

---

# MT Post-editing into the mother tongue or into a foreign language? Spanish-to-English MT translation output post-edited by translation trainees

**Pilar Sánchez-Gijón**

pilar.sanchez.gijon@uab.cat

**Olga Torres-Hostench**

olga.torres.hostench@uab.cat

Tradumàtica Research Group, Departament of Translation, Interpreting and Eastern Studies, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain

---

## Abstract

The aim of this study is to analyse whether translation trainees who are not native speakers of the target language are able to perform as well as those who are native speakers, and whether they achieve the expected quality in a “good enough” post-editing (PE) job. In particular the study focuses on the performance of two groups of students doing PE from Spanish into English: native English speakers and native Spanish speakers. A pilot study was set up to collect evidence to compare and contrast the two groups’ performances. Trainees from both groups had been given the same training in PE and were asked to post-edit 30 sentences translated from Spanish to English. The PE output was analyzed taking into account accuracy errors (mistranslations and omissions) as well as language errors (grammatical errors and syntax errors). The results show that some native Spanish speakers corrected just as many errors as the native English speakers. Furthermore, the Spanish-speaking trainees outperformed their English-speaking counterparts when identifying mistranslations and omissions. Moreover, the performances of the best English-speaking and Spanish-speaking trainees at identifying grammar and syntax errors were very similar.

## 1. Introduction

*V.(d) A translator should, as far as possible, translate into his own mother tongue or into a language of which he or she has a mastery equal to that of his or her mother tongue.*

UNESCO Recommendation, November 22<sup>nd</sup> 1976

Since UNESCO issued its recommendation, more and more translation companies and translation faculties have been adopting this “mother-tongue principle”, with excellent results. However, various authors have questioned this principle. Campbell (1998:212) argues that the “dynamics of immigration, international commerce and the postcolonial world make it inevitable that much translation is done into a second language, despite the prevailing wisdom that translators should only work into their mother tongue.” Kelly (2003) defends the same arguments of necessity, and Pokorn (2005: X) is perhaps the most critical. The latter argues that the traditional view “according to which translators should translate only into their mother tongue in order to create linguistically and culturally-acceptable translations (...) stems from an aprioristic conviction unsupported by any scientific proof that translation into a mother tongue is ipso facto superior to translation into a non-mother tongue.”

Is the same principle applicable to post-editing (PE)? Should PE also adhere blindly to this principle? Marcel Thelen, a supporter of non-native translators (2005:250), argues that the principle is too rigid and questions the UNESCO recommendation. The following extract from Thelen's book unintentionally became the starting point for the research presented in this paper.

“Applying the mother tongue principle seems to have become a sort of quality assurance, part of a guarantee of specialisation. Sticking to the native speaker rule is, however, not necessary in many cases, especially since clients do not all require the same quality of translations depending on the envisaged purpose. (...) In addition, with the implementation of technology and different kinds of translation tools, it becomes increasingly ‘easy’ for non-natives speakers to produce good English through post-editing.”

Is this true? Is it really so easy for non-native speakers to produce “good English”? In what sense would PE quality be affected if it were carried out by non-native speakers? This study attempts to discover whether non-native translation trainees could provide as good PE (in terms of accuracy and language) as native translation trainees. We conducted an empirical study in which a PE task from Spanish into English was carried out by two groups of subjects: non-native translation trainees and native translation trainees. The two groups were asked to post-edit several sentences from the user interface and help file of the OpenOffice software package. The aim of this study was to compare the results of the PE carried out by the two groups in terms of accuracy and language, and thus determine whether non-native translation trainees are able to meet the expected quality standards.

Traditionally, it was supposed there were two levels of PE —light PE and full PE— although TAUS prefers to talk about “good enough quality” and “quality similar to a human translator”. In our study, we expected non-native translation trainees to achieve “good enough quality” (TAUS 2010). The above list shows that in PE that is considered “good enough quality” expectations of the quality of language used are low, whereas accuracy is very important. Accuracy is non-negotiable both in light and full PE. So, if non-native speakers are able to provide accurate PE then we would need to bring into question the mother-tongue principle for PE.

## **2. Related work**

Native translation professionals seem to be the best suited for any PE job. Guerberof (2008, 2009, 2012) analysed the productivity and the quality of PE from the translation memories (TM) and machine translation (MT) output of native professional translators; Plitt and Mas-selot (2010) tested productivity by comparing MT+PE with traditional translation by native professional translators; Almeida and O'Brien (2010) compared PE performance with professional translation experience, and Temizoz's (2013) compared the differences in PE performance between engineers and professional translators. Other interesting studies of different post-editor profiles are Koehn's (2010) on PE by monolingual users and Mitchell, Roturier and O'Brien's (2013) who compare PE by monolingual users vs. that of bilingual users.

Many companies and organizations also rely on native speaking professional translators: the Commission of the European Communities worked with professionals on Systran PE (Wagner 1985); Sybase worked with professionals on PangeaMT PE (Bier and Herranz, 2011); and Continental Airlines worked with professionals on SDL PE (Beaton and Contreras, 2010).

Some organizations, however, are exploring other post-editor profiles. Computer Associates, for instance, is developing a PE crowdsourcing platform where any person who

knows two languages could become a post-editor and quality would be assessed by ranking the PE output (Muntés-Mulero and Paladini, 2012; Muntés-Mulero et al., 2012).

In an academic context, some researchers have carried out studies on PE using native translation students. Sutter and Depraetere (2012) analysed the relationship between PE, distance and fluency using translation trainees and O'Brien (2005) observed the correlation between PE effort and MT translatability. Especially relevant for our study is Garcia (2010) whose study on PE quality and the time taken for the task, used Chinese non-English native translation trainees, comparing their MT+PE with a translation made using a TM.

In light of the related literature, our contribution aims to explore a factor in the post-editors' profile that has been largely unexplored so far (except in the cases mentioned above): their mother tongue.

### **3. Method**

In this paper, we will check the following hypothesis: "PE jobs performed by native translation trainees will be more accurate and linguistically correct than those performed by non-native translation trainees". In order to investigate whether this hypothesis is valid, we will try to answer the following research questions:

- To what extent is PE performed by non-native translation trainees accurate?
- To what extent is PE performed by non-native translation trainees linguistically correct?

As stated in the introduction, our main focus was to establish what level of PE accuracy and linguistic correctness non-native translation trainees can produce taking into account their presumed poorer use of the foreign language compared with that of native speaking translation trainees. Accuracy was analysed by evaluating post-edits of mistranslations and omissions; language was analysed by evaluating post-edits of grammatical and syntax errors. The results of this study may be useful when taking decisions on PE training programs.

#### **3.1. Preparation of the corpus**

The sentences to be post-edited were taken from the English-Spanish bitext of OpenOffice (Tiedemann 2009). We downloaded the TMX file for the en\_GB and es languages (50.6k). The characteristics of this corpus made it a good choice for our study:

- All post-editors had computer skills, so the subject matter of the text did not pose a major challenge to them.
- All the sentences were easy to understand even though post-editors were not given the context.

We began by collecting sentences in Spanish from the English-Spanish bitext. We then back-translated the Spanish sentences into English using Google Translate. Finally, we compared the machine-translated sentences against the original English sentences and selected machine-translated sentences containing specific types of translation errors, as explained in section 3.2.

Twenty correct sentences, chosen from the corpus, were inserted into a table in two columns, Spanish on the left, English on the right, so that students could familiarise themselves with the genre, grammar and syntax of the text. Students read these sentences as a warm up task. We made sure that the sentences chosen were well translated and easy to understand without additional context.

A second table contained the sentences to be post-edited. The first column contained the original segment in Spanish, while the second and third columns showed the Spanish sentences translated into English using MT. Post-editors were asked to enter their changes in the third column of the document. The table below shows the first sentence as it appeared to the post-editors.

| ES (do not change)  | EN (do not change)   | EN (for post-editing)  |
|---|--|--|
| Crea un vínculo al arrastrar y colocar un objeto del Navegador en un documento. | Create a link to drag and drop an object from the Navigator into a document. | Create a link to drag and drop an object from the Navigator into a document. |

Table 1. Extract from the PE test.

The sentences in this second table were chosen in line with the objectives of our research. In the example above, for instance, there is a mistranslation (“to drag and drop” is not the same as “al arrastrar”) and a grammar mistake (“create” instead of “It creates”).

Each sentence was independent from the others, but they were clear and comprehensible even without any context. All the English sentences given to the participants were raw back-translations from Spanish into English produced by Google Translate. These back-translations contained the errors listed in the error typology presented above: 10 mistranslations, 5 omissions, 5 grammatical errors, and 5 syntax errors.

### 3.2. Error typology

We defined our error typology based on the error types listed in the “List of MQM Issue Types”, the QT LaunchPad project (2013), and the TAUS Error typology guidelines (as drafted by Sharon O’Brien for TAUS Labs and reviewed and endorsed by a large number of companies and organizations in 2013). The TAUS typology has four main categories: accuracy, language, terminology and style. As defined in the TAUS Error typology guidelines, “the category of ‘Accuracy’ is applied when incorrect meaning has been transferred or there has been an unacceptable omission or addition in the translated text.” In our test, we divided the TAUS Accuracy category into two subcategories, mistranslations (when an incorrect meaning has been transferred) and omissions. We did this in order to observe which mistakes made by post-editors were due to misunderstanding the message (accuracy) or which were due to lack of sufficient attention (omissions). Many words in our MT output were misplaced, resulting in mistranslations and changes of meaning. When preparing the test, we also found omissions that affected the meaning of the target sentence. Additions were not included in the study as there were no additions in the MT output analysed. For the study we selected ten sentences containing a mistranslation and five containing an omission.

As for Language errors, we identified two subcategories: grammar and syntax. Although the two are closely related, we wanted to analyse them separately since grammatical errors may reduce comprehensibility more than syntax errors. For the study we selected 5 sentences with a grammatical error and 5 with a mistake in the syntax.

We disregarded stylistic mistakes because this is a rather subjective category and is largely irrelevant when dealing with “good enough” PE. We also disregarded terminology mistakes because they are applied only “when a glossary or other standard terminology source has not been adhered to” (TAUS, 2013) and we did not provide a separate glossary for students. Moreover, we selected more mistranslation errors (10) than errors in the other categories (5) because, in our view, this category is key in translation quality reports. Mistranslation

errors are heavily penalised and translation trainees must be aware of the dangers of mistranslation in post-editing.

### **3.3. Participants**

Fifteen translation trainees participated in the research: 12 Spanish non-native English speakers (group A) and 3 native English-speakers from the USA (group B). The group of non-native speakers is larger than the group of native speakers because we were especially interested in the results of this group and its variability. As a matter of fact, the results for group A were initially compared with the correct solution, rather than with the results of group B. We are aware that, statistically, the analysis of the performance of these two small groups of participants will not yield solid results, but they should show if this issue is worthy of further research.

#### **3.3.1. Group A**

Of the 12 participants in group A, 11 were female and 1 male, (9 were aged 21 or 22, with the remaining 3 aged between 27 and 38). All were students in their seventh semester of a translation and interpreting degree in Spain. English was the first foreign language of 10 of the students, the second foreign language of 1 and the third foreign language of 1. Nine of the students had taken a Spanish-English translation course in Spain and 2 had done so in England via an Erasmus grant.

All members of the group attended an introductory 2-hour seminar on machine translation PE and worked on a 10-hour group project to compare and post-edit output from different MT systems into their mother tongue. Their attitude towards PE was generally positive. Only 1 participant said she was “not at all interested” in PE; 3 said they were “a little interested” and 7 said they were “quite interested”. The group was fairly used to using technology. Most of them were active computer users (for professional and personal purposes) but were not computer experts (only one said she had ever developed and created new computing solutions).

#### **3.3.2. Group B**

Group B was composed of 3 native English-speaking American translation trainees aged 30-40, all of whom were female. All three were very successful students taking a distance master’s programme in professional translation with grades of 95% or higher. All had taken Spanish-English and English-Spanish translation courses and had some specific training in localization. Two of them had completed professional internships. All members of the group received the same introductory seminar on machine translation and PE as Group A, part of an online course in translation technologies. Afterwards, they worked on a 5-hour individual project to post-edit a text produced by MT. Two participants said they were “very interested” or “extremely interested” in PE, while 1 participant said she was only “a little interested”. Two of them used IT resources both for professional and personal purposes, while 1 stated that she did not like IT resources but had learned to use them professionally.

### **3.4. Post-editing test procedure**

Both groups participated from the same location where they had received their translation classes, and training in MT and PE. Group A participants worked in a computer lab and group B participants worked online. Members of the two groups were given only one Microsoft Word document with all the information they needed for the test.

Firstly, students read a general explanation of the task to be carried out. They were asked to focus particularly on mistranslations, omissions, and grammar and syntax errors that compromised translation accuracy and comprehensibility. Once they had read the instructions they were invited to answer a profile questionnaire. They were then instructed to read the first table with the 20 correct bilingual sentences, and were told to pay close attention so they could become accustomed to the genre. After this warm up task, they noted the time and then started the PE. The sentences appeared in random order. After PE, they noted the time again and signed a research authorization form granting us permission to use the test data anonymously and only for research purposes.

The post-edited segments were analyzed anonymously. When processing the post-edited segments we considered only those translation errors that we used as indicators. Correctly edited segments were counted as successful edits; segments in which the error had not been detected were counted as unsuccessful edits; and sentences in which changes introduced by post-editors did not make it clear whether the error had been detected were ignored. The post-edited segments were compared with the correct and published translation of the sentences of the corpus.

## 4. Results

### 4.1. Results for non-native Spanish-to-English participants (group A)

Table 2 below shows the overall success rates for group A. The results show that non-native participants were most successful at detecting mistranslations, followed by omissions, syntax errors and finally grammatical errors.

| Error category  | Success rate | Error category | Success rate |
|-----------------|--------------|----------------|--------------|
| Mistranslations | 72.5%        | Accuracy       | 72%          |
| Omissions       | 71.67%       |                |              |
| Syntax          | 66.67%       | Language       | 59%          |
| Grammar         | 51.67%       |                |              |

Table 2. Success rates for group A

It is worth mentioning that non-native participants performed much better correcting mistranslations than language errors. In other words, their command of written English is not as good as their attention to details with respect the accuracy of sentences. We will now analyse each category separately.

a) *Mistranslations*. The test contained ten previously identified mistranslations. Variability in this category is quite high. The least corrected mistranslation was corrected by 5 of the 12 participants, while one mistranslation – the final one – was corrected by all 12 participants, perhaps because by that point the participants were more familiar with the task.

b) *Omissions*. The test contained five omissions that we identified in the MT output. The least-corrected omission was corrected by only 5 of the 12 post-editors, probably because MT produced a perfectly coherent, grammatically correct, factually accurate sentence.

c) *Syntax*. The test contained five previously identified syntax errors. In this category, the least corrected error was corrected by 7 of the 12 participants. The inability to correct this syntax error might be due to a lack of knowledge of the finer rules of English grammar by the participants. The most corrected mistake was corrected by 11 of the 12 participants.

d) *Grammar*. The test contained five previously identified grammatical errors. Uncorrected grammatical errors revealed that non-native participants had some problems with English grammar. The least corrected sentence was corrected by only 5 of the 12 participants, while the most corrected sentence was corrected by 9 of them.

Results for group A have been compared in line with the time devoted to the task by each participant. Figure 2 shows the number of post-editings by each post-editor alongside the time they devoted to the test (upper line). Interestingly, in some cases, there seems to be a correlation between the time devoted to the task and the results. For instance, the best performer of the group spent more time than the rest (except for one other person who spent even more time on the task) while the worst performer was the fastest in the group.

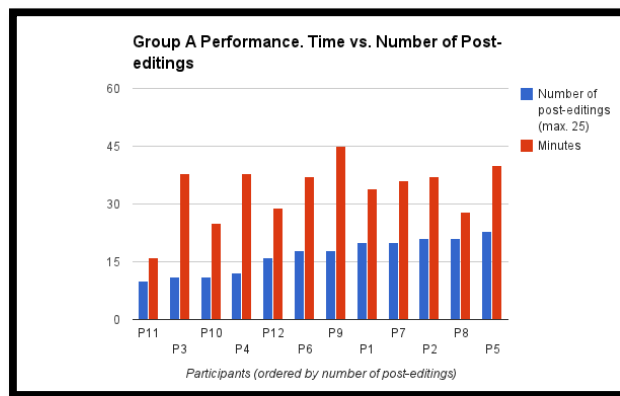


Figure 1. Number of post-editings vs. time spent - Group A.

In Figure 1, we can also see the variability of results among participants. Variability ranges from only 8 correct post-edits by the worst performer to 24 by the best performer. The mean number of correct edits per person was 16.75 out of 25.

Finally, in these results it is interesting to note the results of participants who had a slightly different background to the rest. For instance, all but two participants had English as their first foreign language. For participant 3, English was her third foreign language and she scored 13 out of 30. However, three other participants scored less despite English being their first foreign language. English was participant 9's second foreign language, and she scored 20 out of 30, outscoring five participants whose first foreign language was English. We can see, then, that participants whose first foreign language was not English were not the worst in the group. However, their results suggest they would not be trustworthy post-editors for translations into English.

If we only look at the overall results we might conclude that working with non-native speakers would not be advisable. However, good non-native post-editors could be suitable for the job. In this PE test, if only the participants having a 70%+ success rate were selected then the overall results would change dramatically, because in average their performance would be comparable to group B.

#### 4.2. Results for Spanish-to-English post-editing carried out by native speakers (group B)

Table 3 below shows that native participants corrected almost all the syntax errors. They also performed very well at correcting mistranslations, followed by omissions and grammar. It is

worth mentioning that their results for Accuracy and Language almost match, indicating a more equal level of skills than non-native participants.

| Error category  | Success Rate | Error category | Success Rate |
|-----------------|--------------|----------------|--------------|
| Mistranslations | 80%          | Accuracy       | 76%          |
| Omissions       | 66.67%       |                |              |
| Syntax          | 93.33%       | Language       | 77%          |
| Grammar         | 60%          |                |              |

Table 3. Success rates for participants in group B

From the individual questionnaires we know that participants who had completed a professional internship were slightly more successful. Let us look at the results of each error category in more detail:

a) *Mistranslations*. Overall performance by this group in this category was good; as was expected. None of the participants in this group corrected all the mistranslations, but they came very close. Two participants, both of whom had completed a professional internship, successfully corrected 90% of them.

b) *Omissions*. None of the participants corrected all the omissions, and one of the incorrect sentences was missed by all the participants.

c) *Syntax*. Two of the participants in group B corrected all 5 syntax errors, while the third missed just one of the errors. Syntax is closely related to style and English native speakers are expected to perform very well at correcting this kind of error. Our results suggest this is the key error category that makes the difference between native and non-native speakers.

d) *Grammar*. Grammatical errors were not very obvious. They were grammatical errors caused by a bad translation, but sometimes they resulted in an apparently correct sentence with a different meaning. As was the case with omissions, native speakers of English may have missed some of these errors because the target sentence did not seem to be ungrammatical. Only 1 participant detected all the grammatical errors, while 1 detected just one of the incorrect sentences.

Results for group B have been compared using the time devoted to the task by each participant. Figure 2 shows the number of post-editings by each post-editor alongside the time they devoted to the test (upper line). In this group, there is no obvious correlation between the time devoted to the task and the results of each participant.



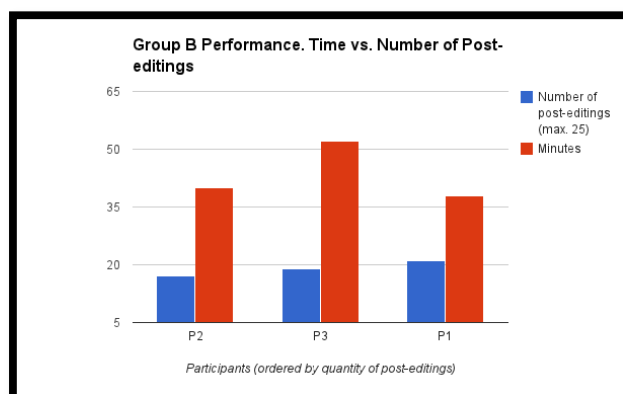


Figure 2. Number of post-editings vs. time spent: Group B

#### 4.3. Comparison of results

As Table 4 shows, group B (native speakers) performed better than group A (non-natives) in all error categories except Omissions. The biggest difference between group A and B is found in the Syntax category, due to the better command of the English language by native participants. However, Accuracy results are more similar in both groups. Upon a closer examination it is worth observing deviation results. Needless to say that the low number of participants, especially in group B, made it difficult to extract valid conclusions about deviation, other than the fact that there was huge variability among participants.

|                 | Group A                           | Group B                           |
|-----------------|-----------------------------------|-----------------------------------|
| Mistranslations | 72.5% ( $\sigma = \pm 21.1\%$ )   | 80% ( $\sigma = \pm 17.32\%$ )    |
| Omissions       | 71.67% ( $\sigma = \pm 24.8\%$ )  | 66.67% ( $\sigma = \pm 23.09\%$ ) |
| <b>ACCURACY</b> | 72.22% ( $\sigma = \pm 20.26\%$ ) | 75.56% ( $\sigma = \pm 19.24\%$ ) |
| Grammar         | 51.67% ( $\sigma = \pm 21.67\%$ ) | 60% ( $\sigma = \pm 40\%$ )       |
| Syntax          | 59.17% ( $\sigma = \pm 27.12\%$ ) | 93.33% ( $\sigma = \pm 11.55\%$ ) |
| <b>LANGUAGE</b> | 59,17% ( $\sigma = \pm 19.29\%$ ) | 76.67% ( $\sigma = \pm 15.28\%$ ) |
| Total           | 64.4% ( $\sigma = \pm 23.38\%$ )  | 75.37% ( $\sigma = \pm 22.11\%$ ) |

Table 4. Average success rates achieved by each group for each error category

General results seem to validate our hypothesis “PE jobs performed by native translation trainees will be better than those performed by non-native translation trainees”. Nevertheless, a more detailed analysis per category, looking at each group’s average scores for each category, as well as the highest and the lowest score in each group could shed light on the strengths and weaknesses of each group in order to find some answers to our research questions: “(1) To what extent is PE performed by non-native translation trainees accurate?; (2) To what extent is PE performed by non-native translation trainees linguistically correct?”

a) *Mistranslations*. Both groups achieved high scores when correcting mistranslations.

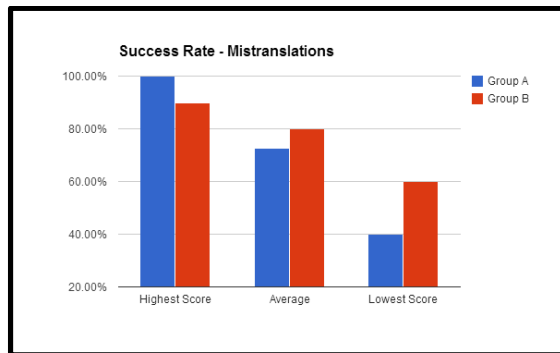


Figure 3. Success rates for each group in identifying mistranslations

Despite group B being the most successful at detecting mistranslations as a whole, it is worth noting that the best performer at correcting mistranslations belonged to group A (as well as the worst performer). Group B's results were more homogeneous. Group B participants were less prone to missing mistranslations, with the lowest scorer scoring considerably more than the lowest scorers in group A. In this case, being a native speaker or not does not seem to be essential in order to deliver a good post-editing job.

b) *Omissions*. As can be observed in Figure 4, group A was more successful at detecting omissions than group B. This might have been because many raw MT sentences were linguistically correct, but with a different meaning to the original source sentence. Since the target sentences seemed to be correct, native speakers of English may have been misled into believing they were valid translations. This is an important warning for native translation trainees doing PE.

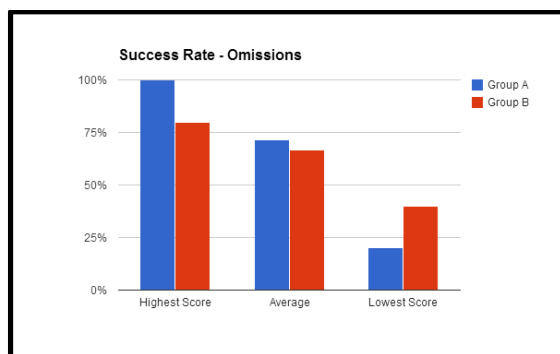


Figure 4. Success rates for each group in identifying omissions

c) *Grammar*. A priori, we may have expected all participants to have been successful at correcting grammatical errors. However, grammatical errors produced by MT systems range from the very obvious (for instance where a pronoun does not match the gender of a nearby subject) to the inconspicuous (for instance where the gender is given in a previous sentence). In other words, although the MT system might propose a sentence that is grammatically correct, the sentence might not be an appropriate equivalent of the source sentence in

terms of its grammar. Only a careful reading of the source sentence would detect this kind of problem.

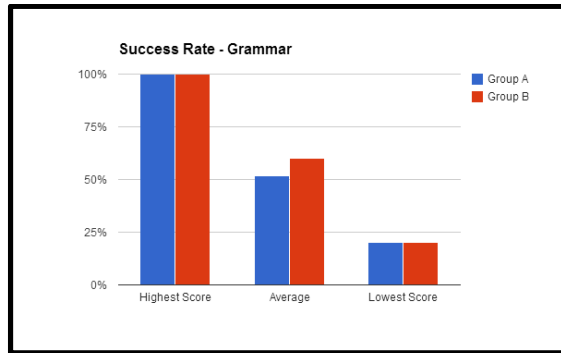


Figure 5. Success rates for each group in identifying grammatical errors

In this case, results from both groups were very similar. Detecting and correcting inconspicuous grammatical errors seems to be more related to the personal skills of each participant rather than the general skills of each group. In both groups there are participants who solved 100% of errors and only 20% of errors.

d) *Syntax*. Language command seems to make a difference in detecting syntax errors.

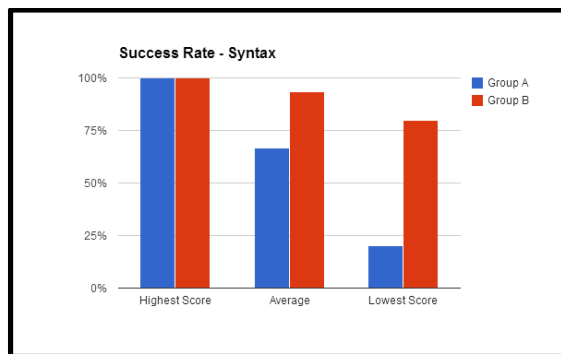


Figure 6. Success rates for each group in identifying syntax errors

As shown in Figure 6, group B performed better and more homogeneously than group A. Although the textual genre chosen for this study (text embedded in a software UI and text from the software's help module) has just a few very distinctive features (it is relatively poor stylistically), participants with less competence in the target language had more difficulty correcting errors related to word order. This suggests that native speakers of the target language would perform better than non-native speakers when post-editing texts belonging to a syntactically and stylistically more complex genre. If we look at the lowest scores from each group in this area we see that the lowest-scoring participants in group B considerably outperformed the lowest-scoring participant in group A.

So far we have presented the results for the best and the worst performer noting that there is considerable variability between them. Now it is worth commenting on variability in individual results. If we look at the results of individual participants we find that there were very good post-editors in both groups (see Figure 7). This information is very useful when

qualifying the validity of the hypothesis as from the detailed results in Figure 7, we would argue that it cannot be automatically inferred that PE jobs performed by native translation trainees are more accurate and linguistically correct than PE jobs performed by the best non-native translation trainees. Indeed, quality seems to depend on the person, regardless of their mother tongue.

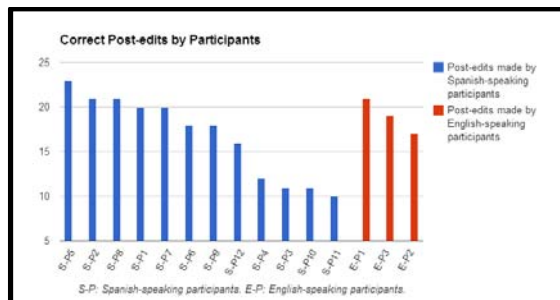


Figure 7. Correct post-edits by participants

When it comes to quality, the same principle is applicable for the research questions “(1) To what extent is PE performed by non-native translation trainees accurate? (2) To what extent is PE performed by non-native translation trainees linguistically correct?” When comparing the best Spanish-speaking participant with the best English-speaking participant (see Figure 8), it is revealing that one Spanish-speaker’s post-editing job is more accurate and even more linguistically correct than that of the best English-speaking participant.

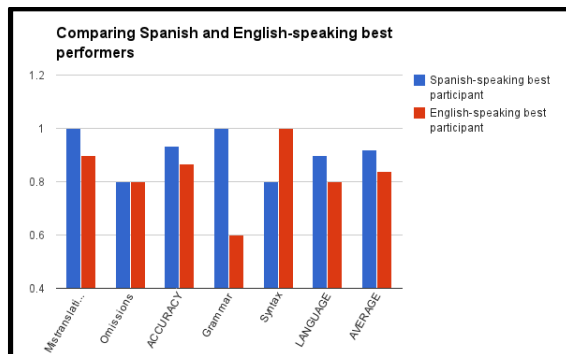


Figure 8. Correct post-edits by participants

## 5. Conclusion and future work

Our study suggests that good performers who are not native speakers of the target language can do “good enough” PE jobs from Spanish to English. The results for the best participants from group A (non-native) were very similar to those of group B (native) in a “good enough” PE task. These results question the validity of the “mother-tongue principle” for “good-enough” PE, as there are better and worse participants regardless of their mother tongue. The results also show that, while not all non-native participants were suitable for PE tasks, good non-native participants can indeed be suitable.

In light of the results, we need to ask whether non-native translation trainees require a different kind of PE training to that offered to native speaking translation trainees. These

results suggest that non-native translation trainees need more practice in identifying syntax and grammatical errors, while native translation trainees need to develop greater observational skills and pay more attention to detail in order to detect inconspicuous errors in natural sounding, grammatically correct sentences. From a cognitive point of view, it would be useful to prepare training material designed to develop observational and error detection skills for PE. These skills have proved to be more decisive in detecting accuracy errors in PE than the established mother-tongue principle. Besides providing an argument in favour of the creation of exercises related to PE guidelines, error typology, etc., these results suggest that PE training should also include exercises oriented towards the development of observational and error detection skills. In future work, it would be interesting to identify what skills good native and non-native post-editors have in common so that these may be improved through training.

Finally, it could be very useful for the translation industry to set up a standardized test to help identify which translators perform better as post-editors and whether they are best at post-editing into their mother tongue or into a foreign language. The test could be adapted for different PE projects or clients, with different types of errors. Such a test would be a suitable way of evaluating both native and non-native post-editors' skills, instead of merely disregarding non-natives a priori. In fact, such a test may even completely disregard if candidates post-edit into their mother tongue.

### **Acknowledgement**

This research is supported by the *Red de formación de poseedores* (Post-editor Training Network) (Ref. FFI2011-16021-E), funded by the Spanish State Secretariat of Research, Development and Innovation), and by the research project *Accesibilidad lingüística y sensorial: tecnologías para las voces superpuestas y la audiodescripción* (Linguistic and sensory accessibility: technologies for voiceover and audio description) (Ref. 2012-31024), funded by the Spanish State Secretariat of Research, Development and Innovation).

### **References**

- Beaton, A. and Contreras, G. (2010) Sharing the Continental Airlines and SDL post-editing experience. Proceedings, Ninth Conference of AMTA, Denver, Colorado.
- Bier, K. and Herranz, M. (2011) MT Experiences at Sybase. Presentation, Localization World Barcelona, July 2011. <http://www.slideshare.net/manuelherranz/loc-world2011-kbiermherranz-8730502>. Accessed 3 January 2014.
- Campbell, S. (1998) Translation into the second language. Addison Wesley Longman, New York.
- De Almeida, G. and O'Brien, S. (2010) Analysing post-editing performance: correlations with years of translation experience. Proceedings, 14th Annual Conference of the EAMT, S. Rafael, France.
- De Sutter, N., Depraetere, I. (2012) Post-edited translation quality, edit distance and fluency scores: report on a case study. Proceedings, Journée d'études "Traduction et qualité Méthodologies en matière d'assurance qualité", Université Lille 3, Sciences humaines et sociales, Lille.
- García, I. (2011) Translating by post-editing: is it the way forward? *Mach Transl*, 25:217–237.
- Guerberof, A. (2009) Productivity and quality in MT post-editing. Proceedings, MT Summit XII workshop: beyond translation memories: new tools for translators, Ottawa, Canada.

- 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.
- Guzmán, R. (2007) Manual MT Post-editing: if it's not broken, don't fix it. *Trans J*, 11, <http://www.bokorlang.com/journal/42mt.htm>. Accessed 24 January 2014.
- Kelly, D. (2003) *La direccionalidad en la traducción e interpretación: perspectivas teóricas, profesionales y didácticas*. Atrio, Granada.
- Koehn, P. (2010) Enabling Monolingual Translators: Post-Editing vs. Options. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, pages 537–545.
- Proceedings, AMTA Workshop on MT Research and the Translation Industry, pp 21-31.
- Enabling Monolingual Translators: Post-Editing vs. Options
- Roturier, M. and O'Brien, S. (2013) Community-based post-editing of machine-translated content: monolingual vs. bilingual. *Proceedings, MT Summit XIV Workshop on Post-editing Technology and Practice*, Nice, pp 35-43.
- Muntés, V., Paladini, P. (2012) Bringing the crowd in the post-editing process. Presentation, TAUS European Summit, Paris, <http://www.youtube.com/watch?v=1bgIKPULcsc>. Accessed 3 January 2014.
- Muntés-Mulero, V., Paladini, P., Solé, M. and Manzoor, J. (2012) Multiplying the Potential of Crowdsourcing with Machine Translation. *Proceedings, Tenth Biennial Conference AMTA*, San Diego.
- OPUS, the open parallel corpus website, <http://opus.lingfil.uu.se/>. Accessed 3 January 2014.
- O'Brien, S. (2005) Methodologies for measuring the correlations between post-editing effort and machine translatability. *Mach Transl*, 19:37–58
- O'Brien, S. (2009) Researching and teaching post-editing. Presentation, Seminar Post-Editing MT Output: Views from the researcher, trainer, publisher and practitioner, CNGL, <http://www.mt-archive.info/MTS-2009-OBrien-ppt.pdf>. Accessed 3 January 2014
- Plitt, M., Masselot, F. (2010) A productivity test of statistical machine translation post-editing in a typical localisation context. *Prague Bull Math Linguist*, 93:7–16.
- Pokorn, NK (2005) *Challenging the traditional axioms: translation into a non-mother tongue*, John Benjamins, Philadelphia, PA / Amsterdam.
- QT Launch Pad (2013) <http://www.qt21.eu/launchpad/content/list-mqm-issue-types>. Accessed 3 January 2014
- TAUS (2013) Quality evaluation using an error typology approach. <https://evaluation.taus.net/resources/error-typology-guidelines>. Accessed 3 January 2014

- TAUS (2010) Machine translation post-editing guidelines.  
<https://evaluation.taus.net/resources/guidelines/post-editing/machine-translation-post-editing-guidelines>. Accessed 3 January 2014
- Temizoz, O. (2013) Postediting machine translation output and its revision: Subject-Matter Experts versus Professional Translators. Doctoral thesis, Universitat Rovira i Virgili, Tarragona.
- Thelen, M. (2005) Translating into English as a non-native language: the Dutch connection. In: Anderman G, Rogers M (eds) In and out of English: for better, for worse, Multilingual Matters, Clevedon, pp 242-255.
- Tiedemann, J. (2009) News from OPUS – A collection of multilingual parallel corpora with tools and interfaces. In: Nicolov N, Mitkov R (eds.) Recent advances in natural language processing (vol V), John Benjamins, Amsterdam/Philadelphia, pp 237-248.
- Wagner, E. (1985) Rapid post-editing of Systran. In: Lawson V (ed) Tools for the trade: translating and the computer 5. Aslib, London, pp 199-213.