Chapter 15. Pre-Editing and Post-Editing

Ana Guerberof Arenas

Abstract

This chapter provides an accessible introductory view of pre-editing and post-editing as the starting-point for research or work in the language industry. It describes source text pre-editing and machine translation post-editing from an industrial as well as academic point of view. In the last ten to fifteen years, there has been a considerable growth in the number of studies and publications dealing with pre-editing, and especially post-editing, that have helped researchers and the industry to understand the impact machine translation technology has on translators' output and their working environment. This interest is likely to continue in view of the recent developments in neural machine translation and artificial intelligence. Although the latest technology has taken a considerable leap forward, the existing body of work should not be disregarded as it has defined clear research lines and methods, as it is more necessary than ever to look at data in their appropriate context and avoid generalizing in the vast and diverse territory of human and machine translation.

Keywords: pre-editing; post-editing; MTPE; NMT; SMT; controlled language; CAT tools; productivity; quality; AEM; cognitive effort.

I. Introduction

This chapter describes pre-editing and post-editing concepts and processes derived from the use of MT in professional and academic translation environments as well as post-editing research methods. It will also summarize in two sections industry and academic research findings; however, a strict separation of this research is somewhat artificial since industry and academia often collaborate closely to gather data in this new field.

Although the pre-editing and post-editing of raw output (see the following section for a description of both editing concepts) have been implemented in some organizations since as early as the 1980s (by the Pan American Health Organization and the European Commission, for example), it is only in the last ten years that machine translation post-editing (PE) has been increasingly introduced as part of the standard translation workflow in most localization agencies worldwide (Lommel and DePalma 2016a).

The introduction of MT technology has caused some disruption in the translation community mainly due to the quality of the output to post-edit, on occasions too low to be of benefit, the time allowed for the PE task, with sometimes overly optimistic deadlines, and the price paid for the PE assignment, where frequently a level of discount on the rate per word is applied.

Since this was, and to some extent still is, recent technology, there is a need to gain more knowledge about how it affects the standard translation workflow and the agents involved in this process. Although this is a continuous endeavour, some key concepts and processes have emerged in recent years. This first section will offer a description of such concepts: controlled language and pre-editing, post-editing, types of PE (full and light), types of quality expected (human and “good enough”) and PE guidelines. It will close with a brief description of an MT output error typology.

1. Description of key concepts

MT is constantly evolving, and additional terms will be necessarily created. The concepts explained here are therefore derived from current uses of MT technology.

Controlled language and pre-editing

In this context, controlled language refers to certain rules applied when writing technical texts to avoid lexical ambiguity and complex grammatical structures, thus making it easier for the user to read and understand the text and, consequently, easier to apply technology such as translation memories (TMs) or MT. Controlled language focuses mainly on vocabulary and grammar and is intended for very specific domains, even for specific companies (O'Brien 2003). Texts have a consistent and direct style as a result of this process, and they are easier and cheaper to translate.

Pre-editing involves the use of a set of terminological and stylistic guidelines or rules to prepare the original text before translation automation to improve the raw output quality. Pre-editors follow certain rules, not only to remove typographical errors and correct possible mistakes in the content, but also to write shorter sentences, to use certain grammatical structures (simplified word order and less passive voice, for example) or semantic choices (the use of consistent terminology), or to mark certain terms (product names, for example) that might not require translation. There is software available, such as Acrolinx[1], that helps pre-edit the content automatically by adding terminological, grammatical and stylistic rules to a database that is then run in the source text.

Post-editing and automatic post-editing

To post-edit can be defined as 'to edit, modify and/or correct pre-translated text that has been processed by an MT system from a source language into (a) target language(s)'

[1] https://www.acrolinx.com/

(Allen 2003: 296). In other words, PE means to review a pre-translated text generated by an MT engine against an original source text, correcting possible errors to comply with specific quality criteria.

It is important to underline that quality criteria should be defined with the customer before the PE task. Another important consideration is that the purpose of the task is to increase productivity and to accelerate the translation process. Thus, PE also involves discarding MT segments that will take longer to post-edit than to translate from zero (scratch) or to process using an existing TM. Depending on the quality provided by the MT engine and the level of quality expected by the customer, the output might require translating again from scratch (if it is not useful), correcting many errors, correcting a few errors or simply accepting the proposal without any change.

Finally, automatic PE (APE) is a process by which the output is corrected automatically before sending the text for human PE or before publication, for example by using a set of regular expressions based on error patterns found in the MT output or by using data-driven approaches (Chatterjee et al. 2015) with the goal of reducing the human effort on the final text.

Types of post-editing: light and full

In MT and PE, it is common to classify texts as texts for assimilation —to roughly understand the text in another language— or texts for dissemination — to publish a text for a wide audience into several languages. Depending on this classification, the level of PE will vary. In the first case, assimilation, the text needs to be understandable and accurate, but style is not fundamental and some grammatical and spelling errors are even permitted. In the second case, dissemination, the text needs to be understandable and accurate, but also the style, grammar, spelling and terminology need to be comparable to the standard produced by a human translator.

Therefore, PE is also classified according to the degree of editing: full PE —human quality— and rapid or light PE —minimal corrections for text 'gisting'. Between full and light PE, there might be a wide range of options that are a result of discussions with the end customer before the task begins. In other words, PE can be as thorough or as superficial as required, although the degree of PE might be difficult to implement by a human post-editor used to correcting all errors in a text.

Types of quality expected and post-editing guidelines

Establishing the quality expected by the customer will help to determine the price as well as the specific instructions for post-editors. If this is not done, some translators might correct only major errors, thinking that they are obliged to utilize the MT proposal as much as possible, while others will correct major, minor and even acceptable proposals because they feel the text must be as human as possible. In general terms, customers know their users and the type of text they want to produce. In consequence, post-editors should be made aware of the expected quality with specific PE guidelines.

The TAUS Guidelines [2] are widely followed in the localization industry, with slight changes and adaptations. They distinguish between two types of quality: 'high-quality human translation,' where full PE is recommended, and 'good enough' or 'fit for purpose,' where light PE is recommended. Way (2013) explains that this dichotomy is too narrow to explain the different uses of MT and the expected quality ranges. He

[2] www.translationautomation.com/joomla/

refers to various levels of quality according to the fitness for purpose of translations and the perishability of content; he emphasizes that the degree of PE is directly related to the content lifespan (Way 2013: 2).

Whatever the result expected, the quality of the output is key in determining the type of PE: light PE applied to very poor output might result in an incomprehensible text, while light PE applied to very good output might result in a publishable text. Therefore, since the quality of the output is the main determiner of the necessary PE effort, the TAUS PE guidelines suggest focusing on the final quality of the text.

The guidelines for good enough quality are:

a. Aim for semantically correct translation.

b. Ensure that no information has been accidentally added or omitted.

c. Edit any offensive, inappropriate or culturally unacceptable content.

d. Use as much of the raw MT output as possible.

e. Basic rules regarding spelling apply.

f. No need to implement corrections that are of a stylistic nature only.

g. No need to restructure sentences solely to improve the natural flow of the text.

While for human quality, they are:

a. Aim for grammatically, syntactically and semantically correct translation.

b. Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of "Do Not Translate" terms.

c. Ensure that no information has been accidentally added or omitted.

d. Edit any offensive, inappropriate or culturally unacceptable content.

e. Use as much of the raw MT output as possible.

f. Apply basic rules regarding spelling, punctuation and hyphenation.

g. Ensure that formatting is correct.

These are general guidelines; the specific guidelines will depend on the project specifications: the language combination, the type of engine, the quality of output, the terminology, etc. It is necessary that the guidelines are language specific to provide real examples to post-editors of what can or cannot be changed. The guidelines could cover the following areas:

a. Description of the type of engine used.

b. Description of the source text, its type and structure.

c. Brief description of the quality of output for that language combination. This could include automatic scores or human evaluation scores.

d. Expected final quality to be delivered to the customer.

e. Scenarios for when to discard segments, so that post-editors should have an idea of how much time to spend to 'recycle' a segment or whether to discard it altogether.

f. Typical types of errors for that language combination that should be corrected, including reference to difficult areas like tagging, links or omissions.

g. Changes to be avoided in accordance with the customer's expectations; for example, certain stylistic changes that might not be necessary.

h. How to deal with terminology. The terminology provided by MT could be perfect or it could be obsolete, or a combination of the two. The advantages of PE, in terms of productivity, could be offset by terminological changes if the engine has been trained with old terminology or if the engine is not delivering accurate terms.

MT output error typology

There are many MT error classifications, the aim of which is not only to improve MT output by providing feedback to the engines but also to raise awareness amongst post-editors on the types of errors they will need to correct. If they know the types of errors frequently found in that output before performing this task, it is easier for post-editors to spot and correct them. Moreover, unnecessary changes are avoided and less frustration is generated when errors that are unusual for human translations are found.

In the last twenty years, there have been different error classifications depending on the type of engine, language pair and aim of the study. For example, Laurian (1984), Loffler-Laurian (1983), Krings (2001), Schäffer (2003), Vilar et al, (2006) and Farrús et al. (2011) offer very extensive error classifications adapted to the type of engine and language combination in use with the aim of understanding in which linguistic areas the machine encounters problems. However, these lists, while useful for the purpose intended, are not practical for the evaluation of MT errors in commercial environments.

Moreover, as the quality of the MT output improves, and MT becomes fully integrated in the standard localization workflow, error typologies like the ones applied to human translations are being used, such as the LISA Quality Model[3], SAE J2450[4], or BS EN15038[5]. Recently, and in a similar vein, TAUS has created its Dynamic Quality Evaluation Model[6] (see also Görög 2014), which is widely used in the localization industry and for PE research. The model allows users to measure MT productivity, rank MT engines, evaluate adequacy and fluency and/or classify output errors. Finally, the Multidimensional Quality Metrics (MQM)[7] is an error typology metric that was developed as part of the EU-funded QTLaunchPad project based on the examination and extension of more than one hundred existing quality models.

[3] Localization Industry Standards Association (LISA). http://dssresources.com/news/1558.php
[4] Society of Automotive Engineers Task Force on Translation Quality Metric. http://www.sae.org/standardsdev/j2450p1.htm
[5] EN-15038 European quality standard http://qualitystandard.bs.en-15038.com

[6]Dynamic Quality Framework (TAUS). https://dqf.taus.net/
[7] Multidimensional Quality Metrics. http://www.qt21.eu/launchpad/content/multidimensional-quality-metrics

These frameworks have common error classifications: Accuracy (referring to misunderstandings of the source text), Language (referring to grammar and spelling errors), Terminology (referring to errors showing divergence from the customer's glossaries), Style (referring to errors showing divergence from the customer's style guide), Country Standards (referring to errors related to dates, currency, number formats and addresses), Format (referring to errors in page layout, index entries, links and tagging). Severity levels (from 1 to 3, for example) can be applied according to the importance of the error found. These frameworks can be customized to suit a project, specific content, an LSP or a customer.

Recently, a new metric for error typology has been developed based on the harmonization of the MQM and DQF frameworks, and available through an open TAUS DQF API. This harmonization allows errors to be classified, firstly, according to a broader DQF error typology and, subsequently, by the subcategories as defined in MQM.

II. Research focal points

This section will give an overview of the research carried out in this field. It will start with a description of methods used to gather data, both quantitative and qualitative, such as screen recording, keyboard logging, eye-tracking, think aloud protocols, interviews, questionnaires and specifically designed software. Following this description, the principal areas of research will be discussed.

1. Research methods

In the last decade, there has been an increase in PE research as the commercial demand for this service has increased (O'Brien and Simard 2014: 159). As PE was a relatively new area of study, there were many factors to analyze that intended to answer if MT, at its core, was a useful tool for translators, and if by using this tool translators would be faster while maintaining the same quality. At the same time, new methods were needed to gather relevant data since all actions were happening 'inside a computer' and 'inside the translators' brain'. Therefore, methods were sometimes created specifically for Translation Studies (TS) and, at other times, methods from related areas of knowledge (primarily psychology and cognitive sciences) were applied experimentally to translation. This is a summary of the methods used for PE research.

Screen recording

With this method, software installed in a computer can record all screen and audio activity and creates a video file for subsequent analysis. Screen recording is used in PE experiments so that post-editors can use their standard tools (rather than an experimental tool designed specifically for research) and at the same time record their actions (some examples of screen-recording in PE research can be found in De Almeida and O'Brien 2010, Teixeira 2011 and Läubli et al. 2013). There are several software applications available in the market such as CamStudio[8], Camtasia Studio[9] and Flashback[10]. These applications change and improve as new generations of computers come onto the market.

[8] http://camstudio.org/
[9] https://camtasia-studio.softonic.com/
[10] https://www.flashbackrecorder.com/

Keyboard logging

There are many software applications that allow the computer to register the users' mouse and keyboard actions. Translog-II[11] is a software program created by CRITT[12] that is used to record and study human reading and writing processes on a computer; it is extensively used in translation process research. Translog-II produces log files which contain user activity data of reading, writing and translation processes, and which can be evaluated by external tools (Carl 2012).

Eye-tracking

This is a method that allows researchers to measure the user's eye movements and fixations on a screen using a light source and a camera generally integrated in a computer to capture the pupil centre corneal reflection (PCCR). The data gathered is written to a file that can be then read with eye-tracking software. Eye data (fixations, gaze direction) can indicate the cognitive state and load of a research participant. One of the most popular software applications available on the market is Tobii Studio Software,[13] used for the analysis and visualization of data from screen-based eye trackers.

Eye-tracking has been used in readability, comprehension, writing, editing and usability research. More importantly, it is regarded as an adequate tool to measure cognitive effort in MT and PE studies (Doherty and O'Brien 2009; Doherty et al. 2010) and has been used (and is being used) in numerous studies dealing with the PE task

[11] https://sites.google.com/site/centretranslationinnovation/translog-ii

[12] Center for Research and Innovation in Translation and Translation Technology

[13] https://www.tobiipro.com/product-listing/tobii-pro-studio/

(Alves et al. 2016; Carl et al. 2011; Doherty and O'Brien 2014; O'Brien 2006, 2011; Vieira 2016).

Think Aloud Protocols

Think aloud protocols (TAP) have been used in psychology and cognitive sciences, and then applied to TS to acquire knowledge on translation processes from the point of view of practitioners. They involve translators thinking aloud as they perform a set of specified tasks. Participants are asked to say what comes into their mind as they complete the task. This might include what they are looking at, thinking, doing and feeling. TAPs can also be post-hoc, that is the participants explain their actions after seeing an on-screen recording or gaze replay of their task. This is known as Retrospective Think Aloud (RTA) or Retrospective Verbal Protocols (RVP).

All verbalizations are transcribed and then analyzed. TAPs have been used extensively to define translation processes (Jääskeläinen and Tirkonnen Condit 1991; Lörscher 1991; Krings 2001; Kussmaul and Tirkonnen Condit 1995; Séguinot 1996). Over time, some drawbacks have been identified regarding TAPs, mainly that by having to explain the actions, the time to complete these actions is obviously altered, the fact that professional translators think faster than their verbalization and that the act of verbalizing the actions might affect cognitive processes (Jakobsen 2003; O'Brien 2005). However, this is not to say that TAPs have been discarded as a research tool. Vieira (2016), for instance, shows that TAPs correlate with eye movements and subjective ratings as measures of cognitive effort.

Interviews

Researchers use interviews to elicit information through questions from translators. This technique is used in PE research especially to find out what translators think about a process, a tool or their working conditions (for example, in Guerberof 2013; Moorkens and O'Brien 2013; Sanchís-Trilles et al. 2014; Teixeira 2014). Evidently, this does not necessarily mean that the interviewees say what they do or what they really believe in, but interviews are especially useful when using a mixed method approach to supplement hard data with translators' opinions and thoughts. This technique is increasingly used in TS and thus also in PE research as 'the discipline expands beyond the realm of linguistics, literature and cultural studies, and takes upon itself the task of integrating the sociological dimension of translation' (Saldanha and O'Brien 2013: 168). For example, a shift in the focus from texts and translation processes to status in society, working spaces and conditions, translators' opinions and perceptions of technology, to name but a few. Interviews, like questionnaires, require careful planning to obtain meaningful data. As with other methods, there is software to help researchers transcribe interviews (such as Google Docs, Dedoose[14], Dragon[15]), and to then carry out qualitative analysis by coding the responses (NVivo[16], MAXQDA[17], ATLAS.ti[18], and Dedoose).

---

[14] http://www.dedoose.com/
[15] http://www.nuance.es/dragon/index.htm
[16] http://www.qsrinternational.com/nvivo-spanish
[17] http://www.maxqda.com/
[18] http://atlasti.com/

Questionnaires

Another way of eliciting information from participants in PE research is to use questionnaires. As with interviews, creating a solid questionnaire is not an easy task, and it needs to be carefully researched, planned and tested. Questionnaires have been used extensively in PE research to find out, through open or closed questions, more about a translator's profile (gender, age, background and experience, for example), to gather information after the actual experiment on opinions about a tool, or to know a participant's attitude towards or perception of MT. The results can be used to clarify quantitative data (see Castilho et al. 2014; De Almeida 2013; Gaspari et al. 2014; Guerberof 2009; Krings 2001; Mesa-Lao 2014; Sanchís-Trilles et al. 2014). Google Forms or Survey Monkey[19] are two of the applications frequently used because they facilitate data distribution, gathering and analysis.

On-line tools for post-editing

Several tools have been developed as part of academic projects that were designed specifically to measure PE activities or to facilitate this task for professional post-editors. These include: CASMACAT (Ortiz-Martínez et al. 2012), MateCat (Federico et al. 2012), Appraise (Federmann 2012), iOmegaT (Moran 2012), TransCenter (Denkowski and Lavie 2012), PET Post-Editing Tool (Aziz et al. 2012), ACCEPT Post-Editing Environment (Roturier et al. 2013) and Kanjingo (O'Brien et al. 2014)

[19] https://www.surveymonkey.com/

As we saw in the section on MT error typology, there are also private initiatives that have led to the creation of tools used in PE tasks that have been also used in research such as the TAUS Dynamic Quality Framework (Görög 2014).

As well as these tools, standard CAT (Computer Assisted Translation) tools such as SDL Trados[20], Kilgray MemoQ[21], MemSource[22] or Lilt [23]are also used for research, especially if the objective is to have a working environment as close as possible to that of a professional post-editor. Moreover, these tools have implemented the use of APIs that allow users to connect to the MT engine of their choice, bringing MT closer to the translators and allowing them to integrate MT output seamlessly into their workflows.

2. Research areas

Until very recently, a lot of information about PE was company specific, since companies carried out their own internal research using their own preferred engines, processes, PE guidelines and desired final target text quality. This information stayed within the company as it was confidential. TS as a discipline was not particularly interested in the world of computer aided translation or MT until around the early 2000s.

Between the 1980s and 1990s, there were several articles published with information on MT implementation in different organizations describing the different tasks, processes, and profiles in PE (Senez 1998; Vasconcellos 1986, 1989, 1992, 1993; Vasconcellos and León 1985; Wagner 1985, 1987,) and specifying the various levels of PE (Loffler-Laurain 1983). These articles were descriptive in nature, as they intended to explain how the implementation of MT worked within a translation cycle. They also

[20] http://www.sdl.com/software-and-services/translation-software/sdl-trados-studio/
[21] https://www.memoq.com/en/
[22] https://www.memsource.com/
[23] https://lilt.com/

gave some indication of the productivity gains under certain restricted workflows. The most extensive research on PE at the time was carried out by Krings in 2001. In his book *Repairing texts*, Krings focuses on the mental processes used in MT PE using TAP. He defines and tests temporal effort (words processed in a given time), cognitive effort (post-editors' mental processing) and technical effort (post-editors' physical actions such as keyboard or mouse activities) in a series of tests.

Since this initial phase, interest in MT and PE has grown exponentially, focusing mainly on the following topics: controlled authoring, pre-editing and PE effort – the effort here can be temporal, technical or cognitive; quality of post-edited material versus quality of human translated material; PE effort versus TM-match effort and translation effort – the effort here can also be temporal (to discover more about productivity), technical or cognitive; the role of experience in PE: experienced versus novice translators' performance in relation to quality and productivity; correlations between automatic evaluation metrics (AEM) and PE effort; MT confidence estimation (CE); interactive translation prediction (ITP); monolingual versus multilingual post-editors in community-based PE (with user-generated content); usability and PE; and MT and PE training.

III. Informing research through the industry

In the early 90s, the localization industry needed new avenues to deliver translated products for a very demanding market that required bigger volumes of words, a higher number of languages with faster delivery times and lower buying prices. Because the content was often repetitive –translators would recycle content by cutting and pasting—the idea of creating databases of bilingual data (TMs) began taking shape

(Arthern 1979; Kay 1980; Melby, 1982), before Trados GmbH commercialized the first versions of Trados for DOS and Multiterm for Windows 3.1 in 1992.

At the same time, and as mentioned in the previous section, institutions and private companies were implementing MT to either increase the speed or lower the costs of their translations while maintaining quality. Seminal articles appeared that paved the way to future research on PE (see 'Research areas' above).

Once the use of TMs was fully functional in the localization workflow, further avenues to produce more translations, in less time and at lower costs while maintaining quality, were also explored. In the mid-2000s, companies like Google, Microsoft, IBM, Autodesk, Adobe and SAP, to name but a few, not only created and contributed to the advance in MT technology (see section 2 above) but they also designed new translation workflows with the dedicated support of their language service providers (LSPs). These new workflows included pre-editing and PE of raw output in different languages, the creation of new guidelines to work in this environment and the training of translators to include this technology without disrupting production or release dates. Moreover, these changes impacted on how translators were paid.

Since then, many companies have presented results from the work done internally using MT in combination with PE (or without PE) in conferences such as Localization World, GALA and the TAUS forums, or in more specialized conferences such as AMTA, EAMT or the MT Summit. Often, because of confidentiality issues, the articles that have subsequently appeared are not able to present detailed data to explain the results; in other cases, measurements have been taken in live settings with too many variables to properly isolate causes. Nevertheless, the companies involved have access to big data with high volumes and many languages from real projects to test the use of MT and PE;

the results are therefore highly relevant in so far as they represent the reality of the industry.

In general, companies seek information on either productivity (with a specific engine, language or type of content), quality (with a focus on their own needs or their customers' needs if they are LSPs), post-editor profiling, systems improvement or MT ecosystems. Some relevant published articles are those by Adobe (Flournoy and Duran, 2009), Autodesk (Plitt and Masselot 2010; Zhechev 2012), Continental Airlines (Beaton and Contreras 2010), IBM (Roukos et al. 2011), Sybase (Bier and Herranz 2011), PayPal (Beregovaya and Yanishevsky 2010), CA (Paladini 2011) and WeLocalize (Casanellas and Marg 2014; O'Curran 2014). These articles report on high productivity gains in the translation cycle thanks to the use of MT and PE while not compromising the final quality. They reflect on the diversity of results depending on the engine, language combination, translator and content. Initial reports came from large corporations or government agencies, understandably since the initial investment at the time was high, but more recently, LSPs have also been reporting positive findings on the use of the technology as part of their own localization cycle.

As well as large companies, consultancy firms have shed some light on the PE task and on the use of MT within the language industry. These consultancy companies have used data and standard processes to support business decisions made in the localization industry, sometimes in collaboration with academia. This is the case of TAUS, a community of users and providers of translation technologies and services that have worked intensively, as mentioned above, on defining a common quality framework to evaluate MT output (see DQF). Common Sense Advisory (CSA), a research and consulting firm, has also published several articles about PE (De Palma 2013), MT technology, pricing models, MT technology providers (Lommel and De Palma 2016b), neural MT (Lommel 2017) and surveys on the use of PE in LSPs (Lommel and De Palma 2016a). The aim is to provide information for large and small LSPs on how to implement MT and to gather common practices and offer recommendations on PE. At the same time, they provide academia with information on the state of the art in MT and PE in the localization industry. Recently, Slator[24], a company that provides business intelligence for the language market through a website, publications and conferences, has released a report on Neural Machine Translation (NMT).

IV. Informing the industry through research

The language industry often lacks the time to invest in research, and, as mentioned in the previous section, their measurements are taken in live settings with so many variables that conclusions might be clouded. Therefore, it is essential that academic research focuses on aspects of the PE task where more time is available and a scientific method can be applied.

Academic researchers have frequently worked together with industry to analyze in depth how technology shapes translation processes, collaborating not only with MT providers and LSPs, but also with freelancers and users of technical translations. The bulk of the research presented here has therefore not been carried out in an isolated academic setting but together with commercial partners. The research presented below is necessarily incomplete and subjective as it is difficult to present all the research on PE done in the last ten years given the sharp increase in the number of publications. In the following, we therefore present a summary of the findings of those studies considered

[24] https://slator.com/

important for the community, innovative at the time or to be stepping stones for future research.

With respect to pre-editing and controlled language, studies indicate that, while pre-editing activities reduce the PE effort in the context of specific languages and engines, these changes are not always easy to implement, that there is an initial heavy investment and that, sometimes, PE is still needed after the pre-editing has been performed (Aikawa et al. 2007; Bouillon et al. 2014; Gerlach 2015; Gerlach et al. 2013; Miyata et al. 2017; O'Brien 2006; Temnikova 2010). Therefore, if pre-editing is to be applied to an actual localization project, careful evaluation of the specific project is necessary to establish if the initial investment on the pre-editing task will have a real impact on the PE effort.

A close look at PE in practice shows that the combination of MT and PE increases translators' productivity (Aranberri et al. 2014; Carl et al. 2011; De Sousa et al. 2011; De Sutter and Depraetere 2012; Federico et al. 2012; García 2010, 2011; Green et al. 2013; Guerberof 2012; Läubli et al. 2013; O'Brien 2006; Ortíz and Matamala 2016; Parra and Arcedillo 2015b; Tatsumi 2010) but this is often applicable to very specific environments with highly customized engines, to suitable content, to certain language combinations, to specific guidelines and/or to suitably trained translators and to those with open attitudes to MT. It is important to note the high inter-subject variability when looking at post-editors' productivity. This is relevant because it signals the difficulty of establishing standards for measuring time, which adds to the complexity of individual pricing as opposed to a general discount per project or language.

Even if productivity increases, an analysis of the quality of the product is needed. Studies have shown that reviewers see very little difference between the quality of post-edited segments (edited to human quality) and human translated segments (Carl et al.

2011; De Sutter and Depraetere 2012; Fiederer and O'Brien 2009; Green et al. 2013; Guerberof 2012; Ortíz and Matamala 2016), and in some cases the quality might be higher with MT.

In view of high inter-subject variability, other variables have been examined to explain this phenomenon; for example, experience is always identified as an important variable when productivity and quality are measured. However, there is no conclusive data when it comes to experience. It seems that post-editors work in diverse ways not only regarding the time, type and number of changes they make, but also according to experience (Aranberri et al. 2014; De Almeida 2013; De Almeida and O'Brien 2010; Guerberof 2012; Moorkens and O'Brien 2015). In some cases, professionals work faster and achieve higher quality. In others, experience is not a factor. It has also been noted that novice translators might have an open attitude towards MT that could facilitate training and acceptance of MT projects. Experienced translators, on the other hand, might do better in following PE guidelines, work faster and produce higher final quality, although this is not always the case.

Since TMs have been fully implemented in the language industry and prices have long been established, the correlation between MT output and fuzzy matches from TMs have also been studied to see if MT prices could be somehow mirrored (Guerberof 2009, 2012; O'Brien 2006; Parra and Arcedillo 2015a; Tatsumi 2010). The results suggest that there are correlations between MT output and TM segments above 75% in terms of productivity. It is surprising that overall processing of MT matches seems to correlate with high fuzzy matches rather than with low fuzzy matches. However, the fact that CAT tools highlight the required changes in TM matches, and not in the MT proposals, facilitates the work done by translators when using TMs.

In a comparable way, the correlation between automatic estimation metrics (AEM) and PE effort has been examined to see if, by analyzing the output automatically, it is easier to infer the post-editor's effort. This correlation, however, seems to be more accurate globally than on a per segment basis or for certain types of sentences, such as longer ones (De Sutter and Depraetere 2012; Guerberof 2012; O'Brien 2011; Parra and Arcedillo 2015b; Tatsumi 2010; Vieira 2014), and there are also discrepancies in matching automatic scores with actual productivity levels. Therefore, these metrics, although they can give an indication of the PE effort, might be difficult to apply in a professional context.

To bridge this gap and simulate the fuzzy-match scores that TMs offer, there have been several studies on MT confidence estimations (CEs) (He et al. 2010; Huang et al. 2014; Specia 2011; Specia et al. 2009a, 2009b; Turchi et al. 2015). CE is a mechanism by which post-editors are informed about the quality of the output in a comparable way to that of TMs. It enables translators to determine rapidly whether an individual MT proposal will be useful, or if it would be more productive to ignore it and translate from scratch, thus enhancing productivity. The results show that CEs can be useful to speed up the PE process, but this is not always the case. Translators' variability regarding productivity, the fact that these scores might not be fully accurate due to the technical complexity and the content itself have made it difficult to fully implement CE in a commercial context.

Another method created to aid post-editors is Interactive Translation Prediction (ITP) (Alves et al. 2016; Pérez-Ortíz et al. 2014; Sanchís-Trilles et al. 2014; Underwood et al. 2014). ITP helps post-editors in their task by changing the MT proposal according to the textual choices made in real time. Although studies show that the system does not necessarily increase productivity, it can reduce the number of keystrokes and cognitive effort without impacting on quality.

Even if productivity increases while quality is maintained, actual experience shows that PE is a tiring task for translators. Therefore, PE and cognitive effort have been explored from different angles. There are studies that highlight the fact that PE requires higher cognitive effort than translation and that describe the cognitive complexity in PE tasks (Krings 2001; O'Brien 2006, 2017). There is also research that explores which aspects of the output or the source text require a higher cognitive effort and are therefore problematic for post-editors. This effort seems to vary according to the target language structure; more instances of high effort are noted under these categories: incorrect syntax, word order, mistranslations and mistranslated idioms (Daems et al. 2015, 2017; Koponen 2012; Lacruz and Shreve 2014; Lacruz et al. 2012; Popovic et al. 2014; Temnikova 2010; Temnikova et al. 2016; Vieira 2014).

Users of MT, however, are not only translators. Some research looks at monolingual PE and finds that it can lead to improved fluency and comprehensibility scores like those achieved through bilingual PE; fidelity, however, improved considerably more with bilingual post-editors. As is usual in this type of research, performance across post-editors varies greatly (Hu et al. 2010; Koehn 2010; Mitchell et al. 2013; Mitchell et al. 2014; Nietzke 2016). MT and PE can be a useful tool for lay users with an appropriate set of recommendations and best practices. At the same time, even bilingual PE by lay users can help bridge the information gap in certain environments, such as that of health organizations (Laurenzi et al. 2013).

MT usability and PE has not been extensively explored yet, although there is relevant work that indicates that usability increases when users read the original text or even a lightly post-edited text as opposed to reading raw MT output (Castilho 2016; Castilho et al. 2014; Doherty and O'Brien 2012, 2014; O'Brien and Castilho 2016). However, users can complete most tasks by using the raw MT output even if the experience is less satisfactory, though the research results again vary considerably depending on the language and, therefore, the quality of the raw output.

Regarding MT and PE training for translators, the skills needed have been described by O'Brien (2002), Rico and Torrejón (2012) and Pym (2013), while syllabi have been designed, and courses explained and described, by Doherty et al. (2012), Doherty and Moorkens (2013), Doherty and Kenny (2014), Koponen (2015) and Mellinger (2017). The suggestions made include teaching basic MT technology concepts, MT evaluation techniques, Statistical MT (SMT) training, pre-editing and controlled language, monolingual PE, understanding various levels of PE (light and full), creating guidelines, MT evaluation, output error identification, when to discard unusable segments and continuous PE practice.

V. Concluding remarks

As this section has shown, research in PE is an area of considerable interest to academia and industry. As MT technology changes, for example through the development of NMT, findings need to be revisited to test how PE effort is affected and, indeed, if PE is necessary at all for certain products or purposes (for example, when dealing with forums, chats and knowledge bases). Logic would indicate that, as MT engines improve, the PE task for translators will become 'easier'; that is, translators would be able to process more words in less time and with a lower technical and cognitive effort. However, this will vary depending on the engine, the language combination and the domain. SMT models have been customized and long implemented in the language industry, with adjustments for specific workflows (customers, content, language combinations). It is still early days for NMT from a PE perspective, and even from a developer's perspective NMT is still in its infancy. Results have so far been encouraging (Bentivogli et al. 2016; Toral et al. 2018), though inconsistent (Castilho et al. 2017), when looking at NMT post-editing and human evaluation of NMT content. The inconsistent results are precisely due to SMT engine customizations, language combinations, domains and the way NMT processes certain segments that might sound fluent but contain serious errors (omissions, additions and mistranslations) that are less predictable than in SMT. Therefore, further collaboration between academia and industry will be needed to gain more knowledge on SMT versus NMT performance, NMT error typology, NMT treatment of long sentences and terminology, NMT integration in CAT tools, NMT for low resource languages and NMT acceptability to translators as well as end-users, to mention just some areas of special interest. Far from the hype of perfect MT quality[25], evidence to date shows that to deliver human quality translations using MT, human intervention is still very much needed.

Funding

This article was partially written under the Edge Research Fellowship programme that has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Marie Sklodowska-Curie grant agreement No. 713567, and by the ADAPT Centre for Digital Content Technology, funded under the SFI Research Centres Programme (Grant 13/RC/2106) and co-funded under the European Regional Development Fund.

[25] https://blogs.microsoft.com/ai/machine-translation-news-test-set-human-parity/?wt.mc_id=74788-mcr-tw

References

Aikawa, T., L. Schwartz, R. King, M. Corston-Oliver and C. Lozano (2007), 'Impact of Controlled Language on Translation Quality and Post-editing in a Statistical Machine Translation Environment', in *Proceedings of the MT summit XI*, 1–7, Copenhagen.

Allen, J. (2003), 'Post-editing', in Somers H. (ed.), *Computers and translation: A translator's guide*, 297–317, Amsterdam: John Benjamins.

Alves, F., A. Koglin, B. Mesa-Lao, M. G. Martínez, N. B. de Lima Fonseca, A. de Melo Sá, J. L. Gonçalves, K. Sarto Szpak, K. Sekino and M. Aquino (2016), 'Analysing the Impact of Interactive Machine Translation on Post-editing Effort', in M. Carl, S. Bangalore and M. Schaeffer (eds), *New directions in empirical translation process research*, 77–94, Cham: Springer.

Aranberri, N., G. Labaka, A. Diaz de Ilarraza and K. Sarasola (2014), 'Comparison of Post-editing Productivity Between Professional Translators and Lay Users', in S. O'Brien, M. Simard and L. Specia (eds), *Proceedings of the 3rd workshop on post-editing technology and practice*, 20–33, Vancouver: AMTA.

Arthern, P. J. (1979), 'Machine Translation and Computerized Terminology Aystems: A Translator's Viewpoint', B. M. Snell (ed.), *Translating and the Computer: Proceedings of a seminar,* 77–108, London: North-Holland.

Aziz, W., S. Castilho and L. Specia (2012), 'PET: A Tool for Post-editing and Assessing Machine Translation', in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, 3982–3987, Istanbul: European Language Resources Association (ELRA).

Beaton, A. and G. Contreras (2010), 'Sharing the Continental Airlines and SDL Post-editing Experience', in *Proceedings of the 9th Conference of the AMTA*, Denver: AMTA.

Bentivogli, L., A. Bisazza, M. Cettolo and M. Federico (2016), 'Neural Versus Phrase-based Machine Translation Quality: A Case Study', arXiv preprint arXiv:1608.04631.

Beregovaya, O. and A. Yanishevsky (2010), 'PROMT at PayPal: Enterprise-scale MT Deployment for Financial Industry Content', in *Proceedings of the 9th Conference of the AMTA*, Denver: AMTA.

Bier, K. and M. Herranz (2011), 'MT Experience at Sybase', *Localization World Conference*, Barcelona.

Bouillon, P., L. Gaspar, J. Gerlach, V. Porro Rodriguez and J. Roturier (2014), 'Pre-editing by Forum Users: A Case Study', *Workshop (W2) on Controlled Natural Language Simplifying Language Use*, in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 3–10, Reykjavik: European Language Resources Association (ELRA).

Carl, M. (2012), 'Translog-II: A Program for Recording User Activity Data for Empirical Reading and Writing Research', in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, 4108–12, Istanbul: European Language Resources Association (ELRA).

Carl, M., B. Dragsted, J. Elming, D. Hardt and A. L. Jakobsen (2011), 'The Process of Post-editing: A Pilot Study', in *Proceedings of the 8th International NLPSC Workshop. Special Theme: Human-machine Interaction in Translation,* Vol. 41, 131–42, Copenhagen: Copenhagen Studies in Language.

Casanellas, L. and L. Marg (2014), 'Assumptions, Expectations and Outliers in Post-editing', in *Proceedings of the 17th Annual Conference of the EAMT*, 93–7, Dubrovnik: EAMT.

Castilho, S. (2016), 'Measuring Acceptability of Machine Translated Enterprise Content', Doctoral thesis, Dublin City University, Dublin.

Castilho, S., S. O'Brien, F. Alves and M. O'Brien (2014), 'Does Post-editing Increase Usability? A Study with Brazilian Portuguese as Target Language', in *Proceedings of the 17th Conference of the EAMT*, 183–90, Dubrovnik: EAMT.

Castilho, S., J. Moorkens, F. Gaspari, I. Calixto, J. Tinsley and A. Way (2017), 'Is Neural Machine Translation the New State of the Art?', *The Prague Bulletin of Mathematical Linguistics*, 108 (1), 109–20.

Chatterjee, R., M. Weller, M. Negri and M. Turchi (2015), 'Exploring the Planet of the Apes: A Comparative Study of State-of-the-Art Methods for MT Automatic Post-editing', in *Proceedings of the 53rd Annual Meeting of the ACL and the 7th International Joint Conference on NLP,* 2, 156–61, Beijing: ACL.

Daems, J., S. Vandepitte, R. Hartsuiker and L. Macken (2015), 'The Impact of Machine Translation Error Types on Post-editing Effort Indicators', in S. O'Brien, M. Simard and L. Specia (eds) *Proceedings of the 4th Workshop onPpost-editing Technology and Practice (WPTP4)*, 31–45, Miami: AMTA.

Daems, J., S. Vandepitte, R. J. Hartsuiker and L. Macken (2017), 'Identifying the Machine Translation Error Types with the Greatest Impact on Post-editing Effort', *Frontiers in Psychology*, *8*, 1282. Available online: https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01282/full (accessed 21 September 2018).

De Almeida, G. (2013), 'Translating the Post-editor: An Investigation of Post-editing Changes and Correlations with Professional Experience Across two Romance Languages', Doctoral thesis, Dublin City University, Dublin.

De Almeida, G. and S. O'Brien (2010), 'Analysing Post-editing Performance: Correlations with Years of Translation Experience', in *Proceedings of the 14th Annual Conference of the EAMT,* 27–8, St. Raphaël: EAMT.

De Palma, D. A. (2013), 'Post-edited Machine Translation Defined', Common Sense Advisory. Available online: http://www.commonsenseadvisory.com/AbstractView/tabid/74/ArticleID/5499/Title/Post-EditedMachineTranslationDefined/Default.aspx (accessed 21 September 2018).

De Sousa, S. C., W. Aziz and L. Specia (2011), 'Assessing the Post-editing Effort for Automatic and Semi-automatic Translations of DVD Subtitles', in G. Angelova, K. Bontcheva, R. Mitkov and N. Nikolov (eds), *Proceedings of RANLP*, 97–103, Hissar.

De Sutter, N. and I. Depraetere (2012), 'Post-edited Translation Quality, Edit Distance and Fluency Scores: Report on a Case Study', Presentation in *Journée d'études Traduction et qualité Méthodologies en matière d'assurance qualité*, Universtité Lille 3, Sciences humaines et sociales, Lille.

Denkowski, M. and A. Lavie (2012), 'TransCenter: Web-based Translation Research Suite', in *AMTA 2012 Workshop on Post-editing Technology and Practice Demo Session,* San Diego: AMTA.

Doherty, S. and D. Kenny (2014), 'The Design and Evaluation of a Statistical Machine Translation Syllabus for Translation Students', *The Interpreter and Translator Trainer*, *8* (2), 295–315.

Doherty, S. and J. Moorkens (2013), 'Investigating the Experience of Translation Technology Labs: Pedagogical Implications', *Journal of Specialised Translation*, 19, 122–36.

Doherty, S. and S. O'Brien (2009), 'Can MT Output be Evaluated through Eye Tracking?', in *Proceedings of the MT Summit XII*, 214–21, Ottawa: AMTA.

Doherty, S. and S. O'Brien (2012), 'A User-based Usability Assessment of Raw Machine Translated Technical Instructions', in *Proceedings of the 10th Biennial Conference of the AMTA*, San Diego: AMTA.

Doherty, S. and S. O'Brien (2014), 'Assessing the Usability of Raw Machine Translated Output: A User-centered Study Using Eye Tracking', *International Journal of Human-Computer Interaction*, 30 (1), 40–51.

Doherty, S., D. Kenny and A. Way (2012), 'Taking Statistical Machine Translation to the Student Translator', in *Proceedings of the 10th biennial conference of the AMTA*, San Diego: AMTA.

Doherty, S., S. O'Brien and M. Carl (2010), 'Eye Tracking as an MT Evaluation Technique', *Machine Translation,* 24(1), 1–13, Netherlands: Springer.

Farrús, M., M. R. Costa-Jussa, J. B. Marino, M. Poch, A. Hernández, C. Henríquez and J. A. Fonollosa (2011), 'Overcoming Statistical Machine Translation Limitations: Error Analysis and Proposed Solutions for the Catalan–Spanish Language Pair', *Language Resources and Evaluation*, 45(2), 181–208, Netherlands: Springer.

Federico, M., A. Cattelan and M. Trombetti (2012), 'Measuring User Productivity in Machine Translation Enhanced Computer Assisted Translation', in *Proceedings of the 10th Conference of the AMTA,* 44–56, Madison: AMTA.

Federmann, C. (2012), 'Appraise: An Open-source Toolkit for Manual Evaluation of MT Output', *The Prague Bulletin of Mathematical Linguistics*, *98*, 25–35.

Fiederer, R. and S. O'Brien (2009), 'Quality and Machine Translation: A Realistic Objective? ', *The Journal of Specialised Translation* 11, 52–74.

Flournoy, R. and C. Duran (2009), 'Machine Translation and Document Localization at Adobe: From Pilot to Production', *Proceedings of the 12th MT Summit,* Ottawa.

García, I. (2010), ‘Is Machine Translation Ready Yet?’, *Target*, (22–1), 7–2, Amsterdam: John Benjamins.

García, I. (2011), ‘Translating by Post-editing: Is it the Way Forward?’, *Machine Translation*, 25(3), 217–37, Netherlands: Springer.

Gaspari, F., A. Toral, S. K. Naskar, D. Groves and A. Way (2014), ‘Perception vs Reality: Measuring Machine Translation Post-editing Productivity’, in S. O’Brien, M. Simard and L. Specia (eds), *Proceedings of the 3rd Workshop on Post-editing Technology and Practice*, 60–72, Vancouver: AMTA.

Gerlach, J. (2015), ‘Improving Statistical Machine Translation of Informal Language: A Rule-based Pre-editing Approach for French Forums’, Doctoral thesis, University of Geneva, Geneva.

Gerlach, J., V. Porro Rodriguez, P. Bouillon and S. Lehmann (2013), ‘Combining Pre-editing and Post-editing to Improve SMT of User-generated Content’, in S. O’Brien, M. Simard and L. Specia (eds), *Proceedings of the 2nd Workshop on Post-editing Technology and Practice*, 45–53, Nice: EAMT.

Görög, A. (2014), ‘Quality Evaluation Today: The Dynamic Quality Framework’, in *Proceedings of Translating and the Computer 36*, 155–64, Geneva: Editions Tradulex.

Green, S., J. Heer and C. D. Manning (2013), ‘The Efficacy of Human Post-editing for Language Translation’, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 439–48, Paris: ACM.

Guerberof, A. (2009), ‘Productivity and Quality in MT Post-editing’, *MT summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT*, Ottawa: EAMT.

Guerberof, A. (2012), ‘Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation’, Doctoral thesis, Universitat Rovira i Virgili, Tarragona.

Guerberof, A. (2013), ‘What do Professional Translators Think about Post-editing’, *The Journal of Specialized Translation* 19, 75–95.

He, Y., Y. Ma, J. Van Genabith and A. Way (2010), ‘Bridging SMT and TM with Translation Recommendation’, in *Proceedings of the 48th Annual Meeting of ACL,* 622–30, Uppsala: ACL.

Hu, C., B. B. Bederson and P. Resnik (2010), ‘Translation by Iterative Collaboration Between Monolingual Users’, in *Proceedings of Graphics Interface,* 39–46, Ottawa: Canadian Information Processing Society.

Huang, F., J. M. Xu, A. Ittycheriah and S. Roukos (2014), ‘Improving MT Post-editing Productivity with Adaptive Confidence Estimation for Document-specific Translation Model’, *Machine Translation*, *28*(3–4), 263–80, Netherlands: Springer.

Jääskeläinen, R. and S. Tirkkonen-Condit (1991), ‘Automatised Processes in Professional vs. Non-professional Translation: A Think-aloud Protocol Study’, *Empirical Research in Translation and Intercultural Studies,* 89–109, Tübingen: Narr.

Jakobsen, A. L. (2003), ‘Effects of Think-aloud on Translation Speed, Revision and Segmentation’, F. Alves (ed.)*, Triangulating translation: Perspectives in Process-oriented Research*, 45, 69–95, Amsterdam: John Benjamins.

Kay, M. (1980/97), ‘The Proper Place of Men and Machines in Language Translation’, Research Report CSL-80-11, Xerox Palo Alto Research Center. Reprinted in *Machine Translation,* 12(1997), 3–23, Netherlands: Springer.

Koehn, P. (2010), ‘Enabling Monolingual Translators: Post-editing vs. Options’, in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL,* 537–545, Los Angeles: ACL.

Koponen, M. (2015), ‘How to Teach Machine Translation Post-editing? Experiences from a Post-editing Course’, in S. O’Brien, M. Simard and L. Specia (eds), *Proceedings of the 4th Workshop on Post-editing Technology and Practice (WPTP 4*), 2–15, Miami: AMTA.

Koponen, M., W. Aziz, L. Ramos and L. Specia (2012), ‘Post-editing Time as a Measure of Cognitive Effort’, in S. O’Brien, M. Simard and L. Specia (eds), *Proceedings AMTA 2012 Workshop on Post-editing Technology and Practice*, 11–20, San Diego: AMTA.

Krings, H. P. (2001), *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes*, (Vol. 5). G. S. Koby, (ed.), Kent, OH: Kent State University Press.

Kussmaul, P. and S. Tirkkonen-Condit (1995), ‘Think-aloud Protocol Analysis in Translation Studies’, in TTR 8, 177–99, Association canadienne de traductologie.

Lacruz, I. and G. M. Shreve (2014), ‘Pauses and Cognitive Effort in Post-editing.’ in *Post-editing of Machine Translation: Processes and Applications*, S. O’Brien, L. W. Balling, M. Simard, L. Specia and M. Carl (eds), 246–72, Newcastle upon Tyne: Cambridge Scholars Publishing.

Lacruz, I., G. M. Shreve and E. Angelone (2012), ‘Average Pause Ratio as an Indicator of Cognitive Effort in Post-editing: A Case Study.’ in S. O’Brien, M. Simard and L. Specia (eds), *Proceedings of the AMTA 2012 Workshop on Post-editing Technology and Practice*, 21–30, San Diego: AMTA.

Läubli, S., M. Fishel, G. Massey, M. Ehrensberger-Dow and M. Volk (2013), ‘Assessing Post-editing Efficiency in a Realistic Translation Environment’, in *Proceedings of the 2nd Workshop on Post-editing Technology and Practice*, 83–91, Nice: EAMT.

Laurenzi, A., M. Brownstein, A. M. Turner and K. Kirchhoff (2013), 'Integrated Post-editing and Translation Management for Lay User Communities', in S. O'Brien, M. Simard and L. Specia (eds), *Proceedings of the 2nd Workshop on Post-editing Technology and Practice*, 27–34, Nice: EAMT.

Laurian, A. M. (1984), 'Machine Translation: What Type of Post-editing on What Type of Documents for What Type of Users', *Proceedings of the 10th International Conference on Computational Linguistics and 22nd Annual Meeting on ACL*, 236–8, Stanford: ACL.

Loffler-Laurian, A. M. (1983), 'Pour une typologie des erreurs dans la traduction automatique'. [Towards a Typology of Errors in Machine Translation], *Multilingua (*2), 65–78.

Lommel, A. R. (2017), 'Neural MT: Sorting Fact from Fiction', Common Sense Advisory. Available online: http://www.commonsenseadvisory.com/AbstractView/tabid/74/ArticleID/37893/Title/NeuralMTSortingFactfromFiction/Default.aspx (accessed 21 September 2018).

Lommel, A. R. and D. A. De Palma (2016a), 'Post-editing Goes Mainstream', Common Sense Advisory. Available online: http://www.commonsenseadvisory.com/Portals/_default/Knowledgebase/ArticleImages/1605_R_IP_Postediting_goes_mainstream-extract.pdf (accessed 21 September 2018).

Lommel, A. R. and D. A. De Palma (2016b), 'TechStack: Machine Translation', Common Sense Advisory. Available online: http://www.commonsenseadvisory.com/Portals/_default/Knowledgebase/ArticleImages/1611_QT_Tech_Machine_Translation-preview.pdf (accessed 21 September 2018).

Lörscher, W. (1991), 'Thinking-aloud as a Method for Collecting Data on Translation Processes', *Empirical Research in Translation and Intercultural Studies: Selected Papers of the TRANSIF Seminar*, Savonlinna 1988, 67–77, Tübingen: G. Narr.

Melby, A. K. (1982), 'Multi-level Translation Aids in a Distributed System', *Proceedings of the 9th Conference on Computational Linguistics,* 215–20, Academia Praha.

Mellinger, C. D. (2017), 'Translators and Machine translation: Knowledge and Skills Gaps in Translator Pedagogy', *The Interpreter and Translator Trainer*, 1–14, Taylor & Francis Online.

Mesa-Lao, B. (2014), 'Speech-enabled Computer-aided Translation: A Satisfaction Survey with Post-editor Trainees', in *Proceedings of the Workshop on Humans and Computer-assisted Translation*, 99–103, Gothenburg: ACL.

Mitchell, L., S. O'Brien and J. Roturier (2014), 'Quality Evaluation in Community Post-editing', *Machine translation*, *28* (3–4), 237–62, Netherlands: Springer.

Mitchell, L., J. Roturier and S. O'Brien (2013), 'Community-based Post-editing of Machine-translated Content: Monolingual vs. Bilingual', in S. O'Brien, M. Simard and L. Specia (eds), *Proceedings of the 2nd Workshop on Post-editing Technology and Practice,* 35–45, Nice: EAMT.

Miyata, R., A. Hartley, K. Kageura and C. Paris (2017), 'Evaluating the Esability of a Controlled Language Authoring Assistant', *The Prague bulletin of mathematical linguistics*, *108* (1), 147–58.

Moorkens, J. and S. O'Brien (2013), 'User Attitudes to the Post-editing Interface', in S. O'Brien, M. Simard and L. Specia (eds), *Proceedings of 2nd Workshop on Post-editing Technology and Practice,* 19–25, Nice: EAMT.

Moorkens, J. and S. O'Brien (2015), 'Post-editing Evaluations: Trade-offs Between Novice and Professional Participants', in *Proceedings of EAMT*, 75–81, Antalya: EAMT.

Moran, J. (2012), ‘Experiences Instrumenting an Open-source CAT Tool to Record Translator Interactions’, in *Expertise in Translation and Post-editing. Research and Application*, L. W. Balling, M. Carl and A. L. Jakobsen, Copenhagen: Copenhagen Business School.

Nitzke, J. (2016), ‘Monolingual Post-editing: An Exploratory Study on Research Behaviour and Target Text Quality’, *Eyetracking and Applied Linguistics*, *2*, 83–108, Berlin: Language Science Press.

O’Brien, S. (2002), ‘Teaching Post-editing: A Proposal for Course Content’, in *Proceedings for the 6th Annual EAMT Conference. Workshop Teaching Machine Translation,* 99–106, Manchester: EAMT.

O’Brien, S. (2003), ‘Controlling Controlled English: An Analytical of Several Controlled Language Rule Sets’, in *Proceedings of EAMT-CLAW 2003*, 105–14, Dublin: EAMT.

O’Brien, S. (2005), ‘Methodologies for Measuring the Correlations Between Post-editing Effort and Machine Translatability’, *Machine translation,* 19 (1), 37–58, Netherlands: Springer.

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

O’Brien, S. (2011), ‘Towards Predicting Post-editing Productivity’, *Machine translation*, *25*(3), 197–215, Netherlands: Springer.

O’Brien, S. (2017), ‘Machine Translation and Cognition’, *The Handbook of Translation and Cognition*, J. W. Schwieter and A. Ferreira (eds), 311–31, Wiley-Blackwell.

O’Brien, S. and S. Castilho (2016), ‘Evaluating the Impact of Light Post-editing on Usability’, in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, 310–6, Portorož: European Language Resources Association (ELRA).

O'Brien, S. and M. Simard (2014), 'Introduction to Special Issue on Post-editing', *Machine Translation*, 28 (3–4), Netherlands: Springer.

O'Brien, S., J. Moorkens and J. Vreeke (2014), 'Kanjingo–A Mobile App for Post-editing', in S. O'Brien, M. Simard and L. Specia (eds), *Proceedings of the 3rd Workshop on Post-editing Technology and Practice*, 125–7, Vancouver: AMTA.

O'Curran, E. (2014), 'Translation Quality in Post-edited Versus Human-translated Segments: A Case Study', in S. O'Brien, M. Simard and L. Specia (eds), *Proceedings of the 3rd Workshop on Post-editing Technology and Practice*, 113–18, Vancouver: AMTA.

Ortiz Boix, C. and A. Matamala (2016), 'Implementing Machine Translation and Post-editing to the Translation of Wildlife Documentaries through Voice-over and Off-screen Dubbing', Doctoral thesis, Universitat Autònoma de Barcelona, Barcelona.

Ortiz-Martınez, D., G. Sanchís-Trilles, F. Casacuberta, V. Alabau, E. Vidal, J. M. Benedı and J. González (2012), 'The CASMACAT Project: The Next Generation Translator's Workbench', *Proceedings of the 7th jornadas en tecnología del habla and the 3rd Iberian SLTech Workshop (IberSPEECH)*, 326–34.

Paladini, P. (2011), 'Translator's Productivity Increase at CA Technologies', *Localization World Conference*, Barcelona.

Parra Escartín, C. and M. Arcedillo (2015a), 'A Fuzzier Approach to Machine Translation Evaluation: A Pilot Study on Post-editing Productivity and Automated Metrics in Commercial Settings', in *Proceedings of the 4th Workshop on Hybrid Approaches to Translation HyTra@ ACL*, 40–5, Beijing: ACL.

Parra Escartín, C. and M. Arcedillo (2015b), 'Living on the Edge: Productivity Gain Thresholds in Machine Translation Evaluation Metrics', in S. O'Brien, M. Simard and L. Specia (eds), *Proceedings of the 4th Workshop on Post-editing Technology and Practice (WPTP4)*, 46, Miami: AMTA.

Pérez-Ortiz, J. A., D. Torregrosa and M. L. Forcada (2014), 'Black-box Integration of Heterogeneous Bilingual Resources into an Interactive Translation System', in *Workshop on Humans and Computer-assisted Translation*, 57–65, Gothenburg: ACL.

Plitt, M. and F. Masselot (2010), 'A Productivity Test of Statistical Machine Translation Post-editing in a Typical Localisation Context', *The Prague Bulletin of Mathematical Linguistics*, 93, 7–16.

Popović, M., A. R. Lommel, A. Burchardt, E. Avramidis and H. Uszkoreit (2014), 'Relations Between Different Types of Post-editing Operations, Cognitive Effort and Temporal Effort', *The 7th Annual Conference of the EAMT,* 191–198, Dubrovnik: EAMT.

Pym, A. (2013), 'Translation Skill-sets in a Machine-translation Age', *Meta*, 58 (3), 487–503.

Rico, C. and E. Torrejón (2012), 'Skills and Profile or the New Role of the Translator as MT Post-editor. Postedició, Canvi de Paradigma?', *Revista Tradumàtica: technologies de la traducció*, 10, 166–78.

Roturier, J., L. Mitchell, D. Silva and B. B. Park (2013), 'The ACCEPT Post-editing Environment: A Flexible and Customisable Online Tool to Perform and Analyse Machine Translation Post-editing', in S. O'Brien, M. Simard and L. Specia (eds), *Proceedings of 2nd Workshop on Post-editing Technology and Practice*, 119–28, Nice: EAMT.

Roukos, S. R., F. M. Xu, J. P. Nesta, S. M. Corriá, A. Chapman and H. S. Vohra (2011), 'The Value of Post-editing: IBM Case Study', Localization World Conference, Barcelona.

Saldanha, G. and S. O'Brien (2013), *Research Methodologies in Translation Studies,* London: Routledge.

Sanchís-Trilles, G., V. Alabau, C. Buck, M. Carl, F. Casacuberta, M. García-Martínez and L. A. Leiva (2014), 'Interactive Translation Prediction Versus Conventional Post-editing in Practice: A Study with the CasMaCat Workbench', *Machine Translation*, 28 (3–4), 217–35, Netherlands: Springer.

Schäffer, F. (2003), 'MT Post-editing: How to Shed Light on the "Unknown Task". Experiences at SAP', *Controlled Language Translation, EAMT-CLAW-03*, 133–40, Dublin: EAMT.

Séguinot, C. (1996), 'Some Thoughts about Think-aloud Protocols', *Target* 8, 75–95.

Senez, D. (1998), 'The Machine Translation Help Desk and the Post-editing Service', *Terminologie et Traduction 1*, 289–95.

Specia, L. (2011), 'Exploiting Objective Annotations for Measuring Translation Post-editing Effort', *Proceedings of the 15th annual EAMT Conference*, 73–80, Leuven: EAMT.

Specia, L., N. Cancedda, M. Dymetman, M. Turchi and N. Cristianini (2009a), 'Estimating the Sentence-level Quality of Machine Translation Systems', *Proceedings of the 13th Annual Conference of the EAMT,* 28–35, Barcelona: EAMT.

Specia, L., C. Saunders, M. Turchi, Z. Wang and J. Shawe-Taylor (2009b), 'Improving the Confidence of Machine Translation Quality Estimates', in *Proceedings of the 12th MT Summit,* 136–43, Ottawa: AMTA.

Tatsumi, M. (2010), 'Post-editing Machine Translated Text in a Commercial Setting: Observation and Statistical Analysis', Doctoral thesis, Dublin City University, Dublin.

Teixeira, C. (2011), 'Knowledge of Provenance and its Effects on Translation Performance in an Integrated TM/MT Environment', in *Proceedings of the 8th International NLPCS workshop–Special theme: Human-Machine Interaction in Translation. Copenhagen Studies in Language* 41, 107–18, Copenhagen: Samfundslitteratur.

Teixeira, C. (2014), 'Perceived vs. Measured Performance in the Post-editing of Suggestions from Machine Translation and Translation Memories', in S. O'Brien, M. Simard and L. Specia (eds), *Proceedings of the 3rd* W*orkshop on Post-editing Technology and Practice*, 45–59, Vancouver: AMTA.

Temnikova, I. P. (2010), 'Cognitive Evaluation Approach for a Controlled Language Post-editing Experiment', in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10),* 3485–90, Malta: European Language Resources Association (ELRA).

Temnikova, I. P., W. Zaghouani, S. Vogel and N. Habash (2016), 'Applying the Cognitive Machine Translation Evaluation Approach to Arabic', in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 3644–51, Portorož: European Language Resources Association (ELRA).

Toral, A., M. Wieling and A. Way (2018), 'Post-editing Effort of a Novel with Statistical and Neural Machine Translation', *Frontiers in Digital Humanities*, 5, 9.

Turchi, M., M. Negri, M. Federico and F. F. B. Kessler (2015), 'MT Quality Estimation for Computer-assisted Translation: Does it Really Help?', in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 530–5, Beijing: ACL.

Underwood, N. L., B. Mesa-Lao, M. García-Martínez, M. Carl, V. Alabau, J. González-Rubio and F. Casacuberta (2014), 'Evaluating the Effects of Interactivity in a Post-editing Workbench', *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, 553–9, Reykjavik: European Language Resources Association (ELRA).

Vasconcellos, M. (1986), ‘Post-editing on Screen: Machine Translation from Spanish into English’, in C. Picken (ed.), *Translating and the Computer 8*: *A Profession on the Move*, 133–46, London: Aslib.

Vasconcellos, M. (1989), ‘Cohesion and Coherence in the Presentation of Machine Translation Products’, in J. E. Alatis (ed.), *Georgetown University Round Table on Languages and Linguistics*, 90–105, Washington D.C.: Georgetown University Press.

Vasconcellos, M. (1992), ‘What Do We Want from MT?’, *Machine Translation*, 7(4), 293–301, Netherlands: Springer.

Vasconcellos, M. (1993), ‘Machine Translation. Translating the Languages of the World on a Desktop Computer Comes of Age’, *BYTE*, 153–64, McGraw Hill.

Vasconcellos, M. and M. León (1985), ‘SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization’, *Computational Linguistics*, Vol. 11 (2–3), 122–36, Cambridge (MA): MIT Press.

Vieira, L. N. (2014), ‘Indices of Cognitive Effort in Machine Translation Post-editing’, *Machine Translation*, *28* (3–4), 187–216, Netherlands: Springer.

Vieira, L. N. (2016), ‘Cognitive Effort in Post-editing of Machine Translation: Evidence from Eye Movements, Subjective Ratings, and Think-aloud Protocols’, Doctoral thesis, Newcastle University, Newcastle.

Vilar, D., J. Xu, L. F. d’Haro and H. Ney (2006), ‘Error Analysis of Statistical Machine Translation Output’, *Proceedings of Fifth Edition European Language Resources Association*, 697–702, Genoa: European Language Resources Association (ELRA).

Wagner, E. (1985), ‘Rapid Post-editing of Systrans’, in V. Lawson (ed.), *Translating and the Computer 5: Tools for the Trade*, 199–213, London: Aslib.

Wagner, E. (1987), ‘Post-editing: Practical Considerations’, in C. Picken (ed.), *ITI Conference I: The Business of Translating and Interpreting*, 71–8, London: Aslib.

Way, A. (2013), ‘Traditional and Emerging Use-cases for Machine Translation’, in *Proceedings of Translating and the Computer 35*, 12, London: Aslib.

Zhechev, V. (2012), ‘Machine Translation Infrastructure and Post-editing Performance at Autodesk’, in S. O’Brien, M. Simard and L. Specia (eds), *Proceedings AMTA 2012 Workshop on Post-editing Technology and Practice (WPTP 2012)*, 87–96, San Diego: AMTA.