
Machine Translation as an Academic Writing Aid for Medical Practitioners

Carla Parra Escartín

carla.parra@adaptcentre.ie

Sharon O'Brien

sharon.obrien@dcu.ie

ADAPT Centre, SALIS, Dublin City University, Dublin, Ireland

Marie-Josée Goulet

marie-josee.goulet@uqo.ca

University of Quebec in Outaouais, Gatineau, Canada

Michel Simard

Michel.Simard@cnrc-nrc.gc.ca

National Research Council of Canada

Abstract

In this paper we explore the utility of Machine Translation as a writing aid and its impact on the quality of the text produced. We focus on medical practitioners who are native speakers of Spanish and who need to publish their scientific work in English as a foreign language. After carrying out a general survey to determine whether Spanish-speaking medical practitioners already use MT as a writing aid, we engaged five participants in an experiment where we asked them to write a paper in Spanish that was subsequently machine translated. They were then asked to post-edit the MT output. We analyse their post-edits and further attempt to evaluate the overall quality of their texts by engaging a professional proofreader. Our results suggest that the texts produced with the help of MT+post-editing still require many edits in order to be considered of acceptable quality. In the conclusion, we identify several avenues worthy of future investigation and that could help achieve better quality.

1. Introduction

In recent times two developments have led to a new type of Machine Translation (MT) deployment, i.e. MT for personal use. Those two developments are: (1) freely available online MT systems and (2) increasing quality of MT output, for some language pairs at least. The 'average' internet user can now take advantage of MT to assist with various tasks such as school homework, translating website content for service and product reviews, and so on. Embedding of MT widgets in all sorts of websites has also contributed to personal MT usage.

One user type that might avail of MT for personal, and professional, purposes is the academic whose first language is not English, but who, in order to widely disseminate his or her work, wishes to publish in English. It is our belief that some who write in English as a Foreign Language (henceforth: EFL writers) are using freely available online MT systems as an aid to the writing process, first writing passages of text in their L1 (or first language) and translating those into English as they produce academic articles.

Despite increasing quality from MT engines, it is still accepted that MT output generally requires post-editing before it is of publishable quality. The focus of our research is on the use of MT as an aid by EFL writers in specialised fields. As this topic appears to have not been researched in any detail, as outlined below, we aim to explore the utility of MT as a writing aid and its impact on the quality of the text produced.

English is the undisputed *lingua franca* of academia (Bennett, 2013, 2014, 2015). This forces those who are not native speakers of English to publish in English in order to disseminate their research and progress their careers. As we have reported elsewhere (O'Brien et al., forthcoming; Goulet et al., forthcoming), this leads to a considerable disadvantage, especially for those who do not master English as a foreign language. The disadvantage touches on the cognitive level (Breuer, 2015), as well as on the career level, if journal acceptance is taken into consideration (Benfield and Feak, 2006), and on the economic level, if cost of additional translators or proofreaders is factored in (Lillis and Curry, 2010). Using MT as a writing aid might ease some of these disadvantages by (1) allowing people first to write in their L1 and then use MT as an aid to produce text in English, thereby tackling some of the cognitive demands of writing in a foreign language and (2) reducing costs by eliminating the need for translators or proofreaders, who often do not possess the specialised domain vocabulary in any case (Willey and Tanimoto, 2015).

Of course, there are several assumptions here that need to be examined. For example, does writing in L1, Machine Translating, and post-editing by the author (which we term 'self-post-editing') reduce the cognitive burden on the EFL writer? Can non-translators (authors in our context) post-edit their own work to an adequate level of quality? Does this method lead to higher quality in the English text, such that journal acceptance is a smoother process? Does it eliminate the need for a proofreader? We cannot tackle all of these questions here, but we have begun to address the questions regarding the quality of the English text (see below), the need for a proofreader, and the feasibility of self-post-editing.

2. Related Research

We report more fully in O'Brien et al. (forthcoming) and Goulet et al. (forthcoming) on related research and so will just summarise here. To put it succinctly, there is little work that focuses on this topic. Some work has been done on MT and second-language writing (for example, Niño, 2008, Garcia and Pena, 2011; and O'Neill, 2012) that demonstrates that MT can be useful as a second-language writing support. This previous work focuses mostly on university students who were learning languages. To the best of our knowledge, no work has been done concerning MT as an aid for professional writing.

In O'Brien et al. (forthcoming), we made a first attempt to explore the utility of MT for the EFL academic cohort. This exploration found that the median times for drafting abstracts were not substantially different between L1 and EFL, however the revision times and number of revisions implemented were greater for the L1 (+MT) sections. The participants were split more or less down the middle in terms of their perceptions of ease of task, while six (out of nine) felt that the quality produced was equal for both and three thought that writing in EFL produced better quality. A professional proofreader was hired to evaluate the quality of the texts, and her assessment supported the authors' perception of quality. In short, we found that there was encouraging support for the assumption that MT could be used as a writing aid by EFL writers without taking up significantly more time and without impacting on quality.

In Goulet et al. (forthcoming), we analysed this data set in more detail, comparing the edits implemented by the proofreader across both halves of the abstract in order to ascertain whether the editing required for text produced in EFL was different from that written in L1 and subsequently machine translated and self-post-edited. In summary, we found the number of edits to be similar (5% and 6% of the total word count in EFL and MT respectively), but that for the authors with Arabic and Chinese as L1, the number of edits to the MT'd parts were higher than for languages such as French or Spanish. Overall, there were no very outstanding differences in terms of the proofreader's edits when one part of the abstract was compared to the other,

again indicating that MT as an academic writing aid certainly does not have a negative impact on the quality of the text produced.

3. Motivation

The exploratory study summarised in the previous Section 2 provided impetus for a follow-up study, which is the focus of this paper. Having previously recruited participants from a broad range of disciplines and languages, it was decided that it would be relevant to focus on one language pair and on one domain for a more in-depth analysis. Knowing anecdotally that medical practitioners seek to, and often have to, publish their research findings, we decided to focus on them as a new cohort. Moreover, we had anecdotal evidence that medical practitioners with Spanish as an L1 sometimes struggle to write in English. Add to this the fact that MT is known to perform relatively well between Spanish and English and so users might be encouraged by its output, we decided to recruit and analyse self-post-editing within this cohort. Our focus of attention this time was to understand more fully the nature of the self-post-editing task as well as MT usage among medical practitioners in general. We consequently asked the following questions:

1) *Are Spanish-speaking medical practitioners using MT as a personal writing support already?*

This question sought to explore whether or not our assumption about personal MT usage was correct.

2) *Without any training in MT or post-editing, what type of edits do medical practitioners make when they write in Spanish and then machine translate into English and self-post-edit?*

a) *Are essential edits implemented or ignored?* (See the Methodology discussion in Section 4.2.3 below for a definition of ‘Essential Edits’ and ‘Essential Edits not Implemented’)

b) *How much non-essential (or preferential) editing is carried out?*

c) *Are errors introduced via self-post-editing?*

Our goal here is to understand the natural competence for self-post-editing without any training whatsoever and to move towards developing potential supports for post-editing for such cohorts. By analysing essential and preferential edits as well as errors introduced we aim at establishing the degree of quality achieved in our experimental setup.

3) *How much editing is required by a professional proofreader on top of the post-edited documents and what type of edits are implemented?*

With this question we investigate whether L1+MT+self-post-editing actually requires another round of proofreading or whether the proofreader could be eliminated from this cycle. Again, this taps into a measurement of the quality produced during the self-post-editing setup.

4. Methodology and Experimental Setup

In order to address our initial research questions (cf. Section 2), we combined different research tools: questionnaires, active writing and post-editing, proofreading, and annotation of the edits made under each condition (self-post-editing and professional proofreading). In what follows we describe the methodology and experimental setup.¹

¹ The research reported here was granted ethical approval by the relevant Research Ethics Committees.

4.1. General Survey

To address the question: “*Are Spanish-speaking medical practitioners using MT as a personal writing support already?*”, we surveyed medical practitioners in Spain. The survey was run between 19 December 2016 and 27 January 2017 and the link to our questionnaire was sent to many organisations, including medical specialised associations, medical unions, universities and research institutes in Spain. We had a total of 50 responses. The questionnaire addressed several questions, including whether or not the respondents already used MT as a writing aid.

This general questionnaire also helped us to identify potential participants for the experiment. At the very end of our questionnaire, we asked the respondents whether they would be willing to participate in experiments using MT and collected their e-mail addresses. Although 31 of the respondents provided us with their e-mail addresses, only five were finally available for the first experimental cycle, carried out between February and the beginning of April 2017.

4.2. Experimental Setup

4.2.1. Participant Profiles

As stated earlier, only five of our questionnaire respondents (3 men and 2 women) were available to engage in the experiment reported here. Four of them are in an early stage of their careers, are between 20 and 30 years old, and are engaged in their residencies. The fifth one is a researcher at a university or research centre and is between 30 and 40 years old. One specialises in Neurosurgery, another in Internal Medicine, two of them are gynecologists and the fifth one works in Immunology. All of them have Spanish as their mother tongue and all of them speak other languages besides English (Catalan, French, German, Italian and/or Portuguese). Table 1 summarises their self-reported level of English using the Common European Framework of Reference for Languages (CEFR), as well as the level of English as established by an online English test on the Cambridge English website.² P01 rated his level of English as lower than what the placement test revealed, whereas P03 rated his level higher. The remaining participants had a fairly accurate self-assessment of their English level.

Participant	Self-Assessment (CEFR, writing)	English Level Test
P01	B2	C1-C2
P02	B1	B1
P03	C1	B2
P04	B2	B2-C1
P05	B1	B1-B2

Table 1: Participants' Level of English

Except for P01, all of the other participants had published a paper before, their number of publications ranged from 1 (P03) to 15 (P05), and only two of them (P02 and P05) had

² In order to cross-check their self-assessment with their actual English level, participants were asked to complete an English level test of 25 questions and let us know their final results. The test can be found here: <http://www.cambridgeenglish.org/test-your-english/general-english/>

published papers in English before. While P02 had only published one paper, P05 had published up to 5 papers in English. In both cases, their reported strategy for publishing was the same: they wrote directly in English and subsequently carried out a self-revision. P02 acknowledged having used Google Translate as a writing aid to confirm the translation of individual words or sentences.

4.2.2. Phase 1 : Publication Drafting in Spanish

We asked our participants to send us a publication or a section of a publication of approximately 750 words that they had originally written in Spanish. We additionally asked them to try, whenever possible, to avoid writing sentences longer than 20 words as this should help to achieve better quality from the MT system. As we could not expect them to count the words in each single sentence, we gave them a visual indication of 20 words in Spanish as being more or less equal to 1.5 lines in MS Word (Times New Roman, font size 12). We aimed at analysing the discussion section, or the section most similar to that, as it is more discursive than other sections (Skelton and Edwards, 2000). We deemed that this section may be one of the most challenging to write, especially for non-native speakers, and that therefore it is a good section to use in testing the use of MT as a writing aid.

4.2.3. Phase 2 : MT and Self-Post-Editing

Upon reception of the texts, we used Google Translate to translate them into English and sent them back to their respective authors asking them to correct the MT output. If they had sent the whole paper, we returned the whole paper translated, and asked them to review the specific section we had selected for our study. We asked them to carry out the revision using the “track changes” functionality in MS Word with the aim of being able to study the types of edits they had made.

After they had returned their self-post-edited texts, we asked them to fill in a post-task questionnaire about their experience. The results of this questionnaire are summarised, together with our analysis, in Section 5.2.

Upon reception of all files, we sought to answer our second research question: “*Without any training in MT or post-editing, what type of edits do medical practitioners make when they write in Spanish and then MT into English and self-post-edit?*”. To do so, we annotated all edits made by the medical practitioners. One author annotated the files and highlighted any unclear cases, and subsequently another author went through all the annotations and we carried out a negotiation phase to determine the final annotations in each case. Unclear cases were further discussed with a third author. As at this stage we were mainly interested in determining whether or not medical practitioners were implementing essential or preferential edits and whether new errors were introduced in the self-post-editing process, we chose to adopt the typology proposed by de Almeida (2013), who was interested in the nature of post-edits implemented by professional translators in an attempt to describe what a ‘good’ post-editor did. De Almeida reviewed many typologies for the analysis of post-editing activity and concluded that there was no internationally adopted model for classifying this type of task. She customised the LISA (2004) and GALE (NIST, 2007) typologies for her own purposes and then layered a number of ‘master categories’ over this typology. The master categories entail:

- **Essential edits:** if the edit is not implemented, the sentence (or part of it) is either:
 - a) Grammatically incorrect (i.e. it obviously breaches a grammatical rule specified in accepted grammar books), or

- b) Grammatically correct, but not accurate in comparison to the source text (i.e. it does not contain all the information that is present in the source text, i.e. an omission, or it contains extra information that is not present in the source text, i.e. an unnecessary addition).
- **Preferential edits:** an edit is considered preferential if the sentence from the raw MT output would still be grammatically correct, intelligible and accurate in relation to the source text, even if the edit in question was not implemented.
- **Essential edits not implemented:** This is an essential edit (as defined above) that was not implemented by the author.
- **Introduced errors:** The error was not present in the raw MT output, and it was introduced by the post-editor while editing a sentence. Because of this edit, the sentence (or part of it) is grammatically incorrect and/or inaccurate.

For this paper, we decided to slightly modify these master categories, and thus instead of categorising edits as ‘introduced errors’, we deemed it important to distinguish between ‘introduced errors’ that were attempting to correct something (i.e. an edit was essential, but the medical practitioner failed at fixing the problem), or those in which the edit was preferential and resulted in an error. That is: we treated the master category “introduced error” as a subcategory of either “essential edits” or “preferential edits”. A more detailed annotation of the types of edits is foreseen for the future.

4.2.4. Phase 3 : Professional Proofreading

As a last stage of our experiment, we recruited a professional translator and proofreader specialised in the medical domain to proofread the texts, after the medical practitioners had post-edited them. We confirmed all the changes made using the “track changes” functionality, and subsequently sent the proofreader the post-edited texts for revision (i.e. we did not provide her with the original Spanish text, nor did we explain the origin of those English texts). As we wanted to avoid over-editing, we also provided her with the following instructions:

“The texts are written in English and we are looking for a surface revision, that is, pay attention to grammar, orthography, punctuation, syntax, and major stylistic problems. We would like the texts to read well enough to be submitted to a scientific conference, for example.

The texts belong to the medical domain, and are all parts of scientific papers written by doctors.”

In order to be able to analyse the proofreader’s edits, we requested that the “track changes” functionality in MS Word be used. We then proceeded to annotate these edits using the same typology that we had used to annotate the edits made by the medical practitioners. Although it is true that the typology was meant to be used for the annotation of post-edited texts, we deemed that a classification of essential and preferential edits was also applicable to a proofread text. Thus, we removed the translation dimension from the typology and focused only on the correctness of the text, using the same categories. This strategy allowed us to reply to our last research question: “*How much editing is required by a professional proofreader on top of the post-edited documents and what type of edits are implemented?*”.

5. Data Analysis

5.1. General Questionnaire Response

As explained in Section 4.1, we conducted a general survey (in Spanish) seeking to gather information as to how Spanish medical practitioners currently write their publications.

The gender spread was 28 female, 21 male and 1 undeclared. 18 of the respondents were between 20 and 30, 8 between 30 and 40, 4 between 40 and 50, 14 between 50 and 60, and 6 were older than 60.

Most of the respondents were at the beginning of their careers and work in public hospitals. Although there were replies for most of the medical specialties, 20% of replies were from gynaecologists, 16% from cardiologists and another 16% from neurologists.

94% of the respondents indicated that their mother tongue is Spanish. For those who indicated a different mother tongue, all of them stated “Catalan”. 84% indicated that they speak English and the remaining 16% indicated that they did not. For self-assessment of English writing skills using the CEFR, only one indicated a C2 level, 4 a C1, 14 indicated a B2 and another 14 B1, 8 chose A2 and 1 A1.

74% of the respondents indicated that they have published scientific papers before and 26% indicated that they have never published.

76% (28 respondents) indicated that they had published papers in English and 24% (nine respondents) said they had no publications in English. The 28 respondents that indicated they had published in English were subsequently asked how those publications were drafted.³ Nine respondents (32%) indicated that they write directly in English and subsequently ask a colleague or friend with a better level of English to do the corrections. 29% (eight people) indicated that they directly write in English and self-revise, 25% (seven people) indicated that they write in Spanish and hire a professional translator, and another 25% (seven people) indicated that they write directly in English and subsequently hire a proofreader. Two people (7%) said that they hire a professional proofreader if they could do so, and another two acknowledged asking a colleague or friend who is a native speaker of English to do the proofreading. These figures support the claim that some EFL writers feel that they need to seek support from others in order to publish in English. This support is sought from colleagues and/or paid for through professional services.

Of the 28 respondents that had published in English, 19 (68%) indicated they had used Machine Translation for writing academic papers and nine (32%) said they had not. Those who said that they did not use MT (8 respondents, 89%) indicated as the main reason that they did not trust the quality.⁴ Two (22%) said that they did not know of the existence of MT, another two indicated that they have problems with terminology, and one indicated “other” and explained that for the type of texts they wrote they were confident enough in English and did not feel the need to use MT.

17 people (89%) indicated they use MT services to check how something is expressed in English, while five (26%) said they used it after drafting a document in Spanish to obtain a preliminary English version they could subsequently post-edit.

Though our questionnaire had a limited number of responses (50), it allowed us to confirm that some Spanish-speaking medical practitioners feel the need to rely on additional supports to aid them in producing articles in English, that some of them are using MT as a writing aid already and that they rarely use it to translate full documents, but rather short passages of text or individual words.

5.2. Post-Task Questionnaire Response

All five participants in our experiment were also asked to fill in a short post-task questionnaire aiming at gathering information about their experience.

³ This question allowed respondents to select all options that applied to them.

⁴ This question allowed respondents to select all options that applied to them.

We first asked them which method for writing scientific publications in English they deemed easiest after their experience participating in our experiment. 60% of them (three participants), chose the option they had just experienced, i.e. writing their publication in their mother tongue and subsequently post-editing an MT version of it. One participant indicated that s/he found it equally easy to write directly in English or to write publications using the proposed workflow, and the fifth participant indicated that s/he preferred to write his/her publications directly in English.

When asked to comment on the difficulties experienced when correcting the MT output, one participant said that s/he had encountered problems with synonyms, and another that s/he had found the translations to be too literal. The third participant said that the MT output was good, the fourth stated that the MT output was better than his/her own English level and therefore s/he found it difficult to identify errors, and the last one said that s/he had encountered the expected issues: grammar problems, terminology and words that change their meaning depending on their context and that had been translated wrongly.

Despite their complaints and comments about the MT output, three out of the five participants deemed that the overall quality of the MT output was at 3 on a scale from 1 to 4, and two gave it the maximum points.

When asked to rank how likely they were to use the proposed workflow for writing scientific papers in the future on a scale from 1 to 4, four of the five participants replied “3”, while the fifth replied “2”. Three of them further indicated that they thought a second experience like this one would allow them to achieve a better overall quality, whereas two indicated “maybe”. Four of them also stated that, with practice, this type of task would become easier, while one said “maybe”.

5.3. Word Count Statistics

Word Count Statistics						
Participant	P01	P02	P03	P04	P05	TOTAL
Number of words in ES	1413	759	1058	908	686	4824
Number of words MT (EN)	1340	685	959	865	639	4488
Number of words MT+Self-PE (EN)	1364	677	945	857	646	4489
Number of words MT+Self-PE+REV (EN)	1389	685	934	859	611	4478

Table 2: Word counts per experimental condition

As stated earlier, we engaged five medical practitioners in our experiments and asked them to draft a paper or a section of a paper of around 750 words, or to send us a paper they had already written in Spanish and intended to translate into English. Table 2 offers a general overview of the number of words they originally wrote in Spanish as well as the breakdown of the word counts after each stage in the experiment.

5.4. Types of Edits

We aimed at identifying the type of edits that medical practitioners make when they engage in the self-post-editing process without any prior training in MT or post-editing using the typology outlined in Section 4.2.3.

Figure 1 shows the edit rates per participant.⁵ Edits provoked by other edits were counted as a single edit in our analysis, as they would not have happened, if the first edit had not been made.

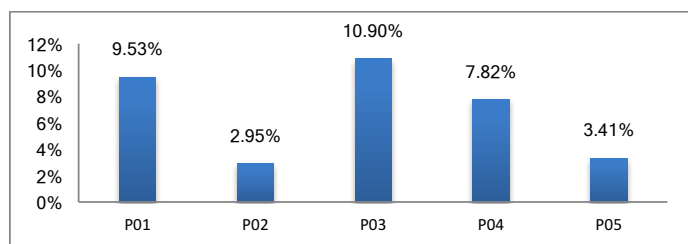


Figure 1: Edit rates per participant

As indicated in Figure 1, P03 was the author who has the highest edit rate, followed by P01 and P04. A potential explanation for this may be related to their English level. Both P02 and P05 had a B1 level of English according to the test (P05 was actually between B1 and B2), and also reported a B1 in their self-assessment. The other participants, on the other hand, had a B2 or C1 level (according to the test, P04 was between B2 and C1, and P01 between C1 and C2). It could therefore be the case, that a lower level of English hampers the ability to post-edit. This was also hinted at by P02 who declared that the MT output outperformed his/her level of English.

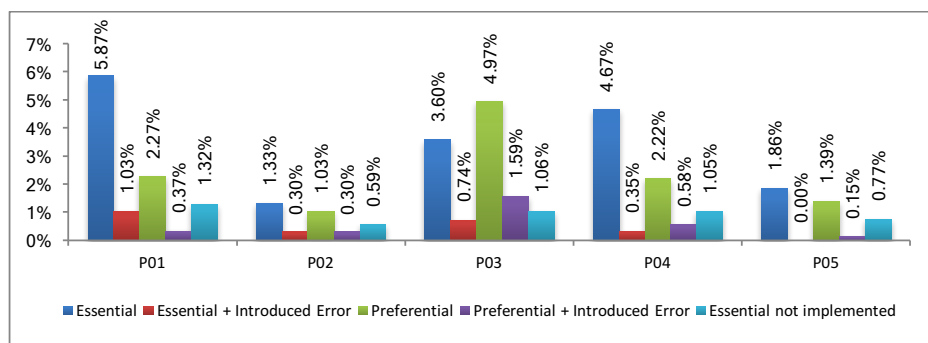


Figure 2: Types of edit per participant measured in edit rates

Our second research question was: “Without any training in MT or post-editing, what type of edits do medical practitioners make when they write in Spanish and then machine translate into English and self-post-edit?”. If we break down the types of edits made (cf. Figure 2), our analysis shows that medical practitioners are able to identify and implement essential edits during post-editing without any prior training. P01 and P04, the two participants with the highest edit rates as per Figure 1, are precisely the two participants who made the highest rate of essential edits (5.87% and 4.67% respectively). However, P01 was also the participant who had the highest rate of essential edits not implemented (1.32%), followed by P03 and P04 (1.06% and 1.05% respectively). This additionally replies to our related question, “Are essential edits implemented or ignored?”.

An interesting observation during the annotation was that in many cases the essential edits in the text were related to spelling and grammar, highlighting the importance of using

⁵ By “edit rate” here we mean the number of edits implemented per 100 words of raw MT output, expressed as a percentage

spelling and grammar checkers as writing aids for non-native speakers of English. It was also surprising to see how Google Translate performed well even with noisy text, as the original texts in Spanish contained some grammar and spelling errors. For example, *típica* (typical) was spelt *típica*; *nuestro* (our), *nuesto*, and *excluida* (excluded), *excluída*, and yet the MT system translated them correctly. In any case, it does seem to hold true that those participants with a higher level of English identified and implemented more essential edits than those with a lower English level.

With regard to the subquestion, “*How much non-essential (or preferential) editing is carried out?*”, we observed that again there is a tendency to implement preferential edits too. In our small cohort, only one participant (P03) implemented more preferential than essential edits. The extent of edits implemented varies however per individual, which has also been observed among professional translators who post-edit (e.g. de Almeida, 2013; Bundgaard, 2017). We found, for instance, that participants had different preferences regarding the use of technical versus colloquial terms, which was reflected in their edits. For example, P02 changed ‘axillae’ to ‘armpits’, whereas P05 seemed to prefer a more formal final text and changed expressions such as ‘hospital stay’ to ‘hospitalization time’.

Our last related question addressed *whether errors are introduced via self-post-editing*. As with professional translators, medical practitioners also introduced some errors while editing, though the rates are relatively low. P01 was the participant who introduced the highest rate of errors when making essential edits (1.03%), followed by P03 (0.74%). P03 was also the participant that introduced more errors when implementing preferential edits (1.59%), and as a result the one who had the highest rate of introduced errors overall (2.33%). Further investigation is needed to identify the nature of the errors introduced and determine whether they could have been avoided (e.g. by means of spell and grammar checkers in the case of introduced typos). However, this might also have to do with the *need* for edits: within the medical domain, there exist several sub-domains and genres. This raises a new research question worth investigating in our future work: *Did Google Translate perform better in some sub-domains than others?*

5.5. Professional Proofreading

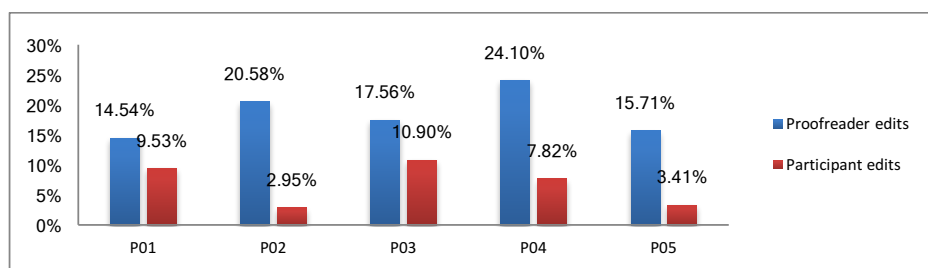


Figure 3: Edit rates per participant (proofreader vs. participants)

We subsequently analysed the edits made by the professional proofreader on the texts already self-post-edited by our participants. Figure 3 shows the overall edit rates per participant. The edit rates of each medical practitioner are provided to allow for an easier comparison. As may be observed, the professional revision of the texts resulted in a higher edit rate in all cases, with P04’s text being the one that recorded the highest edit rate, followed by P02’s. This is an interesting finding, as P04 was precisely one of the participants with a higher level of English, which suggests that the English level is not necessarily correlated with the post-editing ability. In some cases, e.g. P02 and P05, the proofreader introduced a significantly higher number of edits than

the participant, leading to edit rates that are more than five times those of the participants. An obvious question is whether or not the proofreader edits were indeed necessary or, were rather preferential and not strictly required. This is particularly relevant, because, as mentioned earlier, we asked the proofreader to focus on a mere surface revision of the text.

Similar to what we did with the texts undergoing self-PE we also annotated all edits per type of edit. Figure 4 summarises the edit rates of the proofreader classified by type. It is striking how in some cases the rate of preferential edits made was as high as that for essential edits (P04), or even higher (P01, P03 and P05). Only in the case of P01 was the rate of preferential edits lower than that of essential edits. Surprisingly, the proofreader also introduced some errors in the text while implementing edits. It is interesting to note that the number of introduced errors is higher in the case of preferential edits than in the case of essential ones. In some cases, the error introduced may have been caused by the use of “track changes” (e.g. when correcting the spelling of “patient”, she accidentally deleted the space between the word being corrected and the next: “the patientpatientwas urgently...”), but the degree to which this influenced the editing process is difficult to gauge. In the case of P03, some of the errors introduced had to do with the bibliographical style, as the medical practitioner had opted for references between parenthesis and our proofreader changed them to superscript.

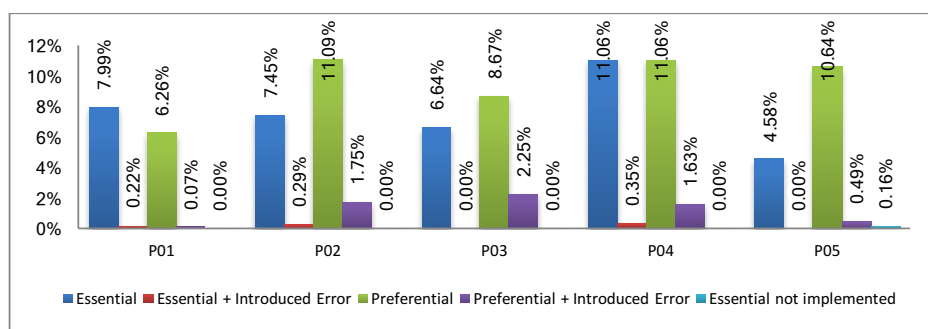


Figure 4: Types of edits by proofreader for each text

This analysis of the professional proofreader edits allows us to answer our third research question: “How much editing is required by a professional proofreader on top of the post-edited documents and what type of edits are implemented?”. Indeed, the proofreader implemented a considerable number of edits. However, according to our typology, the proofreader also implemented a surprising number of preferential edits and even introduced some errors during the proofing process, though these were low in number. This seems to indicate that the proofreader is still required and that the post-editing process by our small cohort of medical practitioners did not render the text to a level of quality such that the proofreader thought that it required little to no editing. As an aside, this question also arises in professional practice and the general practice is still to have a revision after post-editing, which indicates that our findings would not be out of line with normal machine translation workflows.

Although we are not doing a comparison here between the number of required edits after post-editing versus the number of required edits to texts directly written in EFL, in a previous experiment we observed that the proofreader implemented more or less an equal number of edits on text that had been post-edited and text that had been written in EFL (O’Brien et al., forthcoming; Goulet et al., forthcoming). In future work it would be interesting to test if the same findings can be replicated in the medical domain.

6. Conclusion and Future Work

Here, we have reported on a small experiment seeking to explore the usefulness of MT as a writing aid for Spanish medical practitioners that need to publish their work in English. Thanks to the general survey we conducted, we found that Spanish-speaking medical practitioners are already using MT as a writing aid. However, they also showed mixed feelings about its usefulness. Some of the main criticisms had to do with the literalness of the MT output, incorrect use of synonyms, grammar and the lack of terminology. This raises a question as to whether domain-tuned MT engines might solve some of these issues.

Our analysis revealed that medical practitioners perform both essential and preferential edits (3.90% and 2.52% respectively and on average for all participants), and that the professional proofreader hired to proof the self-post-edited texts written by our participants also implemented both types of changes (7.75% and 9.02%). Surprisingly, in the case of the proofreader the rate of preferential edits was higher than that of the essential ones. This seems to agree with what has been observed in professional translation workflows, as demonstrated by Bundgaard (2017). In an investigation of professional translators' edits during the "checking phase" of translations (translators checking their own work) Bundgaard (2017: 205) found the rate of preferential edits to be 43% on average for one text and 66% on average for a second text in her experiment, i.e. of all edits implemented for one text, 43% of them were deemed to be 'preferential'. Bundgaard was also using de Almeida's typology for assessing essential and preferential edits. Bundgaard (2017: 225) also measured the number of essential and preferential edits implemented by a third party during a 'review phase' (an independent translator checking another translator's work) and these ranged from 37% on average for one text and 60% for the second text.

Our analysis of the edits made by medical practitioners and the subsequent engagement of a professional proofreader additionally sought to answer whether medical practitioners would be in a position to carry out self-post-editing without any prior training and whether they were able to achieve an acceptable quality text with MT. Overall, without training, these experts can implement essential edits, but they also implement preferential edits and introduce errors. This raises the question as to whether further training and practice would make medical practitioners better post-editors.

At the same time, the proofreader's intervention demonstrated that an important number of essential edits had not been implemented by the medical practitioners. Yet, the proofreader also implemented a high proportion of preferential edits, according to our typology. It is still to be determined whether the texts produced by our participants would have actually been considered acceptable for publications or presentations in medical conferences where non-native speakers of English also present their work. In future work we plan to engage native speakers to assess this. We may consider, for example, asking them to rank the post-edited version against the proofread version to ascertain whether, and to what extent, the proofread version is acceptable as well as whether, and to what extent, it is superior to the post-edited version.

Similarly to what we did in Goulet et al. (forthcoming), we also plan to carry out a second round of annotation in which we will annotate the type of edit made (insert, delete, move, replace), the type of language unit affected in each case (noun, verb, preposition, etc.), and the linguistic dimension involved (morphology, syntax, semantics, etc.). This will allow us to analyse the edits further, make comparisons across the edits made by the experts and the professional proofreader, and determine whether automatic post-editing could be used to enhance the text prior to the self-post-editing process.

To sum up, our results, while demonstrating that the medical practitioners were capable of post-editing their own texts to some degree, do not seem to indicate that they could produce

their final drafts of scientific papers under the current experimental setup, i.e. with a generic engine, no automated post-editing rules and no intervention by a proofreader. However, the small cohort engaged in our experiment (five participants) does not allow us to draw a general conclusion. This experiment helped us to identify several avenues to improve our experimental setup and we will endeavour to address the issues identified and answer these new questions in our future work.

Acknowledgements

This research is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 713567, and Science Foundation Ireland in the ADAPT Centre (Grant 13/RC/2106) (www.adaptcentre.ie) at Dublin City University.

References

- Benfield J. R., and Feak, C. B. (2006). How authors can cope with the burden of English as an international language. *Chest*, 129(6), 1728-1730.
- Bennett, K. (2013). English as a lingua franca in academia. *The Interpreter and Translator Trainer* 7/2, 169-193.
- Bennett, K. (2014). The political and economic infrastructure of academic practice: the 'semi-periphery' as a category for social and linguistic analysis. In Bennett, K. (dir.), *The semi-periphery of academic writing: Discourses, communities and practices*. London: Palgrave Macmillan, 1-12.
- Bennett, K. (2015). Towards an epistemological monoculture: mechanisms of epistemicide in European research publication. In Plo, R. and C. Pérez-Llantada (dir.), *English as an academic and research language (English in Europe Vol. 2)*. De Gruyter Mouton: Berlin, 9-35.
- Breuer, E. O. (2015). *First language versus foreign language. Fluency, errors and revision processes in foreign language academic writing*. Frankfurt am Main: Peter Lang Edition.
- Bundgaard, K. (2017). *(Post-)Editing - A workplace study of translator-computer interaction at Textminded Danmark A/S*. PhD Thesis. Aarhus University, Denmark.
- De Almeida, Giselle (2013) *Translating the post-editor: an investigation of post-editing changes and correlations with professional experience across two Romance languages*, Unpublished PhD Thesis, Dublin City University.
- Garcia, I., and Pena, M. I. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning*, 24(5), 471-487. doi: 10.1080/09588221.2011.582687.
- Goulet, M.J., Simard, M., Parra Escartín, C. and O'Brien, S (forthcoming). *La traduction automatique comme outil d'aide à la rédaction scientifique en anglais langue seconde : résultats d'une étude exploratoire sur la qualité linguistique*. Anglais de Spécialité (ASp).
- Lillis, T. and Curry, M. J. (2010). *Academic writing in a global context: The politics and practices of publishing in English*. London and New York: Routledge.

- LISA (Localization Industry Standards Association). (2004). *LISA QA Model 3.0. Product Documentation*. Féchy, Switzerland: LISA.
- Niño, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, 21(1), 29-49.
- NIST (National Institute of Standards and Technology) Information Access Division/Speech Group and Linguistic Data Consortium (LDC). (2007). *Post Editing Guidelines for GALE Machine Translation Evaluation*. Version 3.0.2.
- O'Brien, S., Simard, M. and Goulet, M. J. (forthcoming, 2018). Machine Translation and Self-Post-Editing for Academic Writing Support: Quality Explorations. In: Moorkens, J., Castilho, S., Doherty, S. and Gaspari, F, Springer Series on Machine Translation. Springer.
- O'Neill, E. M. (2012). *The effect of online translators on L2 writing in French*, PhD thesis, University of Illinois at Urbana-Champaign, USA. <http://hdl.handle.net/2142/34317>. Accessed May 8 2017.
- Skelton, J. R. and Edwards, S. J. L. The Function of the Discussion Section in Academic Medical Writing. *BMJ: British Medical Journal* 320.7244 (2000): 1269–1270.
- Wiley, I., and Tanimoto, K. (2015). “We’re drifting into strange territory here”: What think-aloud protocols reveal about convenience editing. *Journal of Second Language Writing*, 27, 63-83, (2015). doi: 10.1016/j.jslw.2014.09.010.