Carl, Dragsted, *et al.* (2011) study had no previous post-editing experience, and time constraints had been imposed. The general trend in all studies seems to be that fixations during post-editing are more target text-centred, and those during human translation more source text-centred. Overall, it seems that post-editing is cognitively less demanding than human translation, although there is a difference when the source and target text processing are compared, with post-editors relying more heavily on the target text, presumably because there is already some output from the MT system present, whereas with human translation only the source text is given.

A final aspect we are interested in with regards to translation processes is the use of external resources. Overall, we expect translators to look up in fewer resources or spend less time in external resources when post-editing compared to translation, since the MT output should already provide some lexical elements to start from, whereas there is no support during human translation. Daems, Carl, *et al.* (2016) indeed found that more time was spent in external resources when translating from scratch than when post-editing. They further found no significant difference in the types of resources consulted for both methods. Concordancing tools were heavily used, which was confirmed for post-editing by Zapata (2016) as well.

In addition to the translation process, we wish to study the final product of both methods of translation. Interestingly, post-editing has been found to be beneficial to a translation's quality sometimes. Carl, Dragsted, Elming, *et al.* (2011) established that post-edited sentences were usually ranked as being better than sentences translated from scratch. Comparable results were obtained by Garcia (2011), who found that post-edited texts received better grades than texts translated from scratch. Guerberof (2009) compared human translation with translation from MT and translation from TM, and found that translation from MT led to better final quality than translation from TM, although regular human translation still outperformed translation from MT.

In addition to final quality, translators' attitudes matter as well. Even if post-editing is found to be faster, without having to compromise on quality, it is still important for translators to feel happy about their performance. Fulford (2002) found that, though professional translators are mostly skeptical about MT, they are interested in learning about it. A later survey conducted by Guerberof (2013) indicated that translators' attitudes towards MT were somewhat mixed. Specifically for translations from English into Dutch, post-editing was perceived as more effortful and more time-consuming than human translation, and participants preferred human translation over post-editing (Gaspari, Toral, *et al.* 2014).

From the above-mentioned research, we expect the post-editing process in the present study to take less time than the human translation process, the focus during post-editing to be on the target text, less time being spent in external resources when post-editing, the final products of both tasks to be of comparable quality, and translators' attitudes towards post-editing to be mixed.

1.3. Experience and translation methods

Building on previous research, we expect post-editing to be faster than human translation as well as cognitively less demanding for both groups, although we expect student translators to benefit the most from post-editing. Less experienced translators seem to handle translation as a lexical task (Tirkkonen-Condit 1990), and post-editing provides translators directly with lexical information. Consequently, we expect both groups to consult fewer dictionaries when post-editing, although we expect the difference to be larger for the students. We expect quality to be comparable across methods and both groups of participants (Carl, Dragsted, Elming, *et al.* 2011; Kiraly 1995). We expect students to be somewhat more positive towards post-editing than professionals (Moorkens and O'Brien 2015).

2. Method

2.1. Participants

Participants were 10 master's students of translation (2 male and 8 female) at Ghent University who had passed their final English Translation examination, and 13 professional translators (3 male and 10 female). With the exception of one translator, who had two years of experience, all translators had a minimum of 5 years and a maximum of 18 years of experience working as a full-time professional translator. Median age of students was 23 years (range 21-25). Median age of professional translators was 37 (range 25-51). All participants had normal or corrected to normal vision. Two students wore contact lenses and one student wore glasses, yet the calibration with the eye tracker was successful for all three. Two professional translators wore lenses. Calibration was problematic for one of the professionals. Sessions with problematic calibration were removed from the data. Students were given two gift vouchers of 50 euros each for their work, professional translators were paid 300 euros and their travel costs were refunded.

Students reported that they were aware of the existence of MT systems, and sometimes used them as an additional resource, but they had received no explicit post-editing training. Some professional translators had basic experience with post-editing, although none of the translators had ever post-edited an entire text. Their personal experience with post-editing – if any – was limited to MT output offered by a translation tool whenever the TM did not contain a good match.

All participants performed a LexTALE test (Lemhöfer and Broersma 2012), which is a word recognition test used in psycholinguistic experiments, so as to assess English proficiency. Other than being an indicator of vocabulary knowledge, it is also an indicator of general English proficiency. Proficiency in a language other than a person's native language can be considered to be part of the bilingual sub-competence as defined within the PACTE group's revised model of translation competence (2003), or the communicative competence in at least two languages as defined by Göpferich (2009). As such, it is an important factor to take into account when comparing the translation process of students and professionals. We expected to see a clear difference in proficiency between groups, but we could not find a statistically significant difference in LexTALE scores: t(21)=0.089, p=0.47; $\mu(\sigma^2)$ professionals = 88.27(90.24), $\mu(\sigma^2)$ students = 88(60.01).

2.2. Materials

Fifteen newspaper articles were selected from *Newsela*, a website which offers English newspaper articles at various levels of complexity. We selected 150/160-word passages from articles with comparable Lexile® scores (between 1160L and 1190L[1]). Lexile® measures are a scientifically established standard for text complexity and comprehension levels, providing a more accurate measure than regular readability measures. To control texts further, we manually compared them for readability, potential translation problems and MT quality. Texts with on average fewer than fifteen or more than twenty words per sentence were discarded, as well as texts that contained too many or too few complex compounds, idiomatic expressions, infrequent words or polysemous words. The MT was taken from *Google Translate* (output obtained January 24,

2014), and annotated with our two-step Translation Quality Assessment approach[2]. We discarded the texts that would be too problematic, or not problematic enough, for post-editors, based on the number of structural grammatical problems, lexical issues, logical problems and mistranslated polysemous words. The final corpus consisted of eight newspaper articles, each seven to ten sentences long (see Appendix). The topics of the texts varied, and the texts required no specialist knowledge to be translated.

2.3. Procedure

The experiment consisted of two sessions for each participant in the periods of June/July 2014 for students and April/May 2015 for professional translators. We used a combination of surveys, logging tools, and a retrospection session to be able to triangulate data from different sources. The first session started with a survey, to get an idea of participants' backgrounds, their experience with and attitude towards MT and post-editing, and a LexTALE test. This was followed by a copy task (participants had to copy a text to get used to the keyboard and screen) and a warm-up task combining post-editing and human translation, so participants could get used to the environment, the tools, and the different types of tasks. The actual experimental tasks consisted of two texts that they translated from scratch, and two texts that they post-edited. The second session started with a warm-up task as well, followed by two post-editing tasks and two translation tasks. The order of texts and tasks was balanced across participants within each group in a Latin square design. The final part of the session consisted of unsupervised retrospection (participants received the texts which they just translated and were requested to highlight elements they found particularly difficult to translate) and another survey, to get an idea of participants' attitude after the experiment.

We used a combination of keystroke logging and eye tracking tools to register the translation and post-editing processes. The main logging tool was the *CASMACAT* translators' workbench (Alabau, Bonk, *et al.* 2013). This tool can be used as a translation environment with the additional advantage that it contains keystroke logging and mouse-tracking software suited for subsequent translation process research. Participants only received one text at a time, and the text was subdivided into editable segments, corresponding to sentences in the source text. In addition to *CASMACAT*, we used an *EyeLink 1000* eyetracker to register participants' eye movements. A plugin connected the *EyeLink* with *CASMACAT*, so that the *CASMACAT* logging data contained gaze data in addition to its own process data. The final tool that was used was *Inputlog*, another keystroke logging tool. Though originally intended for writing research within the *Microsoft Word* environment, *Inputlog* (Leijten and Van Waes 2013) is capable of logging external applications as well. Since *CASMACAT* only logs what happens inside the *CASMACAT* interface, *Inputlog* was set to run in the background, to gather information on the resources participants consulted outside of *CASMACAT*.

2.4. Data Exclusion

For each participant, we collected logging data for four post-editing tasks and four regular translation tasks, leading to a total number of 92 post-editing tasks and 92 regular translation tasks. All student sessions could be used for further analysis, but some of the professional translators' data had to be discarded due to technical problems. Something went wrong with the logging files, there was an issue with calibration, or translators accidentally closed the *CASMACAT* interface, leading to a disruption in the logging files. Rather than work with potentially problematic data, we discarded those recordings altogether. In total, five human translation and five post-editing tasks were discarded, leading to an overall total of 87 post-editing tasks and 87 human translation tasks.

2.5. Analysis

The data consisted of the concatenated SG-files as obtained by processing the *CASMACAT* data (Carl, Schaeffer, *et al.* 2016). We normalized several variables and added some additional variables (which will be discussed where relevant) before loading the data file into *R*, a statistical software package (R Core Team 2014). In total, the data file consisted of 1444 observations – i.e., segments. For each analysis, we excluded the segments with incomplete data (due to minor problems with the eye-tracker or keystroke logger). The number of segments retained was never lower than 1412. All analyses discussed below were performed with *R*. We used the *lme4* package (Bates, Maechler, *et al.* 2014) and the *lmerTest* package (Kuznetsova, Brockhoff, *et al.* 2014) to perform linear mixed effects analyses on our data. Mixed effects models contain random effects in addition to fixed effects (= independent variables such as translation method). In our case, the random factors were always the participant (since we expect individual differences across participants to potentially influence the data) and the sentence code (an identifier of the text and the exact sentence in that text, since sentence-inherent aspects may also influence the data). A mixed model is constructed in such a way that it can identify the effect of independent variables on dependent variables while taking these random factors into account. Whenever we discuss mixed models below, the first step is to build a null model, which contains only the dependent variable and random factors. In the next step, the predictor (or independent) variables are added to the model and tested against the null model, to see if the predictor variable is actually capable of predicting the dependent variable. The predictor variables in the following models are always translation method (human translation or post-editing) and experience (student or professional) with interaction. To compare and select models we calculated Akaike's Information Criterion (AIC) value (Akaike 1974). The actual value itself has no meaning, only the difference between values for different models predicting the same dependent variable can be compared. According to Burnham and Anderson (2004), the best model is the model with the lowest AIC value. Their rule of thumb states that if the difference between models is less than 2, there is still substantial support for the weaker model. If the difference is between 4 and 7, there is far less support for the weaker model, and if the difference is greater than 10, there is hardly any support for the weaker model. A summary of all models discussed below can be found in Table 2.

3. Results

A general comparative analysis of two methods of translation (human translation and post-editing) and two groups of subjects (student translators and professional translators) was carried out. We are interested in the following aspects: the differences in process, as recorded by logging tools during the experiment, the quality of the final product, as established by means of translation quality assessment afterwards, and translators' general attitude towards post-editing and experience with it, as recorded by surveys before and after the experiment. These aspects will be discussed in more detail below.

3.1. Process: speed

The first aspect we investigated is translation speed. We built a mixed model with the average duration per word as a dependent variable. The model with predictors performed significantly better than the null model, yet only method had a significant effect, with post-editing reducing the time needed per word by almost a second compared to human translation. The effect is plotted in Figure 1. Students seem to require somewhat more time than professionals, although this effect was not significant.

Figure 1

Effect plot of interaction effect between method (human translation and post-editing, HT and PE, respectively) and experience (professional and student) on translation speed (=average duration per word in ms)
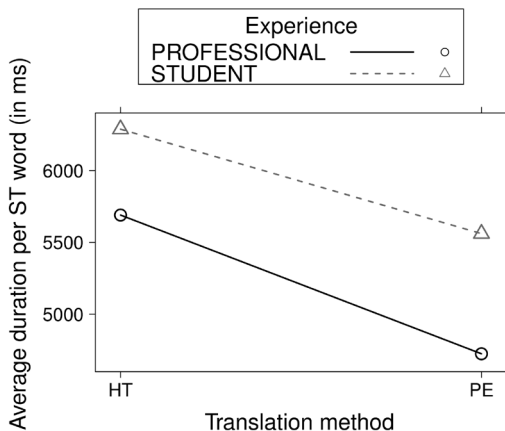

3.2. Process: cognitive effort

In addition to speed, we also calculated average fixation durations and average number of fixations to get an indication of cognitive effort. Average fixation duration was calculated by dividing the total fixation time within a segment divided by the number of fixations for that segment. Average number of fixations was calculated by dividing the number of fixations by the number of source text tokens in a segment.

In order to compare our data to the studies presented in Table 1, we first looked at the average number of fixations on source and target texts for both methods of translation, regardless of translator experience (Figure 2).

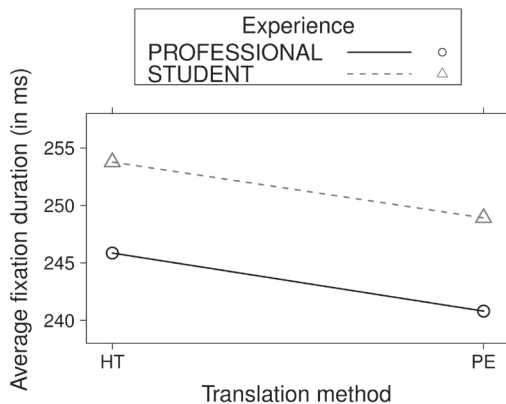**Average fixation number on source text and target text for both translation methods**
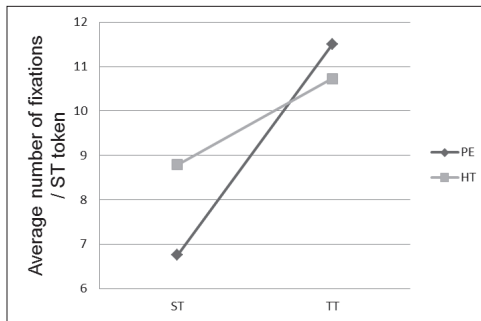
The averages presented in Figure 2 support the findings by Carl, Dragsted, Elming, *et al.* (2011) and Nitzke and Oster (2016) that most attention goes to the target text for both methods of translation and that the difference in attention is greater for post-editing than for human translation. This only contradicts Koglin (2015), who found more attention on the source text for human translation.

We then built mixed models to establish whether experience had any impact on fixation behavior. The first model had average total fixation duration as a dependent variable and method and experience with interaction as possible predictors. Only method was a significant predictor, with the average fixation duration being 5 milliseconds shorter when post-editing compared to human translation. The effect is plotted in Figure 3. As with translation speed, there again seems to be a trend for fixation duration to be longer for students compared to professional translators, but this effect was not found to be significant either.

Figure 3

Effect plot of interaction effect between method and experience for the average fixation duration (in ms) across the whole text

While overall average fixation duration gives us some indication of cognitive load, we also investigated fixations on source and target texts separately. For the analysis of the number of fixations on the source text, the fitted model again performed better than the null model but, again, only method was found to be significant. Processing of the source text during post-editing required fewer fixations per word than for human translation.

In the analysis of average fixation duration, only the interaction between method and experience is significant, showing that – for students only – the average fixation duration on the source text during post-editing is significantly shorter than during human translation (Figure 4).
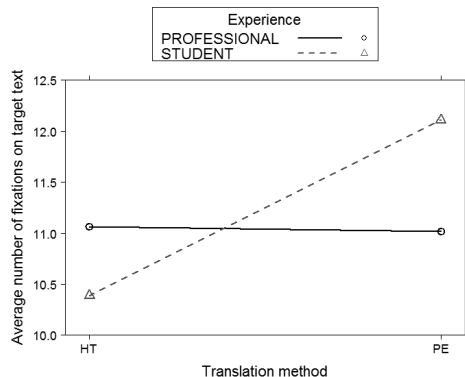
Figure 4

Effect plot of interaction effect between method and experience for the average fixation duration (in ms) on the source text
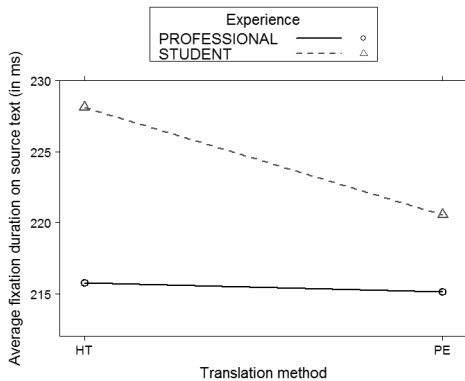
For the average number of fixations on the target text, too, the summary of the fitted model showed only the interaction effect of method and experience to be significant. The effect is plotted in Figure 5 below.

Figure 5

Effect plot of interaction effect between method and experience for the average number of fixations on the target text

There is a higher number of fixations on the target text when post-editing compared to human translation, but only for the students (Figure 5). The number of

fixations on the target text for professional translators seems to be comparable for both methods of translation.

We also looked at the average fixation duration on the target text. The model with fixed effects performed better than the null model, yet only method was found to be a significant predictor, with average fixation duration being 5 milliseconds shorter when post-editing compared to human translation.

3.3. Process: use of external resources

To observe external resource behavior of the translators, we coded the information from *Inputlog*. Each consultation was labeled with the relevant category: dictionary, concordancer; search, encyclopedia, MT or 'other' (grammar or spelling websites, fora, news sites, termbanks, and synonym sites). We added the numbers of times each type of resource was consulted as well as the time spent in each type of resource to the SG-data file and calculated the average number of external resources consulted per source token, as well as the average time spent in external resources per source token. We fitted a mixed effect model with total time spent in external resources as dependent variable, but this model did not outperform the null model. The same holds true for the model predicting the total number of source hits (number of times an external resource was consulted). Here, method was almost significant, but not sufficiently so to justify the model.

While the total time spent in external resources did not significantly differ between groups or translation methods, Figure 6 gives an overview of the percentage of overall time spent in external resources for each type of resource and reveals that for both groups of participants and both methods, *Google search*, concordancers and dictionaries are the most common resources. It can be seen, however, that students rely more heavily on dictionaries than professional translators, as was also confirmed statistically, t(999)=5.96, p<0.001). Professional translators seem to spend somewhat more time on machine translation websites than students, even when post-editing, which seems counterintuitive at first. From the surveys, however, we learned that *Google Translate* is often used to check the translation of a single word and to get alternative translations. Students consulted synonym websites rather than *Google Translate* when looking for alternative translations.

FIGURE 6

Percentage of total time spent in external resources per resource type for both methods and levels of experience
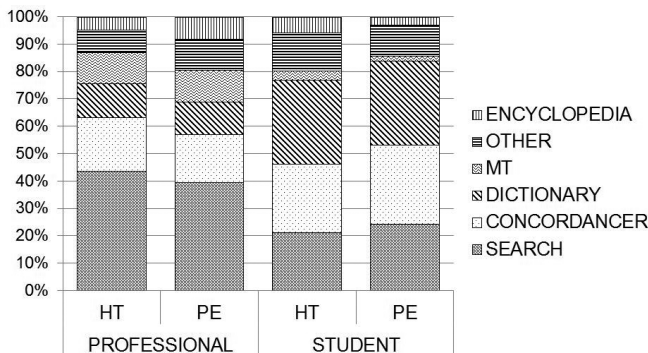
An investigation into the types of resources used within each category, revealed that students used both the *Glosbe* concordancer and *Linguee*, whereas professional translators only used *Linguee*. In total, twenty-two different types of dictionaries were consulted across all participants. Six of those were consulted only by students, whereas nine of those were only consulted by professional translators. The dictionary most commonly used by all participants is *Van Dale*, a classic dictionary for the Dutch language. *Van Dale* was used more frequently than all other dictionaries combined. We also know the language of the search queries, but this seems fairly comparable across groups: 76% of the professional translators' queries in *Van Dale* were English (the source language), the others were Dutch (the target language), compared to 82% of search queries within the student group.

3.4. Product

We used our fine-grained translation quality assessment approach (Daems, Macken, *et al.* 2013) to determine the final quality of the product. Two of the authors annotated all final texts for acceptability (target text, language and audience) and adequacy (correspondence to the source text) issues using the *brat rapid annotation tool* (Stenetorp, Pyysalo, *et al.* 2012). All error classifications were discussed by both annotators, and only the annotations both annotators agreed on were retained for the final analysis. Each error type receives an error weight, corresponding to the severity of the error (for example, a typo would receive a weight of 1, whereas a contradiction would receive a weight of 4, see Note 2 for details).

We fitted a linear mixed effects model with average total error weight per word as dependent variable, but the model with predictor variable did not outperform the null model, although the predictor experience was almost significant. We can derive that students perform somewhat worse than professional translators, but the effect was not statistically significant. As such, we can conclude that there is no difference in overall quality between human translation and post-editing, or between students and professional translators.

In addition to overall quality, it is also interesting to compare the types of errors common for both methods of translation and both groups of participants. Figure 7 shows the percentage of all errors made for the main error categories.

Figure 7

Occurrence of main error types for both methods and levels of experience
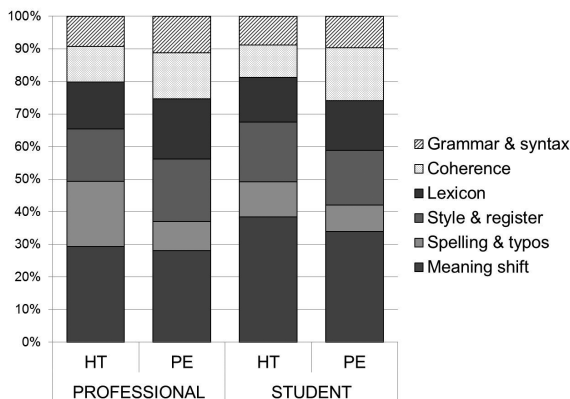
What can be derived from Figure 7 is that the most common errors for students are meaning shifts, i.e., discrepancies in meaning between source and target texts, although this particular error category becomes less common when post-editing. For professional translators, spelling and typos are far less common in post-editing than for human translation. Coherence is somewhat more problematic for post-editing compared to human translation in both groups.

Number of occurrences of each error type making up at least 5% of errors for at least one of the conditions
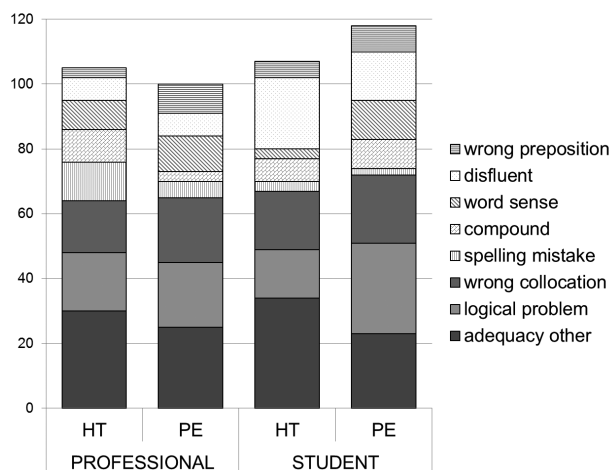

Figure 8 shows an even more fine-grained picture, displaying the number of occurrences of the most common error categories. Only error categories that accounted for minimum 5% of all errors in a specific condition have been included in the graph. The most common category for students' post-editing is 'logical problem,' which groups together those choices that do not make sense in the context of the text, or the world at large. For example, if a text is about snakes, and the translator used the word 'fly' to describe how the snake moved, this is a logical problem in the sense that snakes do not fly. A textual logical problem occurs when, for example, the instruction to 'open a door' is repeated without closing that very door in between instructions. This is a logical problem in the sense that the door referred to in the text cannot be opened twice. Interestingly, there is a lower number of word sense errors (a specific type of adequacy issue) in human translations than in post-edited translations, especially for the students; a higher number of disfluent constructions when comparing students with professional translators, especially in the human translation condition; and an abundance of spelling mistakes in the human translation condition for professional translators.

To verify the assumption that translators specialized in the translation of general texts outperform translators who do not specialize in general text translation, we looked at the number of errors each professional translator made, and compared that with the survey data: their years of professional experience and their level of specialization for the current text type, i.e., percentage of their time translating general texts (Figure 9).

Figure 9

Relationship between professional translators' level of specialization (percentage of time spent translating general text types, plotted on secondary axis), their translation experience (years, plotted on secondary axis), and the total error count for their human translation and post-editing tasks (plotted on primary axis). Labels on x-axis are participant codes
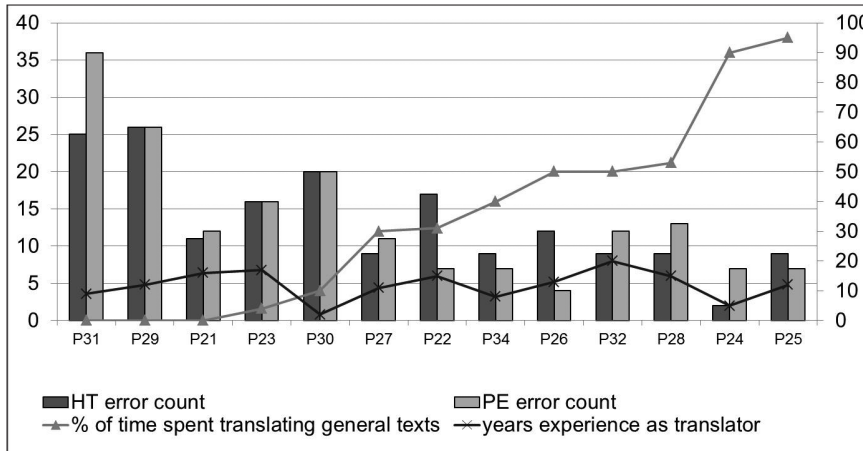
While the number of years of professional experience is not correlated with quality (r=-0.08, p=0.79 for HT; r=-0.17, p=0.57 for PE), there is a negative correlation between level of specialization and quality (r=-0.76, p=0.003 for HT; r=-0.66, p=0.01 for PE), with participants 24 and 25 – spending respectively 90% and 95% of their time translating general texts – producing the highest quality texts.

Table 2

Overview of AIC difference and model summaries

Fig.	dependent variable	AIC difference predictor model – null model	predictors	effect	sig.
1	Time per word (in ms)	-22	method	-965 (± 227)	**< 0.001**
			experience	598 (± 702)	0.4
			method:exp	238 (± 332)	0.47
2	Average fixation duration (in ms)	-24	method	-5.06 (± 1.24)	**< 0.001**
			experience	7.92 (± 15.63)	0.62
			method:exp	0.2 (± 1.84)	0.91
n/a	Average # fixations ST	-52	method	-1.94 (± 0.36)	**< 0.001**
			experience	-0.66 (± 1.54)	0.67
			method:exp	-0.2 (± 0.53)	0.7
3	Average fixation duration ST	-12	method	-0.6 (± 1.67)	0.72
			experience	12.34 (± 12.39)	0.33
			method:exp	-6.95 (± 2.48)	**0.005**
4	Average # fixations TT	-10	method	-0.05 (± 0.4)	0.91
			experience	-0.68 (± 1.18)	0.57
			method:exp	1.77 (± 0.59)	**0.003**

n/a	Average fixation duration TT	-105	method	-11.77 (± 1.64)	**< 0.001**
			experience	7.82 (± 17.1)	0.65
			method:exp	-2.43 (± 2.42)	0.31
n/a	Time spent in external resources per word (in ms)	4	n/a	n/a	n/a
n/a	Average # external resources per word	-3	method	-0.02 (± 0.01)	0.07
			experience	0.03 (± 0.03)	0.36
			method:exp	-0.006 (± 0.02)	0.72
n/a	Average total error weight	2	n/a	n/a	n/a

3.5. Attitude

Each participant filled out a survey before and after the experiment. Surveys were created in Dutch using *Qualtrics*. About half of the students as well as the professionals indicated that they had some experience with post-editing (question: 'I … make use of MT systems while translating'; options: 'never, seldom, sometimes, often, always'). Their additional comments, however, showed that they often considered post-editing to be 'working with a translation tool,' including editing TM matches as well as MT output. The opinions on post-editing taken from the pre-test survey thus encompass issues other than post-editing. Most students who claimed to have some experience with post-editing found it equally rewarding as human translation, or preferred human translation to a small degree. Professional translators with knowledge of post-editing found human translation more rewarding, although they did not mind post-editing. Professionals had different feelings about post-editing: those who enjoyed it mostly enjoyed not having to start from scratch, and noted that it could save them some time, provided the output was of sufficient quality. Those who preferred regular translation mentioned creativity and freedom as an important factor, and they did not believe that post-editing would necessarily save time. One translator explicitly mentioned the reduced per-word fee as a reason not to prefer post-editing. Both students and professionals found MT output 'often' or 'sometimes' useful, saying it gives them some original ideas. With regard to speed, half the number of students expected post-editing to be faster than regular translation, compared to only three out of thirteen professionals. From the students and professionals who claimed to have some knowledge of post-editing, only two (one student and one professional) believed they produced better quality with post-editing. From the participants without post-editing experience, only one professional translator expected this to be the case. The quality concerns listed explicitly were comparable among groups: a product can contain non-idiomatic expressions because of post-editing, and it is harder to control for consistency when post-editing than translating. The last concern seemed to be a valid one, as coherence issues were more common when post-editing compared to human translation (Figure 6).

In the survey taken after the experiment, we asked participants about their preferred translation method for the text type, their perceived speed, and what they thought was the least tiring translation method. Most participants, students and professionals alike, preferred human translation over post-editing. Four professionals

and one student preferred post-editing. With regard to speed, six students and five professionals were convinced that post-editing was faster, compared to only one student and two professionals who believed human translation was faster. The remaining three students and six professionals did not perceive a difference in speed between both methods of translation. For most participants, their perception of speed did not change after the experiment. In each group, only two participants changed their minds in favor of human translation, whereas five professionals and two students believed post-editing to be faster than they thought before the experiment. The question about which translation method participants considered to be the most tiring was included since we are interested in the perceived cognitive load. Responses for the professional translators varied, with a comparable number of participants choosing one of the three options (HT less tiring, PE less tiring, both equally tiring). The result is slightly different with the students, with only one student considering HT to be less tiring, and the others selecting PE or 'equally tiring.' It is interesting to see the students' perceptions correspond to the fixation analysis which showed that post-editing was cognitively less demanding than human translation.

4. Discussion

Overall, we found that students and professional translators are not as different as often thought, while the differences between human translation and post-editing are mostly in line with previous research. This may partly be due to the relatively small number of participants or the text type, and it is possible that other statistically significant effects to surface with larger datasets or more specialized texts. In the following sections, we discuss our most important findings and we formulate more practical suggestions.

4.1. Process: speed

Increased productivity is one of the main reasons for adding post-editing to the translation workflow. While post-editing has been shown to be faster than human-translation for technical texts (Plitt and Masselot 2010), we now also found it to be statistically significantly faster for general text types. There was no significant difference in processing speed between students and professionals, although students do seem to require somewhat more time. While in contrast with Tirkkonen-Condit (1990), this finding is in line with Jääskeläinen's observation (1996) that professional translators do not necessarily translate faster than students.

4.2. Process: cognitive effort

Cognitive effort needs to be taken into account as well, since a higher cognitive effort can cause fatigue and be detrimental to productivity in the long run as well as translators' attitude. We expected post-editing to be cognitively less demanding (O'Brien 2007) because it provides translators with lexical information, and it might help them make decisions in situations where multiple translation options are possible. Seeing how students treat translation as a lexical task (Tirkkonen-Condit 1990), we expected post-editing to be especially beneficial for them. The effect shown in Figure 2 con-

firms that post-editing is cognitively less demanding than human translation, but we did not find an effect for experience.

A more detailed analysis, however, revealed that there are some differences between students and professionals when source text and target text fixations are studied in isolation. Laukkanen (1993) found that insecurity leads to heavier reliance on the source text. As such, we expected students to rely more heavily on the source text than professional translators. We also expected less reliance on the source text when post-editing (Carl, Dragsted, Elming, *et al.* 2011). Processing of the source text during post-editing did indeed require fewer fixations per word than for human translation. The average fixation duration was also shorter when post-editing (Nitzke and Oster 2016), but only for the students. It seems that the cognitive load of professional processing of the source text is equal for both methods. For students, however, there is a clear difference between the cognitive load during translation – which is also higher than that of the professionals – and the load during post-editing (which approaches that of the professionals).

With regard to target text processing, we expect a higher number of fixations during the post-editing process compared to the human translation process, as the machine translation system already provides a target text to work on, whereas with human translation the source text is the main source of information. There was indeed a higher number of fixations on the target text when post-editing compared to human translation (Nitzke and Oster 2016).

In sum, the fixation analysis has shown that post-editing overall is less cognitively demanding than human translation (O'Brien 2007) for professional translators and students alike. When processing the source text, students benefit more from the post-editing condition than professional translators, whereas when processing the target text, post-editing seems to be less cognitively demanding for both groups, although professional translators process the target text differently, with students requiring fewer fixations when translating from scratch compared to post-editing and professional translators requiring a comparable number of fixations for both methods. Further analysis of the actual text production and final translations is needed to get a better idea of what is really happening, and whether or not it is successful. This knowledge can then be used to better train students or provide feedback to professionals. Perhaps the professional translators treat post-editing more as a regular translation task, or they know how to move through a text more efficiently than students, considering they have more experience with spotting and solving translation issues.

4.3. Process: external resources

There was no significant difference for the time spent in external resources by students or professionals, during human translation or post-editing, which is in contrast with our previous findings on a smaller dataset (Daems, Carl, *et al.* 2016). Closer inspection revealed abundant use of dictionaries, search functions and concordancers, the last corresponding to findings by Zapata (2016). Students also relied significantly more on dictionaries than professionals, in line with findings by Jensen (1999) that the use of dictionaries decreases with experience. This goes to show that external resources are crucial for students and professionals alike, independent of translation task.

4.4. Product

Supporting the findings by Jääskeläinen and Tirkkonen-Condit (1991), Kiraly (1995), and Jääskeläinen (1996), we found that the more experienced translators are not necessarily the more successful translators, with students producing products of comparable overall quality. There seems to be no statistically significant difference in quality between human translation and post-editing either, which confirms previous findings that post-editing can produce texts that are at least as good as human translations (Garcia 2011).

The detailed analysis also confirmed other findings. Students seem to struggle with meaning shifts, disfluency and logical problems. This is in line with findings by Séguinot (1991), who characterized structure, cohesion and register as advanced translation issues, which might in part be explained by the fact that students treat translation as a linguistic task (Tirkkonen-Condit 1990).

We further found that professional translators specialized in the translation of general texts – the most common text type in the students' training – outperformed translators who do not specialize in general text translation (Jääskeläinen 2010). We expect to see more significant differences between students and professionals for specialized types of translation: although students are introduced to some specialized text translation in their classes, that would presumably not be enough for them to perform equally well as, let alone outperform, professionals with a few years of specialized experience. Other factors might also provide different insights into the translation and post-editing process, such as 'confidence' (Kiraly 1995), translation styles (Carl, Dragsted, & Jakobsen 2011) or translation patterns (Asadi and Séguinot 2005) rather than experience.

4.5. Attitude

In line with Guerberof (2013) we can tentatively conclude from the surveys that student and professional translators hold similar opinions, and that preferences seem to be caused by individual differences rather than between group differences. Both groups seem to prefer human translation, although they do not mind post-editing, and while they are not always convinced of post-editing quality, they mostly agree that post-editing is faster than human translation, especially after participating in the experiment.

We can only detect one obvious difference between students and professionals when considering their opinions about the least tiring translation method. Professional translators experienced no obvious difference, whereas students seemed to consider post-editing the least tiring method of translation. This might be explained in part by the findings by Tirkkonen-Condit (1990) that non-professional participants treat translation as a linguistic task and mostly rely on dictionaries to solve problems. In a post-editing condition, lexical information is already provided by the MT output, which might reduce the need to look for additional information, and thus make the students experience the process as less tiring than regular human translation.

4.6. Practical suggestions

When integrating external resources into translation tools, we suggest integrating the most often consulted resources: dictionaries, concordancer, and Google search. Dictionaries should be given primary focus for novice translators in particular. Translators, especially novice translators, could further benefit from visual clues in the target text, seeing how the target text received the most visual attention. At the same time, attention could be drawn to the source text to avoid adequacy issues, which could be caused by poor consultation of the source text. For example, polysemous words could be highlighted, especially during post-editing.

Since translators' attitude became somewhat more positive towards post-editing after participating in the experiment, we believe that post-editing should be included in translator training. Students could be taught to detect typical machine translation errors that currently go unnoticed, such as meaning shifts, wrong collocations, logical problems, and word sense issues. In view of translators' present doubts about the final quality of post-edited texts, it might also be a good idea to make translators more aware of its quality, since we found no significant difference between post-edited and human translation quality.

5. Conclusions

We found students and professional translators to be more alike than often thought, even when working with different translation methods (human translation and post-editing). Still, post-editing seemed more beneficial for students than for professionals, with students experiencing it as less tiring than regular translation. Our findings imply that post-editing is a viable alternative for human translation, even for general text types: it is faster without leading to lower quality results, and it is cognitively less demanding. The fact that the professionals did not obviously outperform students from Ghent University might mean that the current translation curriculum prepares students well for the translation of general texts. Looking at the benefits of post-editing and the fact that most participants weren't opposed to post-editing after participating, perhaps specific post-editing training could be added to the translation curriculum to make for an even better future generation of translators.

NOTES

1. The authors would like to thank MetaMetrics® for their permission to publish Lexile scores in the present paper. <https://www.metametricsinc.com/lexile-framework-reading>.
2. Guidelines for the Translation Quality Assessment approach can be found via <http://users.ugent.be/~jvdaems/TQA_guidelines_2.0.html>.

REFERENCES

Akaike, Hirotugu (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control.* 19(6):716-723.

Alabau, Vincent, Bonk, Ragnar, Buck, Christian, *et al.* (2013): CASMACAT: An open source workbench for advanced computer aided translation. *The Prague Bulletin of Mathematical Linguistics.* 100:101-112.

Alves, Fabio and Campos, Tânia Liparini (2009): Translation technology in time: Investigating the impact of translation memory systems and time pressure on types of internal and

external support. *In*: Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees, eds. *Behind the Mind: Methods, Models and Results in Translation Process Research*. Frederiksberg C: Samfundslitteratur, 191-218.

Aranberri, Nora, Labaka, Gorka, Diaz de Ilarraza, Arantza, *et al.* (2014): *Comparison of Post-editing Productivity Between Professional Translators and Lay Users*. Paper presented at the AMTA 2014 3rd Workshop on Post-editing Technology and Practice (WPTP-3), Vancouver, Canada, October 26, 2014.

Asadi, Paula and Séguinot, Candace (2005): Shortcuts, strategies and general patterns in a process study of nine professionals. *Meta*. 50(2):522-547.

Bates, Douglas, Maechler, Martin, Bolker, Ben, *et al.* (2014): Lme4: Linear mixed-effects models using eigen and s4. R package version 1.1-7. Retrieved from <http://CRAN.R-project.org/package=lme4>.

Bowker, Lynne and Buitrago Ciro, Jairo (2015): Investigating the usefulness of machine translation for newcomers at the public library. *Translation and Interpreting Studies*. 10(2):165-186.

Burnham, Kenneth and Anderson, David (2004): Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*. 33:261-304.

Carl, Michael, Dragsted, Barbara, Elming, Jakob, *et al.* (2011): The process of post-editing: A pilot study. *In*: Bernadette Sharp, Michael Zock, Michael Carl and Arnt Lykke Jakobsen, eds. *Proceedings of the 8th International NLPCS Workshop*. Frederiksberg C: Samfundsliterattur, 131-142.

Carl, Michael, Dragsted, Barbara and Jakobsen, Arnt Lykke (2011): A taxonomy of human translation styles. *Translation Journal*. 16(2).

Carl, Michael, Schaeffer, Moritz and Bangalore, Srinivas (2016): The CRITT translation process research database. *In*: Michael Carl, Srinivas Bangalore and Moritz Schaeffer, eds. *New Directions in Empirical Translation Process Research. Exploring the CRITT TPR-DB*. Switzerland: Springer International Publishing, 13-54.

Daems, Joke, Carl, Michael, Vandepitte, Sonia, *et al.* (2016): The effectiveness of consulting external resources during translation and postediting of general text types. *In*: Michael Carl, Srinivas Bangalore and Moritz Schaeffer, eds. *New Directions in Empirical Translation Process Research: CRITT TPR-DB*. Switzerland: Springer International Publishing, 111-133.

Daems, Joke, Macken, Lieve and Vandepitte, Sonia (2013): *Quality as the sum of its parts: A two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE*. Paper presented at the MT Summit XIV 2nd Workshop on Post-editing Technology and Practice (WPTP-2), Nice, France, 2-6 September, 2013.

Doherty, Stephen (2016): The impact of translation technologies on the process and product of translation. *International Journal of Communication*. 10:647-969.

Doherty, Stephen, O'Brien, Sharon and Carl, Michael (2010): Eye tracking as an MT evaluation technique. *Machine Translation*. 24:1-13.

Fulford, Heather (2002): *Freelance translators and machine translation: An investigation of perceptions, uptake, experience and training needs*. Paper presented at the 6th annual conference of the European Association for Machine Translation Workshop on Teaching Machine Translation, Manchester, United Kingdom, 14-15 November, 2002.

Gambier, Yves (2014): Changing landscape in translation. *International Journal of Society, Culture & Language*. 2(2):1-12.

Garcia, Ignacio (2010): Is machine translation ready yet? *Target*. 22(1):7-21.

Garcia, Ignacio (2011): Translating by post-editing: Is it the way forward? *Machine Translation*. 25:217-237.

Gaspari, Federico, Toral, Antonio, Naskar, Sudip Kumar, *et al.* (2014): *Perception vs reality: Measuring machine translation post-editing productivity* Paper presented at the AMTA 2014 3rd Workshop on Post-editing Technology and Practice (WPTP-3), Vancouver, Canada , October 26, 2014.

Gerloff, Pamela. (1988): *From French to English: A look at the translation process in students, bilinguals, and professional translators.* Doctoral dissertation. Cambridge: Harvard University.

Göpferich, Susanne (2009): Towards a model of translation competence and its acquisition: The longitudinal study *transcomp. In*: Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees, eds. *Behind the Mind: Methods, Models and Results in Translation Process Research.* Frederiksberg C: Samfundslitteratur, 11-38.

Guerberof, Ana (2009): *Productivity and quality in mt post-editing.* Paper presented at the MT Summit XII Beyond Translation Memories Workshop (WS3), Ottawa, Ontario, Canada, August 29, 2009.

Guerberof, Ana (2013): What do professional translators think about postediting? *JoSTrans.* 19:75-95.

Hansen, Gyde (2010): Integrative description in translation process research. *In*: Gregory M. Shreve and Erik Angelone, eds. *Translation and Cognition.* Amsterdam: John Benjamins, 189-211.

Jääskeläinen, Riitta. (1990): *Features of successful translation processes: A think-aloud protocol study.* Licentiate thesis. Joensuu: University of Joensuu, Savonlinna School of Translation Studies.

Jääskeläinen, Riitta (1996): Hard work will bear beautiful fruit. A comparison of two think-aloud protocol studies. *Meta.* 41(1):60-74.

Jääskeläinen, Riitta (2010): Are all professionals experts? Definitions of expertise and reinterpretation of research evidence in process studies. *In*: Gregory M. Shreve and Erik Angelone, eds. *Translation and Cognition.* Amsterdam: John Benjamins, 213-262.

Jääskeläinen, Riitta and Tirkkonen-Condit, Sonja (1991): Automatised processes in professional vs. non-professional translation: A think-aloud protocol study. *In*: Sonja Tirkkonen-Condit, ed. *Empirical Research in Translation and Intercultural Studies.* Tübingen: Gunter Narr, 89-109.

Jensen, Astrid (1999): Time pressure in translation. *In*: Gyde Hansen, ed. *Probing the Process in Translation: Methods and Results.* Vol. 24. Frederiksberg C: Samfundslitteratur, 103-119.

Just, Marcel and Carpenter, Patricia (1980): A theory of reading: From eye fixations to comprehension. *Psychological Review.* 87(4):329-354.

Kiraly, Donald (1995): *Pathways to Translation: Pedagogy and Process.* Kent, Ohio: Kent State University Press.

Koglin, Arlene (2015): An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors. *The International Journal for Translation & Interpreting Research.* 7(1):126-141.

Kuznetsova, Alexandra, Brockhoff, Per Bruun and Christensen, Rune Haubo Bojesen. (2014): Lmertest: Tests in linear mixed effects models. R package version 2.0-20. Retrieved from <http://CRAN.R-project.org/package=lmerTest>.

Laukkanen, Johanna. (1993). *Routine vs. Non-routine processes in translation: A think-aloud protocol study.* pro gradu thesis. Joensuu: University of Joensuu, Savonlinna School of Translation Studies.

Leijten, Mariëlle and Van Waes, Luuk (2013): Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes. *Written Communication.* 30(3):325-343.

Lemhöfer, Kristin and Broersma, Mirjam (2012): Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods.* 44:325-343.

Moorkens, Joss and O'Brien, Sharon (2015): *Post-editing evaluations: Trade-offs between novice and professional participants.* Paper presented at the 18th Conference of the European Association for Machine Translation (EAMT 2015), Antalya, Turkey, May 11-13, 2015.

Nitzke, Jean and Oster, Katharina (2016): Comparing translation and post-editing: An annotation schema for activity units. *In*: Michael Carl, Srinivas Bangalore and Moritz Schaeffer, eds. *New Directions in Empirical Translation Process Research. Exploring the CRITT TPR-DB.* Switzerland: Springer International Publishing, 293-308.

O'Brien, Sharon (2007): Eye-tracking and translation memory matches. *Perspectives: Studies in Translatology.* 14(3):185-205.

O'Curran, Elaine (2014): *Translation quality in post-edited versus human-translated segments: A case study.* Paper presented at the AMTA 2014 3rd Workshop on Post-editing Technology and Practice (WPTP-3), Vancouver, Canada, October 26, 2014.

PACTE (2003): Building a translation competence model. *In*: Fabio Alves, ed. *Triangulating Translation: Perspectives in Process Oriented Research.* Amsterdam: John Benjamins, 43-66.

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

R Core Team. (2014): R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. Retrieved from <http://www.R-project.org>.

Séguinot, Candace (1991): A study of student translation strategies. *In*: Sonja Tirkkonen-Condit; ed. *Empirical Research in Translation and Intercultural Studies.* Tübingen: Gunter Narr, 79-88.

Sharmin, Selina, Šparkov, Oleg , Räihä, Kari-Jouko, *et al.* (2008): *Effects of time pressure and text complexity on translators' fixations.* Paper presented at the 2008 Symposium on Eye-Tracking Research and Applications (ETRA'08), Savannah, GA, USA, March 26-28, 2008.

Stenetorp, Pontus, Pyysalo, Sampo, Topic, Goran, *et al.* (2012): *Brat: A web-based tool for NLP-assisted text annotation.* Paper presented at the Demonstrations Session at EACL 2012, Avignon, France, April 23-27, 2012.

Tirkkonen-Condit, Sonja (1990): Professional vs. non-professional translation: A think-aloud protocol study. *In*: M. A. K. Halliday, John Gibbons and Howard Nicholas, eds. *Learning, Keeping and Using Language: Selected papers from the Eighth World Congress of Applied Linguistics.* Amsterdam: John Benjamins, 381-394.

Zapata, Julian (2016): Investigating translator-information interaction: A case study on the use of the prototype biconcordancer tool integrated in CASMACAT. *In*: Michael Carl, Srinivas Bangalore and Moritz Schaeffer, eds. *New Directions in Empirical Translation Process Research. Exploring the CRITT TPR-DB.* Switzerland: Springer International Publishing, 135-152.

APPENDIX

Source texts used

TEXT 1: Listen up. This blue whale's earwax tells the story of its life and locale
A giant plug of earwax pulled from a dead blue whale is providing scientists with a detailed biography of the wild animal's life, from birth to death, in 6-month chapters. The scientists' new technique is described in the journal Proceedings of the National Academy of Science. It arms researchers with a tool to understand a whale's hormonal and chemical biography – and a window into how pollutants, some long discontinued, still remain in the environment today. Whales are often called marine sentinels because they can reveal a lot about the waters they swim through, said Sascha Usenko. He is an analytic environmental chemist at Baylor University. "These types of marine mammals that are long-lived have a great ability to accumulate contaminants, and so they're often perceived as being sentinels of their ecosystem," said Usenko, who helped write the study.

TEXT 2: Done with mirrors: Bringing the sun to a small Norwegian town
Tucked between steep mountains, Rjukan is normally shrouded in shadow for almost six months a year. Residents have to catch a cable car to the top of a nearby precipice to get a dose of midday vitamin D. But on Wednesday, faint rays from the winter sun for the first time reached the town's market square, thanks to three 183-square-foot mirrors placed on a mountain. Cheering families, some on sun loungers, drinking cocktails and waving

Norwegian flags, donned shades as the sun crept from behind a cloud. It hit the mirrors and reflected down onto the faces of delighted children below. «Before when it was a fine day, you would see that the sky was blue and you knew that the sun was shining. But you couldn't quite see it. It was very frustrating,» said Karin Roe, from the local tourist office.

TEXT 3: China's "orphan grandparents" can sue absent children for not visiting
Yan Meiyue, 90, said her 72-year-old daughter rarely visited, even for the annual Spring Festival, when families traditionally reunite. So Yan, a widow since her husband's death nearly a decade ago, spends every weekday at a modest community center near her home, where she plays mahjong and eats meals prepared by a volunteer staff. "The volunteers keep us company," she said with a smile, her voice trailing off. Yan is one of a rapidly growing number of self-described "orphan grandparents" who feel personally or financially abandoned. It's a troubling trend for China where elders have traditionally been among the most respected members of society. For centuries, Chinese households have included many generations, and Chinese elders could count on their children caring for them as they grew frail. But today this ancient social contract is giving way. The booming Chinese economy is prying apart families with job opportunities that lure adult children to distant cities or other countries.

TEXT 4: Lie-detector test comes under fire as FBI hiring tool
Thousands of job seekers come to FBI offices all across the country every year, only to be turned away from the top U.S. law enforcement agency. The reason is not because they do not have the right work experience or education, or because they have a criminal record. They are turned down because they failed their polygraph tests. The polygraph is also known as a lie detector. Many scientists disagree with the FBI's policy of rejecting candidates who fail the tests. They say government agencies should not rely solely on the tests to decide whether to hire or fire someone. Experts say polygraph testing does not reliably show when somebody is actually lying, especially when they are applying for a job. "I was called a lazy, lying, drug-dealing junkie by a man who doesn't know me, my stellar background or my societal contributions," wrote one applicant in Baltimore.

TEXT 5: Huge art pieces, done on an iPad, draw gasps at museum exhibit
Britain's most celebrated living artist, David Hockney, is pioneering in the art world again. Happily hunched over his iPad, he is using his index finger like a paintbrush to create colorful landscapes and richly layered scenes on a touch screen. «It's a very new medium,» said Hockney. So new, in fact, he wasn't sure what he was creating until he began printing his digital images a few years ago. «I was pretty amazed by them actually,» he said, laughing. «I'm still amazed.» A new exhibit of Hockney's work, including many iPad images, opened Saturday in San Francisco's de Young Museum. Located in Golden Gate Park, the museum is just a short trip for Silicon Valley techies who created both the hardware and software for this 21st-century reinvention of finger-painting. The show is billed as the museum's largest ever.

TEXT 6: Some young Iranians ignore officially enforced anger at the West
World leaders in Geneva negotiated the future of Iran's nuclear development program. In exchange for limiting uranium enrichment, Iran will be freed from certain trade restrictions, known as sanctions. But religious hard-liners here continued to warn of a deceitful West scheming to weaken the Islamic Republic. Yet things are different for the mostly young, jeans-clad set in this busy capital city. Among them, chanting denunciations of the United States is as out of date as 1970s fashion. "In art, in fashion, in cinema and in our daily lifestyle, we copycat American culture," said Sarah. She is the proprietor of a cozy cafe in the basement of a high-rise in northwestern Tehran. "There is a big difference between the approved culture and the reality of urban lifestyles in big cities like Tehran."

TEXT 7: Climate change could lead to more wars and civil unrest, a study says
The theory that high temperatures fuel aggressive and violent behavior is only just beginning to be studied. Using examples ranging widely from road rage, ancient wars and Major League Baseball, scientists have taken early steps to quantify the potential influence of climate warming on human conflict. Three researchers at the University of California, Berkeley, have pulled together data from these and other studies. They concluded that outbreaks of war and civil unrest may increase by as much as 56 percent by 2050 because of higher temperatures and extreme rainfall patterns predicted by climate change scientists. Likewise, episodes of personal violence – murder, assault, rape, domestic abuse – could increase by as much as 16 percent. Their study was published on Aug. 1 by the journal Science. "We find strong causal evidence linking climatic events to human conflict … across all major regions of the world," the researchers concluded.

TEXT 8: Scared and scarred by the global crisis, families hoard their money
Although they speak different languages, live in wealthy countries and poor ones, face good job markets and bad, when it comes to money they are acting as one. Families are holding tight to their cash, driven more by a fear of losing what they have than a desire to increase it. An Associated Press study of households in the 10 biggest economies shows that families continue to spend cautiously. They have pulled hundreds of billions of dollars out of the stock market and cut their borrowing for the first time in decades. They are putting their money into savings and investments that offer low interest payments, often too small to keep up with the cost of living increase each year. «It doesn't take very much to destroy confidence, but it takes an awful lot to build it back,» says Ian Bright, a senior economist at ING, a global bank based in Amsterdam.