Translating the post-editor: an investigation of post-editing changes and correlations with professional experience across two Romance languages

Giselle de Almeida

Thesis submitted for the degree of Doctor of Philosophy

School of Applied Language and Intercultural Studies

Dublin City University

January 2013

Supervisors:
Dr Sharon O'Brien and Phil Ritchie

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ ID No.: _____

Date: _____

Abstract

With the growing use of machine translation, more and more companies are also using post-editing services to make the machine-translated output correct, precise and fully understandable. Post-editing, which is distinct from translation and revision, is still a new activity for many translators. The lack of training, clear and consistent guidelines and international standards may cause difficulties in the transition from translation to post-editing. Aiming to gain a better understanding of these difficulties, this study investigates the impact of translation experience on post-editing performance, as well as differences and similarities in post-editing behaviours and trends between two languages of the same family (French and Brazilian Portuguese). The research data were gathered by means of individual sessions in which participants remotely connected to a computer and post-edited machine-translated segments from the IT domain, while all their edits and onscreen activities were recorded via screen-recording and keylogging programs. A mixed-methods approach was employed for the qualitative and quantitative analysis of the data. The findings suggest that there are no clear correlations between translation experience and post-editing performance, or post-editing experience and post-editing performance. However, other aspects such as the opinion regarding machine translation seem to be predictors of post-editing performance. Our analysis enabled us to combine multiple factors in order to identify the 'best' post-editors in our participant group. Finally, similar post-editing trends were observed for both target languages, suggesting that training, guidelines and automated aids could be targeted at language groups rather than at individual languages. The insight gathered will be useful for devising future post-editing guidelines and training programmes.

Acknowledgements

Many people contributed in different ways to make this PhD possible, and I would like to acknowledge their great assistance.

I am deeply grateful to my academic supervisor, Dr Sharon O'Brien, and to my industrial supervisor, Phil Ritchie, whose outstanding support, guidance, patience, knowledge, encouragement, generosity and friendship were crucial for the completion of my research. It has been a privilege and an honour to have you as my supervisors!

I would also like to express my gratitude to VistaTEC and IRCSET, the funders of my PhD, for granting me a scholarship and giving me the opportunity to conduct this research, and for supporting and believing in the project.

Many thanks to the staff at VistaTEC, especially Malgorzata Jarlinska, Nuria Corominas, Treasa Kelly and Ronan Daly, for all the logistical and technical assistance.

I am very grateful to Autodesk for providing the corpus and the post-editing environment for my main experiment, and I would like to thank Mirko Plitt and Francois Masselot for their invaluable support and technical expertise.

Thanks to my two examiners, Dr Dorothy Kenny and Dr Lucia Specia, who made very valuable and helpful recommendations to improve my research.

Thanks also to all the post-editors and the two validators who took part in my experiments, and whose participation was an important element of my research.

To all my colleagues and the staff from SALIS in DCU, thank you for your help and encouragement! Thanks also to all my friends for being a constant source of motivation.

I would like to thank my former teachers at PUC-Rio, Dr Marcia Martins, Dr Maria Paula Frota, Paulo Henriques Britto and André Beckenn, for their

invaluable guidance and support not only during my undergraduate studies, but also on other occasions through the years, and for providing me with the opportunity to give the very first presentation related to my PhD, in 2008.

A very special thank you to my dear friend Maria Elizabeth Cabral de Melo, for helping me gather all the necessary documentation to apply for the PhD, and for her constant support and friendship.

To Lia Wyler, who has always been a great example and inspiration for me since my early days in the translation profession: thank you for your friendship and encouragement, and for all that I have learned from you through the years.

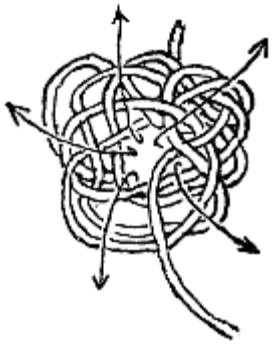
Minha gratidão a minha mãe, Creuza, e a meu avô, Plácido, que sempre me incentivaram e apoiaram: gostaria que vocês tivessem podido compartilhar comigo a conclusão deste desafio.

À ma famille française : merci de votre tendresse et de votre soutien constant.

A todas as pessoas queridas que tenho no Brasil - Cremilda, Lúcia, Lívia, Paquito, Cirênio, Célia, Sebastião (*in memoriam*), Kelly, Betinho, José Manoel, Antônia, Raimundo, Alba, Suleima, Maria, Manoela: sem vocês, eu não teria conseguido chegar até aqui. Muito obrigada, de coração.

Et à Nicolas, qui m'aide toujours à retrouver la capacité de sourire et l'envie et le courage de poursuivre mes rêves : merci de tout.

To the memory of my mother.



Post-editing – untangling threads of meaning:

"To untangle a snarl, loosen all jams or knots and open a hole through the mass at the point where the longest end leaves the snarl. Then proceed to roll or wind the end out through the center exactly as a stocking is rolled. Keep the snarl open and loose at all times and do not pull on the end; permit it to unfold itself. As the process is continued the end gradually emerges. No snarl is too complicated to be solved by this method; only patience is required."

(Clifford W. Ashley – The Ashley book of knots, p. 29)

Table of contents

List of tables

List of figures

List of abbreviations

ALPAC - Automatic Language Processing Advisory Committee

AVI - Audio Video Interleave

BLEU - Bilingual Evaluation Understudy

CAT - Computer-Aided Translation

GTM - General Text Matcher

IT - Information Technology

LSP - Language Service Provider

MT - Machine Translation

NTIs - Negative Translatability Indicators

PAHO - Pan American Health Organisation

PE - Post-Editing

QA - Quality Assurance

SMT - Statistical Machine Translation

TAPs - Think-Aloud Protocols

TAUS - Translation Automaton User Society

TER - Translation Edit Rate

TM - Translation Memory

UI - User Interface

Chapter 1: Introduction

In today's globalised world, quick and accurate language translation is more important than ever before, being essential to business success in many fields. According to a report by Common Sense Advisory, "the language services market is growing in 2012 at an annual rate of 12.17%" (Kelly and DePalma 2012: 6). To meet the increasing demand for the translation of large quantities of information, Machine Translation (MT) is being adopted by companies worldwide, especially in the Information Technology sphere. While machine translation has started to play an important role in the modern translation process, human linguists are still required to verify and correct machine-translated documents. This is done through a process called post-editing (PE), in which the raw output of MT is corrected in order to make documents as accurate and readable as possible. Understanding and clarifying exactly what is involved in post-editing is no simple task, however.

Post-editing is distinct from translation, employing a specific skill set. It involves specific quality requirements and productivity expectations, and the guidelines and constraints adopted can vary significantly from company to company. Post-editing differs from translation because, among other particularities, it requires:

- Dealing with three texts (source, MT output, and target) instead of two
- Changing sentence structure, editing grammatical inflections, and
- Un-translating words (like product names, for example), if needed.

The final quality required must also be taken into consideration to make sure neither too many nor too few resources are devoted to the post-editing process, because one of the main objectives of MT is to reduce translation costs and increase translation productivity.

Still, there is little training available related to post-editing, and there are no international standards designed to provide clear and universal definitions for the different levels of post-editing. Therefore, translators often encounter difficulties when they are required to perform this task.

1.1. Aims of this research

Post-editing has been researched for many years, and research interest has grown in the last decade, as will be illustrated in Chapter 2. Still, we do not fully understand the nuances and complexities of this task, so more research is required.

One leading privately held company in the globalization industry, VistaTEC, has partially funded this research into PE (as further detailed in section 1.3). Like many other language service providers, VistaTEC has faced a growing demand for MT and PE services in recent years, and this led to the inclusion of these services among their offerings, distinct from translation and revision. However, the linguists who normally supplied translation services to VistaTEC did not all have adequate experience with MT and PE.

In consultation with VistaTEC, and informed by the existing research on PE (presented in the next chapter, the Literature review), interesting questions emerged:

I. Does the level of previous experience with translation influence the performance of translators when doing post-editing tasks?

   I.b. If so, does the level of experience have a positive or a negative impact on the performance in terms of time spent and fitness for purpose of the final text product?

II. Are the same post-editing strategies employed across languages of the same family? (Test case: French and Brazilian Portuguese.)

The research questions are crucial to determine the scope and the focus of the research. As pointed out by Flick (2009: 103):

> The way in which research questions are formulated exerts a strong influence on the design of the study.

The development of our research questions was guided by the need to explore in more detail the activities involved in PE because it is still a relatively new aspect of the translation and localisation market. Currently, companies that offer PE services do not all follow the same standards. Additionally, there are no universally accepted PE guidelines or instructions. This, combined with the scarcity of PE training, may mean that the transition to PE projects is more difficult for some translation professionals than for others. Therefore, the research questions were formulated with the intention of helping to shed light on the issues and problem areas involved in PE work.

Aiming to better understand the work involved in post-editing, this research concentrates on exploring some of the main difficulties presented by PE and seeks to uncover strategies for training and preparing linguists for this type of work. The insights gained may also be helpful for improving existing MT systems and developing tools and methods for automating part of the PE work. Attention is given to the perspective of translators so that researchers can investigate how personal biases regarding MT may affect their PE performance and ultimately how their attitudes and opinions may affect the integration of MT systems in the translation process.

Some of the primary studies in the field are examined in the Literature Review (Chapter 2), and their contribution to the field is appraised. The literature available on this topic helps highlight the current perspectives on PE, illustrating common practices and expectations and pointing to some of PE's inherent drawbacks.

Since the specific aim of this research is to perform an investigation of PE, other areas that are pertinent to the field of machine translation (such as the results and limitations of different MT systems and the use of controlled language) are outside the scope of the present study and, therefore, will not be reviewed in detail here.

1.2. Relevance

Post-editing is becoming a widespread activity in the localisation industry. This has been observed by many authors, such as Allen (2003), Wolochwianski (2008) and Yunker (2008), to mention a few. However, there is still relatively little information available on the difficulties faced by

translation professionals when carrying out PE, on the effect of previous translation experience on PE performance or on common PE strategies among languages of the same family.

Therefore, the relevance of the two research questions proposed could be summarised as follows:

- Determining if experienced translators consistently perform better in the PE task when compared with novice translators would be helpful for the selection of possible candidates for PE projects and for training.

- If the same strategies can be identified for PE in the two languages to be used in the current study, this could be seen as an indication that similar PE guidelines and training can be adopted for languages of the same family.

- Determining some of the main difficulties found by post-editors would be helpful for devising PE training and guidelines.

- Ultimately, the data gathered can be helpful for the development and/or improvement of automated PE systems. The main strategies identified can be used as an indication of the areas in which PE tools should focus to make this task more efficient, less time-consuming, less tiring and less tedious for translators.

These items would make the results of the present research relevant for translators, translator trainers and developers of MT systems and editing tools as well as for MT users in general.

1.3. Funding

This research was possible thanks to the funding provided by VistaTEC and the Irish Research Council for Science, Engineering & Technology (IRCSET). The funding was provided through the Enterprise Partnership Scheme.

VistaTEC is a privately held, indigenous Irish Localisation Service Provider. The company has its headquarters in Dublin, business development offices in the United States, Translation Partners across five continents and offshore production facilities in Eastern Europe, India and China. Formed in 1997, the company offers services such as: translation to over 100 languages, internationalisation, software engineering, software testing, document publishing, multilingual SEO and machine translation post editing. VistaTEC is recognized in the industry for its quality of service and technical expertise. VistaTEC co-sponsored this research as it believes that Research and Development is key to delivering leading-edge solutions to its customers.

The Irish Research Council for Science, Engineering & Technology (IRCSET) provides funding for researchers across Masters, Doctoral and Postdoctoral levels in the fields of sciences, engineering and technology, with an emphasis on innovation within Ireland. IRCSET is funded by the National Development Plan of Ireland, under the Department of Education and Science.

1.4. Structure of the thesis

Chapter 2, Literature review, surveys relevant works related to different fields that inform the present research. Next, in Chapter 3, the methodology is presented in detail. Chapter 4 provides the analysis of the data gathered. Finally, Chapter 5 presents the conclusion, including suggestions for further research. Detailed data collected during our experiments are provided in the Appendices.

Chapter 2: Literature review

2.1. Introduction

This chapter reviews the literature on several areas of knowledge that are relevant to MT and to Post-Editing (hereafter abbreviated as PE for brevity) and that inform the present research. The chapter covers the development of machine translation in brief, the different types of MT systems, the definition of post-editing, research on integration between MT and translation memory tools, translators' opinions of MT and PE, PE training and guidelines, and automated PE, and concludes by considering research on the profile of a good post-editor.

2.2. Machine Translation: background and context

In this section we provide an overview of the development of Machine Translation. This overview is necessarily brief since, while it is important to our topic, it is not of central importance. Up until the early 1960s, machine translation (MT), a new field of study that, according to researchers at that time, showed great promise for the goal of furthering human communication across different languages, gradually attracted substantial attention and created many new investment ventures. As a new technology, it was seen as a solution that would soon be capable of handling all kinds of texts, providing the same level of quality as human translation (Hutchins, 2000).

A number of different factors were responsible for the growing interest in this area, such as the expansion of international commerce, the evolution of international collaboration in the scientific sphere and the development of multinational bodies in different domains, accentuating the need to make information available worldwide in multiple languages (Loffler-Laurian, 1996). The world perspective was also changing: with the Cold War, access to military intelligence was considered of vital importance, and the American government saw machine translation as one of the means of obtaining information only available in the Russian language, for instance (Somers, 2001b).

In the 1960s, it gradually became clear that early assumptions about the proficiency of machine translation systems would have to be re-evaluated. Despite substantial investments made in research, the results obtained did not correspond to the initial expectations, and the prospect of achieving fully-automated high quality translation still seemed to be a distant goal (Hutchins, 2000). In 1966, the ALPAC (Automatic Language Processing Advisory Committee) report was published, with an assessment of the results obtained for Russian-English MT in the USA. The report provided a very negative evaluation of MT, describing it as less efficient than human translation in terms of cost and speed. It had a considerable impact, not only in the USA, but also internationally, motivating a global reduction in investment in MT research for the following years (Somers, 2003).

More than a decade went by before the interest in MT was renewed, then no longer motivated by military issues, but mainly by the expansion of international commerce (Hutchins, 2000). Gradually, new initiatives emerged, leading to the success of projects such as the MÉTÉO™ system, adopted in Canada for translating weather reports from English into French from 1981 to 2001 (Canadian International Trade Tribunal, 2001), being described as "the world's only example of a truly fully-automatic MT system" (Slocum, 1984, p. 552), due to its ability to detect its own errors (which it then sends to human editors for correction). The first commercial MT systems became available on the market, such as Systran, which was successfully implemented by the European Commission, as well as by General Motors and Xerox, among other organisations, beginning in the 1970s (op. cit., pp. 551-552). From the 1980s onwards, there was also an increase in the development of computer-aided tools, i.e. translation memory and terminology management software (Somers, 2001b).

As the initial goal of fully-automated high quality translation has proved to be unviable, at least with the technology, methods and tools currently available, the work of human translators has been recognised as an important and necessary component for producing high-quality output in the MT process. An important and ground-breaking move was achieved in the 1990's with the advent of statistical MT (SMT), which has become the most widely adopted MT paradigm in recent years (Koehn, 2010).

As well as SMT, rule-based machine translation (RBMT) is also in use nowadays. Additionally, a new model has emerged, Hybrid MT, which is gradually becoming more adopted. Rule-based MT engines utilise bilingual dictionaries and grammatical rules to automatically translate texts from the source language to the target language. They can perform this process in three different ways. As described by Hutchins and Somers (1992), the Direct Approach replaces terms (in a word by word approach) in the source language with terms in the target language, aided by bilingual dictionaries. There is also the Interlingua method, which first translates terms from the source language to what is called an "Interlingua" (pivot representations that are found to be common to more than one language) and, in a second step, translates the terms from the Interlingua representations to the target language. A third method, called Transfer-based, performs an initial analysis step, transforming terms in the source language into disambiguated representations, taking into account syntactic rules of the source language (Hutchins and Somers, 1992). This is followed by a transfer step, comprising lexical and structural transfer for the terms from the source to the target languages. Finally, there is the generation step, in which the translation into the target language is created. The Transfer-based approach is the most common one for RBMT systems, such as Systran. In the case of RBMT engines, there are several tasks that can be performed by human linguists, such as maintaining and updating the internal dictionaries, making adjustments in the configuration options, and testing the output to identify

possible errors that can be eliminated with adjustments and/or additions to the dictionaries.

Statistical MT systems carry out the translation process by applying algorithms and analysing probabilities in relation to parallel corpora of bilingual texts. The translations for individual terms are selected according to their frequency in the bilingual corpora, taking into account "probability distribution and probability estimation" (Carl and Way, 2003: xix). This approach does not involve the use of grammatical rules, unlike RBMT. SMT systems can be continually "trained" with additional corpora, adding new translations to be used when the process is repeated with new texts (Winiwarter, 2007: 345). Examples of SMT systems currently in use are Language Weaver and Google Translate. SMT systems do not require the intervention of linguists with full knowledge of the source and target languages in order to be trained or adjusted, since the approach used is based on statistical analysis rather than linguistic rules.

As mentioned above, a third type of MT approach also exists: hybrid MT. In this type of MT system, aspects of RBMT and SMT systems are combined in a new model. For instance, syntactic and semantic aspects may be incorporated to a SMT system, such as the Microsoft MT engine (Wendt, 2008).

However, regardless of the type of MT system employed, if the output produced is not destined to be used merely for gisting purposes, and a high level of quality and precision is required, the errors and imprecisions that may

be left in the text need to be corrected. This specific activity is termed *post-editing*.

2.3. Definition of post-editing

As defined by Allen (2001), post-editing consists of correcting texts that have been translated from a source language into a target language by a machine translation system. Another useful definition is provided by Somers (2001a, p. 138), who describes it as "tidying up the raw output, correcting mistakes, revising entire, or, in the worst case, retranslating entire sections".

> As automated translation still has many limitations even nowadays, the corrections made by human linguists remain indispensable to make machine-translated texts more understandable and accessible to readers.

The task of PE can be classified in different ways, depending on the volume of corrections and on the effort required. Loffler-Laurian (1984, 1996) has proposed the following typology: fast PE and conventional PE. According to the author, some of the main characteristics of fast PE would be its quick turnaround, and its focus solely on corrections that are essential. In other words, only issues that could seriously hinder the understanding of the text should be corrected, such as incorrect meaning or grammar. On the other hand, conventional (or full) PE aims to produce a level of quality equivalent to that of a text that has been translated by a human linguist. In this case, all issues should be corrected, instead of only the most critical ones. The type of

PE to be chosen would be determined by how the text is intended to be used: for instance, for publication (requiring a higher level of quality), or for gisting (with lower quality requirements, in this case). Other authors follow the same classification mentioned above or offer slight alternatives (Hutchins, 1992; Newton, 1992). Many companies are also adopting similar guidelines to indicate to linguists the type of PE that needs to be carried out for a given text, based on its future usage.

Elaborating on the categories proposed by Loffler-Laurian, Allen (2003, pp. 297-316) suggests a broader typology. MT with no PE, or raw MT output, would be used for texts destined exclusively for gisting purposes, for the fast dissemination of a basic level of information in different languages. Rapid PE, adopted in the European Commission's PE service with the acquisition of the Systran system in 1976, involves only corrections that are strictly essential for rendering a document understandable by the target readers (and, for this reason, many corrections may be deliberately left out, such as issues related to gender agreement). Partial or minimal PE would be used for documents to be distributed to third parties, so it would involve a higher volume of corrections than rapid PE (and corrections that are not included in rapid PE may be required for this level of PE). The author mentions that because there is no universally adopted PE classification system, different companies and even different linguists have diverse interpretations concerning the level of corrections to be used for this type of PE. Finally, full PE would include all necessary corrections, being the most laborious and time-intensive category.

However, Allen remarks that it might be less time-consuming to fully post-edit documents written in controlled language than in uncontrolled language. The reason for this is that, with controlled language, it is possible to avoid including linguistic structures that are known to potentially cause issues in the raw RBMT output.[2] As a result, with fewer issues to correct, the task of post-editing such texts becomes easier. The author also signals that, in the field of localisation, PE is gradually becoming a more common task (Allen, 2003, p. 300):

> With such an increased demand for translation, many companies are actively seeking ways to meet their translation needs within a reasonably affordable price range. Globalization and localization are significant factors that influence MT, and therefore the use of MT post-editing.

2.4. Studies on post-editing

In the field of Translation Studies, and indeed in Machine Translation, research dealing specifically with PE is relatively limited, though growing, especially since this activity has become more widespread only in recent years.

[2] The effects of controlled language on SMT are, as yet, underinvestigated, but research suggests that it may have some benefits for SMT. The results of Doherty's study (2012) suggest that the use of controlled language results in higher scores of text recall, as well as higher scores for eye-tracking metrics such as fixation count and length.

One of the most extensive works of research in the area of PE has been Krings' study (2001).[3] The author uses Think-Aloud Protocols (TAPs, in which the participants of the experiment verbalise their thoughts while performing the tasks that are being analysed) to investigate the cognitive processes involved in PE, as well as the speed that can be achieved by post-editors, the actual cost-effectiveness of MT, and possible areas in which the usage of MT should concentrate. In order to carry out his analysis, he studies a group of translators who perform PE with and without source texts. The texts used in the experiment are restricted to a well-defined domain of knowledge: instructions for using simple technical appliances (Krings, 2001, p. 186). Krings talks about the frequent reluctance of translators to accept MT systems and the implications that can arise from this issue. Some of the topics covered include the approach used by translators for translating a text from one language into another or for post-editing a machine-translated text, and the component processes that may be observed while these tasks are performed, the distribution of sub-processes in PE, as opposed to translation, and the strategies implemented by translators in solving translation and PE problems. He also touches on the linguistic and non-linguistic problems that may appear in translation and in PE processes, the competences required for translation and for PE (also in an attempt to determine if they intersect or if they are fully independent), and the number of alternatives produced in PE and in translation, as well as what determines this number. The author

3 Although Krings' research was published in English in 2001, it dates from 1994, when the author submitted it as his postdoctoral thesis (in German). Therefore, this research work refers to the MT systems and technology available in 1994.

suggests that MT may decrease the effort required in translation by serving as a bilingual dictionary, thus reducing the need to check reference works. One of his main conclusions is that PE with a source text (as opposed to without reference to a source text) involves more cognitive effort than translating. Two of the reasons for this, he suggests, may be that PE (with a source text) requires more source-related cognitive processes than translation (ibid, p. 319) and, in addition, PE also involves more attention shifts between texts than translation (ibid, p. 320). To reduce the effort, the quality of the MT output would need to be high. The author also indicates that medium quality raw MT may require more cognitive effort during PE. In addition, according to Krings, in the presence of MT, the process of comprehension of the source text by the translator may be different from the comprehension of the source text during translation. The results of the author's study suggest that experienced post-editors might work somewhat more slowly than inexperienced ones, as they might pay more attention to detail. Krings' research is comprehensive, exploring many different aspects of PE and covering numerous issues related to it. It may be argued that some of the data offered might no longer be applicable nowadays as MT systems have evolved and changed substantially since this study was carried out. For instance, Krings devotes part of his research to PE done on paper. In the researcher's own experience, this is no longer the usual practice for translation and/or PE. Taking into account the very tight deadlines and the requirements of the translation market today, working on paper might make

translation and PE virtually unfeasible. Furthermore, some of Krings' comments about the limitations of MT systems (for instance, in relation to compatibility) do not apply to current MT programs available on the market. Nevertheless, this still remains a very useful, complete and relevant study for understanding different aspects of PE.

Loffler-Laurian (1996) provides information about several MT systems, their development and usage, and she talks about the profile of a good candidate for performing PE. This is of particular interest for the present research; in fact, the suggested profile of a good post-editor is delineated in Chapter 3, and this is revisited in Chapter 5. Other areas discussed include the reactions of reviewers towards MT and PE, suggestions for improving the work of post-editors, the characteristics of different types of PE and the criteria for using each of them, categories of linguistic issues that affect MT, types of errors that cause modifications in PE, and the usefulness of MT for teaching foreign languages and translation. The typology for different levels of PE and the detailed discussion about the difficulties faced by translators when performing PE are of particular interest for the present research, as they provide in-depth information that cannot be easily found in other studies in this area. This typology will be revisited later, in Chapter 3, which discusses the methodology used for the present research.

O'Brien (2006) carried out an extensive study on the use of controlled language and its effect on PE effort. The study comprises an experiment in which a group of translators post-edit a text that has been previously

prepared, with some NTIs (negative translatability indicators) being removed. NTIs are linguistic features considered to be "negative sentence properties" (ibid., p. 38) for MT, such as relative clauses or words that can have different parts of speech, for instance. NTIs can be used as a measure to determine the suitability of a text for MT. The results of the experiment suggest that some NTIs may indeed increase PE effort, and their elimination could contribute to streamlining the PE process. As well as investigating translatability issues and the use of controlled language, the author provides ample information on PE itself, covering topics such as types of PE and the quality levels expected for each of them, successful implementations of MT and PE in different companies, training and computer-aided PE.

O'Brien and Fiederer (2009) addressed the topic of quality measurement in a study that involved the evaluation of 30 sentences, with English as the source language and German as the target language. There were three versions for each sentence: in the first version, the source text was translated by human translators; in the second version, the source text was machine-translated, and the raw MT output was not post-edited; in the third version, the source text was machine-translated and post-edited. The three versions of the segments were assessed in terms of clarity, accuracy and style by eleven raters.

O'Brien's study published in 2011 provided an investigation on the possible correlation between automatic scores (General Text Matcher and Translation Edit Rate) and PE productivity (expressed in cognitive effort and

speed). An experiment was conducted using Alchemy Catalyst as the PE environment, in which seven post-editors worked on segments machine translated from English into French. Eye-tracking was included to provide further clues about cognitive effort. It was verified that the automatic scores included in the study, GTM and TER, correlated with PE speed for specific groups of segments. The segments were classified in different levels, according to their GTM values, and those with low GTM scores were associated with the highest number of fixations, as recorded by the eye tracker. O'Brien points out that the findings regarding the correlations with GTM levels and cognitive effort may be helpful to predict the level of effort required for projects involving PE, since it would be possible to make an initial estimate of the effort and extrapolate the values to the remainder of a project.

In a study that has points in common with O'Brien and Fiederer's, García (2010) conducted an experiment in which reviewers were asked to rate the quality of translations produced with the use of MT and translations done from scratch by human translators from English into Chinese. The reviewers did not consider the quality of both types of translations to be significantly different, and even slightly favoured the translations produced by MT.

Using Jeffrey Allen's methodology as a guideline for her study, Guerra Martínez (2003) proposes different tests for benchmarking the speed that may be achieved with PE as opposed to human translation. Guerra Martínez provides a detailed explanation of the categories of PE and the approaches

that may be chosen for its implementation. She also elaborates on the different types of professionals that might be suitable candidates for becoming post-editors. According to Guerra Martínez, experienced translators might not be the best candidates, as they might have a negative attitude in relation to PE and to MT systems (although she acknowledges that other authors, such as Krings and Vasconcellos, are in favour of employing translators for this task). The author argues that revisers (i.e. those who work full-time revising texts) might be more equipped for PE. Other areas covered are the maintenance of dictionaries, a comparison between the human translation cycle and the MT cycle (with and without different types of PE) and a detailed description of the MT software used for the study. Of particular interest for the present research are the PE guidelines suggested by the author, which help delimit the work to be performed by post-editors, and also serve as a good example of how to provide guidance for linguists who work on PE projects. By comparing the total time spent on human translation and the time required for the different types of PE, Guerra Martínez concludes that it is possible to obtain an increase in productivity most particularly if an MT system with PE features is used, as this can help post-editors work faster. The MT program used by the researcher for her study, @Prompt Professional, includes a few PE features, such as displaying a list of translation alternatives from which the post-editor can quickly choose, offering quick access to the dictionary for the correction of terms during PE, and aligning paragraphs and highlighted terms for easy identification of selected

terms, among other options (Guerra Martínez, 2003, p. 46-47). She also suggests that it is crucial to select the appropriate MT approach and to ensure that dictionaries are kept up-to-date with the necessary terminology (in the case of RBMT). It is worth noting that the information provided by this research could have been expanded if the researcher had carried out her study with several participants. The researcher herself was the only participant in the experiment, which restricted the scope of her findings. With a higher number of participants, it might have been possible to compare the answers provided and the approaches used, to identify the differences and the similarities in the results originated by each of the participants, as well as the difficulties found. This would have allowed the researcher to gather more data for drawing further conclusions and for obtaining more insights into the MT cycle.

Giving continuity to the research carried out in the end-of-course project for her specialisation degree in translation, Alfaro (1997, 1998) explores some of the main features of MT systems, providing details about the development of MT technology through the years. The study includes a brief usage test of Globalink Power Translator, in which she compares the time required to manually translate a text from Portuguese into English against the time to automatically translate the same text using this software (with and without a subsequent PE step). By analysing the time in minutes required for each task, she suggests that the adoption of MT could indeed be

helpful for saving time and costs, especially by extrapolating these values to a larger-scale project.

2.5. PE and translation memory (TM) tools

Translation Memory is a de facto technology in many specialised translation domains. Many translators, and their clients, see TM as an essential aid in the execution of their day-to-day business. As MT increases in popularity, it is becoming increasingly common for TM tools to be used in conjunction with MT technology in large-scale translation projects. This section reviews studies dealing with the use of TM tools, their applicability for joint use with MT and some of the implications for linguists.

In their study of 2001, Plitt and Bruckner analyse the possibility of combining translation memories with MT output, and whether this could be advantageous in terms of quality and productivity. In order to test this hypothesis, they carry out an experiment comparing the work of two different groups: the first one translates software texts from English into German with the use of a TM without MT, while the second group carries out the translation of the same texts using a TM that also contains machine-translated segments. Their analysis suggests that the use of MT segments could be beneficial when compared to TM fuzzy matches that are ranked lower than 75%, but it would be more advantageous to use the fuzzy matches themselves if they are 76% or higher. Plitt and Masselot (2010) published a

study on PE productivity in which twelve post-editors had to deal with segments either produced by MT or originating from a TM (without an indication of their origin). The environment used was the same PE workbench that was employed in the main PE project of the present research. The results indicated a higher productivity linked to MT segments, and a higher number of errors was associated with the TM segments. Like the present research, this study also included data on keyboard usage (albeit with different settings), with indications that MT segments involved a lower level of keyboard usage.

The results of O'Brien's research (2006), which investigates the cognitive effort involved when translators deal with TM segments and MT segments, seem to confirm Plitt and Bruckner's findings. In her experiment, which also includes the use of eye tracking for obtaining further insights into cognitive effort, four participants translate segments from English into French and German using both TM and MT segments. This study indicates that the effort involved for post-editing MT segments would be equivalent to the processing of TM fuzzy matches ranked between 80% to 90%, assuming that the raw MT output is reasonably good to begin with.

In her minor thesis for a PhD program in Translation and Intercultural Studies, Guerberof (2008) analyses aspects related to quality and productivity involved in the use of translation memory systems combined with MT. The author carries out an experiment in which translators translate new segments, revise pre-translated segments from a TM and post-edit machine-translated

segments (the origin of the different types of segments is not indicated to the participants). At the end, each participant is asked to fill out a questionnaire. After surveying different types of MT and providing information about TM systems and some of the main studies in the field of translation technology, the author examines differences in productivity and quality related to the three types of segment. She correlates the number of errors found for each participant with their processing speed and their level of technical experience. The conclusions are that post-editing machine-translated segments is faster than editing TM segments (80%-90% matches); more experienced translators achieve higher processing speeds for post-editing MT segments, while less experienced translators seem likely to work at similar speeds when dealing with TM and MT segments. Surprisingly, the number of errors is higher in TM segments than in new or MT segments. The author also concludes that translators' technical experience has an impact on the processing speed, which is of particular interest to the present research. Guerberof defines technical experience as experience in software localisation and knowledge of translation tools, subject matter and PE. According to her findings, when the level of technical experience of the participants is correlated with the processing speed, the most experienced translators in the group achieve the best performance when dealing both with MT and TM segments. On the other hand, when the author correlates technical experience with number of errors, there is no significant difference in the performance of the most experienced translators and the least experienced ones. However, Guerberof mentions

that this study included a limited number of participants (nine people in total), and it would be necessary to test the results in a larger-scale experiment in order to draw further conclusions.

Sousa et at. (2011) carry out a study in which PE effort is measured in terms of time. The study comprises subtitles in English that are translated into Brazilian Portuguese with two different MT systems, with a TM tool, and also without the use of any tools. The results found for each of these approaches are analysed in terms of the total time taken, and also with the use of scoring guidelines, as well as BLEU scores. The participants are fluent in both the source and the target languages, and they have some experience with translation (although there is no indication that they are professional translators). The analysis indicates that translating the subtitles without the use of any tools can take up to 70% longer than post-editing the same subtitles. The post-edited subtitles had an average BLEU score of 69.92 in comparison with the translations done from scratch, which suggests that there was no loss of quality for the post-edited segments (although the study does not include a human evaluation to confirm this finding).

Building on the research presented in her minor thesis, Guerberof (2012) develops an in-depth study focusing on the investigation of correlations between TM fuzzy matches and MT segments. This research project includes a larger number of participants (24 translators and three reviewers) and a higher word count for the experiment (2,124 words) than her previous study, as well as a questionnaire and retrospective interviews. The

objectives are to gather insights about the price value for MT match segments, also taking into account the translators' productivity. In addition, her research examines the impact that the use of MT output can have on the final quality of translated texts, and whether translators' experience can influence their productivity and the quality of their work, which is helpful for defining the profile of good candidates for PE work, and is of particular interest to the present research. Regarding productivity, the results found do not indicate statistically significant differences in processing speed between MT and fuzzy matches (Guerberof 2012: 143). The results for quality indicate that the hypothesis that a higher level of quality (measured by the number of errors) would be achieved when using MT segments did not hold true, since the post-editors made a similar number of errors while processing segments from MT and from the TM (ibid: 185). It was also verified that the speed of the post-editors did not influence the final quality achieved for the segments (ibid: 186).

Rieche (2004) provides ample information on the whole localisation cycle, particularly the features and use of translation memories. She draws on her own experience as a professional translator to provide practical examples of difficulties found when dealing with large localisation projects for international companies. Although the focus of her study is the use of translation memories, her chapter on the theoretical basis for the dissertation is very relevant for the present research, as she provides insights on the concepts of equivalence and errors for quality assessment of translation. After

surveying different theories on the notion of equivalence in translation — following traditional, post-structuralist, descriptivist and functionalist approaches — she proposes an operational concept of equivalence for the localisation field, suggesting that equivalence could be seen as relative instead of absolute, and that it would derive from each different context or situation, instead of being defined *a priori* by a formula or algorithm (Rieche, 2004, p. 77). According to the author, some translation clients might require a more literal translation, depending on the goal and the intended audience, while others might opt for a less literal translation, also depending on the purpose of the translated material. Rieche provides examples of non-literary segments that can be translated differently, according to the above-mentioned factors (purpose of the text and intended audience). The texts chosen are an excerpt from a hardware manual and segments related to management practices. She provides different translations for each segment, indicating several terms that could be chosen according to the translation client, the type of document, the target audience, the purpose of the translation and recommendations from the client. By doing so, the author expands on the notion that there may be more than one correct translation for a text. She concludes her explanation by adding that the functionalist view might be the most adequate for the localisation market, since it takes into account the translation client, their requests and recommendations, and it follows a notion of relative equivalence. It would be possible to apply the same concept to PE activity, in which different projects and different target

users determine the amount of changes required (fast PE or complete PE, for instance), to meet different purposes. The author covers different methods of quality control in the localisation industry and provides detailed information on the results of her research into translation memory use among Brazilian translators, by analysing data gathered from a questionnaire answered by 80 participants from the translation community in Brazil (77% of whom identified themselves as technical translators). Some of the conclusions of the study were: when the study was carried out, in 2004, the majority of the translators who took part in it used or intended to start using translation memory programs (63% of the 80 respondents indicated that they used TM programs, and 11% intended to adopt them), confirming similar results found in an earlier survey carried out by LISA in 2002 (The Localization Industry Standards Association, 2002); the systems most frequently used were dictated by the clients' preferences in many cases; most participants (90%) believed that the use of TM systems increases productivity; and there was not a standard method of maintaining TMs. Finally, the author proposes strategies for avoiding the propagation of errors and for maintaining a high level of quality in TMs. Although this issue is outside the scope of the present research, it is interesting to observe that some of the strategies proposed by Rieche (such as the periodic maintenance and correction of TMs at the end of each project, so that they can be reused without disseminating errors, and the incorporation of this process to the overall localisation cycle, in the case of large projects carried out by translation vendors) could also be useful for

helping to maintain a high level of quality when translation memories are used *in tandem* with machine translation systems.

Exemplifying the current trends in the integration between MT and TM technologies, with the acquisition of Language Weaver by SDL in 2010 (DePalma 2010), SDL further integrated its translation memory products with MT functionality. With the two methods integrated in a single process, when a match is not found in the TMs for a given segment, the MT engine automatically translates it. The output can be subsequently reviewed by linguists, including both the TM matches and the machine-translated segments. Wolochwianski (2008b) warns that one of the possible drawbacks of this approach is that sometimes errors might be overlooked by the post-editors "due to the natural flow of the sentence stored in the TM", and because the TM segments previously produced by human translators might seem more reliable than MT segments and, for this reason, editors might not always check them as thoroughly as they would check machine-translated texts. However, as already mentioned, Guerberof's research suggests that this might not be the case.

2.6. Translators' opinions of MT and PE

Nowadays, as more and more companies invest in MT tools and technology, the worldwide translation market and the work of translators seem to be changing shape very quickly, as observed by different authors (Champollion 2001, Zhuang 2002, Yunker 2008, Seeburg 2008,

Wolochwianski 2008b, Guerberof 2008). Wolochwianski (2008a, p. 14) comments that:

> The possibility of being creative in our everyday tasks is becoming more and more limited: we have to follow the glossary, we have to respect the client's preferences, we have to imitate the style in the TM, we need to use Neutral Spanish (if there is such a thing), we have to unify the style of all the translators in the team... and now, we have to post-edit texts that have been automatically translated.

The changes brought about by the advances in MT technology are seen by some translators as a threat: MT tools could be used to replace them, taking their jobs away. This can have a significantly negative impact on the translators' acceptance of this technology and, consequently, on its implementation. As stated by Krings:

> The decision to acquire such a system can trigger acceptance problems among those affected, especially when the proposed system is regarded as a threat to one's own position or as a "job killer". (Krings 2001, p. 33).

Loffler-Laurian (1996, p. 83) also points out that the reactions to the MT output may be not only of a linguistic nature, but also of psychological origin. There may be an element of fear of being replaced by a machine and, as a result, there may be a total rejection of any text produced by a machine or perceived as such.

Negative attitudes towards MT might also be linked to reasons other than the fear of being replaced by a computer system. Brosnan (1998, p. 10) talks about "technophobia", which describes "individuals who resist using computers when given the opportunity to use them". The author mentions that while this is not a phobia in the strict sense, it is well documented in the

literature. Rather than originating from a fear of being replaced by computers, "technophobia" would be a general negative attitude towards technology, not necessarily based on any concrete reasons. Sinkovics et al. (2002, p. 478) mention that such aversion to technology is not related to age, and argue that it might influence the adoption of technological products.

A study on translators' views and experiences regarding MT was conducted by Fulford (2002). By means of discussions and focus groups involving freelance translators, the author tries to gauge the participants' perceptions, attitudes, experiences and difficulties in relation to MT. A low percentage (7%) of the participants indicated that they actively use MT systems. Although approximately half of the 30 translators that took part in the study already had some experience with post-editing, only 23% of them had received practical training on MT technologies. Concerning the views of the participants on MT, the author comments:

> Among the translators in the sample, there was a mix of views about MT, its capabilities, its potential, and its viability. The predominant view was one of scepticism. When probed on this issue, it seemed that this scepticism was founded not so much on a fear that MT systems might pose a threat to the role of the human translator, but rather on a belief that the task of translation is too complex to be able ever to be undertaken effectively by a machine. Those who were most dismissive about the capabilities of MT were, perhaps inevitably, largely those having had the least exposure to it. (Fulford 2002, p. 120)

The answers provided by the participants hint at the need for more widespread MT training as well as PE guidelines. In addition, according to the findings of this study, the majority of the participants (even those who express uncertainty and scepticism regarding MT) are interested in learning more

about MT technologies. The author concludes by commenting that the provision of training resources could have a beneficial effect, promoting a higher level of collaboration among users, developers and academics.

The study conducted by Doherty et al. (2012) seems to confirm Fulford's considerations. Analysing the insights and the data gathered during the development of an introductory course on statistical machine translation for postgraduate translation students, the authors find that the participants report an increase "in their levels of confidence and knowledge of MT in general, and of SMT in particular" (Doherty et al. 2012: 1), and the introduction to SMT also seems to have a beneficial effect on their technical competence and confidence. These results also highlight the importance of MT and PE training.

Araújo's study (2004) analyses the answers of 19 experienced Brazilian translators on whether they believe that MT will ever replace human translation. The answers are compiled from interviews carried out by Benedetti and Sobral (2003, cited in Araújo, 2004, pp. 3-7). Most of the participants reply that they do not believe that MT will ever replace human translators, expressing their opinions with different levels of emphasis. The author analyses the answers, the arguments provided and the terms used by the participants. He verifies that 89% (17 out of 19) of the respondents do not seem to be sufficiently well-informed of the progress and the results already achieved in this field, and they may still hold the same perception that was widespread in the 1950s: MT technology would be capable of producing

translations very quickly with the same level of quality as human translations. The MT systems currently available do not fully meet this expectation; this causes the respondents' disbelief in the efficiency of automated translation. Araújo concludes that it would be useful and important for both experienced and novice translators to become acquainted with the features and limitations of MT. By doing so, they can make an informed decision on whether and when this technology can help them, and how they can contribute to the further development of the field, if so inclined. Most of the respondents that took part in this study are not technical translators; moreover, the number of participants (19) is limited, so it is not possible to say that this is a representative study of the views held by most Brazilian translators regarding MT. However, the results of this research confirm Fulford's findings, in that they stress the need to make the actual capabilities and limitations of MT technology well-known in the translation community, by means of training or other resources, in order to dispel erroneous notions.

In 2010, TAUS published a report with data from a survey on trends in PE (TAUS 2010), with the participation of 75 LSPs. The report indicates that 50.7% of the participants are not carrying out PE projects on a regular basis. For 86% of those who do work on PE projects, PE represents less than 10% of their revenue. Translators' resistance was indicated as one of the problems related to PE by 28.8% of the participants.

From the studies surveyed, it is possible to conclude that the acceptance of MT systems by the translation community is still limited and, if

MT is to be fostered and put to use as an aid for translation, one of the most important ways of achieving this would be to better understand the task and to make relevant training more widely available than it is nowadays.

2.7. Post-editing training

As can be seen from the information covered in the previous section, MT and PE have not yet been fully accepted by translators. The availability of specialised training might have a positive impact on translator attitudes.

Drawing on observations collected from previous studies, O'Brien's paper (2002) mentions several important considerations that need to be taken into account when planning for PE training. The reasoning behind such training would be that the demand for PE work appears to be growing, and currently not many translators have enough experience with this type of activity. By getting acquainted with the features and the functionality of MT programs, as well as with other relevant areas, such as programming essentials (for writing macros, for instance), the different types of PE and the changes and the approach required for each of them, translators would be well-equipped to deal with PE projects, if they so wished.

O'Brien suggests a PE training programme covering several important topics, such as an introduction to MT technology and to the use of controlled language, terminology management, programming skills and text linguistics. In the future, it would be extremely helpful for translators if such programmes would become more widely available. Currently, in the researcher's own

experience, when faced with the prospect of starting a PE project, most (if not all) translators only receive a limited set of instructions from the client, which may or may not answer their most common questions, and which, in some cases, may actually create more doubts and contribute to their negative perception of MT.

This situation appears to be changing, however. A growing number of universities are adding PE training modules to their translation courses. This might be a new trend, following the expanded use of MT and PE in the translation market. As part of the present research, data were gathered on universities offering such courses in French, Spanish, Portuguese and Chinese. An institution that seems to offer one of the most complete training programmes in PE at the time of writing is the Chinese University of Hong Kong (2009). The Department of Translation offers the course "Editing Skills for Computer Translation", which is described as follows:

> This course introduces the concepts and skills essential to the editing of the source and target texts before, during and after computer translation so as to optimize efficiency and translation quality.

The topics covered in the 13 weeks of the course include lexical, grammatical, semantic, pragmatic and cultural aspects of PE, as well as the integration of MT editing and CAT tools.

In order to investigate what PE guidelines should receive special attention in PE training programmes, Depraetere (2010) carried out an analysis of 2230 words post-edited by translation students. Only the essential instructions were provided for the PE task; this was done on purpose, so that

the researcher could investigate which corrections the participants would implement according to their own judgement. The analysis indicates that, while the students followed the specific instructions for the task, in some cases they left out corrections that should have been implemented. The researcher comments that this could be an indication of how novice and more experienced translators approach a PE task differently. This is relevant for the present research, and this aspect will be revisited in Chapter 5, in the concluding remarks.

As the demand for post-editors is likely to grow in the future, it is possible to anticipate that more universities and organisations related to translation studies will begin to offer PE training courses as well. Recent examples have included the tutorial on PE presented by O'Brien, Roturier and de Almeida at the AMTA Summit in 2009 (O'Brien et al., 2009), the workshop on best practices for PE presented by O'Brien in Amsterdam (O'Brien, 2011), the module on PE that Ana Guerberof teaches at the Universitat Rovira i Virgili (2012-2013) and the PE training she conducted at the Colegio de Traductores Públicos de la Ciudad de Buenos Aires (2012).

2.8. Post-editing guidelines

As PE can still be considered a fairly recent activity in the localisation industry, it is not entirely surprising that standard guidelines have not been developed yet. As mentioned by Allen (2003, p. 305),

> In most cases, there appears to be a missing link between the development of the systems and the training on how to use them and the resulting output. This is definitely an area which requires improvement for enhancing translation and post-editing productivity.

As companies normally develop their own sets of PE guidelines exclusively for internal use, without making them publicly available, it is not possible to comment individually on them. However, what they seem to have in common is an attempt to provide at least a set of general guidelines, so that translators can adequately perform the PE task. This may include instructions aimed at helping translators to quickly decide if a machine-translated segment can be useful or not (e.g. number of seconds to spend deciding whether to post-edit or to re-translate), details about the types of corrections to be made (e.g. whether to fix capitalisation or punctuation errors or not) and about the level of final quality expected for the project, for instance.

Despite the lack of publicly available PE guidelines, a noteworthy exception in the past were the Microsoft® style guides (Microsoft Language Portal 2008), which were made available online in April 2008, with the public launch of the Microsoft Language Portal. Although not all the style guides available in the portal for over 30 languages contained specific instructions about how to deal with machine-translated texts, some of them, like the Spanish style guide, included a full section with detailed information about PE. The instructions covered items such as: definitions of MT and of the different levels of quality that may be required for different projects; how to

deal with several types of lexical, grammatical and other issues (such as synonyms, verb forms, pronouns and capitalisation); how to differentiate what is acceptable or not in PE (for instance, if the word order is grammatically correct, it can be left unchanged in the MT output, even if it might not be the most elegant style). This could be considered a good example of how PE guidelines may be devised to anticipate common issues faced by post-editors and to help them work more efficiently. However, in recent editions of the publicly available style guides (such as the editions available online in 2012), Microsoft has removed all the sections related to PE for all languages.

In a study on the PE of machine-translated output for SAP, Schäfer (2003) proposes a definition of the tasks and cognitive skills involved in PE, as well as discussing a typology of MT errors. The outlined typology is suggested for use with different language pairs, since the author comments that there is a level of similarity among the types of PE corrections required for different languages. Some of the other issues discussed are the importance of having an open mind towards PE and MT, the need to check machine-translated sentences against the source text, in order to avoid overlooking errors, to help linguists develop their skills at recognising recurring MT errors, and to gather examples of areas for improvement in the MT system. The author then provides detailed information about the PE guidelines developed for SAP projects. The objective of the guidelines is to help linguists understand the PE task and develop a positive attitude towards it. The guidelines divide the PE process into the following steps: general

output check, for identifying the main recurring issues in the MT output, such as words to be included in the dictionary; editing the MT output, according to the typology of errors provided; proofreading, to detect semantic errors and to ensure adequacy of style. The typology classifies the errors as: lexical, syntactic, grammatical and due to defective input text. The author provides examples of these categories in different languages and concludes by mentioning that the guidelines are a work in progress, to be complemented with the introduction of controlled language in SAP projects. While the complete PE guidelines are not made available, this is a very useful example of how guidelines can be used to help companies make the MT cycle more efficient, and to assist linguists in the PE task by providing the necessary knowledge, definitions and clearly-defined error categories to be corrected.

However, as mentioned previously, guidelines are not always readily available and, as a result, post-editors may be faced with the prospect of not always being sure of how to proceed, or having to unnecessarily correct the same mistakes over and over again. This may prove to be tedious, discouraging many translators from accepting further PE projects.

When guidelines are provided, sometimes they may be unnecessarily detailed and lengthy, causing confusion. As observed by Allen (2003, p. 313):

> [...] much energy can be wasted on (re)creating principles to tell post-editors to fix up the highly frequent, small MT raw output mistakes that unnecessarily add to the cognitive load on these experienced language experts.

In the researcher's own experience, PE guidelines provided by different companies lack detail, especially taking into account that PE may still

be a new activity for many linguists. Many of the guidelines are not publicly available, being used by internal teams and by language vendors only. This precludes a thorough individual analysis here; however, a few common trends can be detected among the different PE guidelines to which the researcher has access. The instructions tend to focus on the speed required to complete the PE task, highlighting the importance of avoiding unnecessary corrections. Post-editors are given an overview of the main types of errors to be fixed, classified by severity levels. Sometimes practical examples are also included to illustrate the error categories. Ideally, many examples taken from real machine-translated texts should be provided, but this is not always the case. It is also worth highlighting that the guidelines usually do not present a clear distinction between the different types of PE. Many linguists who currently perform PE may not be familiar with the different levels of PE, such as the differences between fast and full PE. Therefore, one area of improvement for existing guidelines would be the inclusion of a section providing details on the different types of PE. Linguists would then be better equipped to deal with the task at hand, and would have a clearer understanding of PE in general terms. An initiative to fulfil this need was the project developed by TAUS regarding PE guidelines (TAUS 2011). Although TAUS' guidelines are generic and high level, they were formulated following the analysis of private guidelines from different companies, so they reflect current guidelines in use, while suggesting ways to improve on them.

2.9. Automated post-editing

In an attempt to minimise the need to repeat the same corrections numerous times, thus making the activity of PE less tedious for translators, and also producing gains in productivity and reducing costs, many solutions for partially automating PE are already available nowadays. For instance, Parton et al. (2012) describe APEs (automatic post-editors) used to correct adequacy errors: "deleted content words, content words that were translated into function words, and mistranslated named entities" (Parson et al. 2012, p. 111). Different APE techniques are explored, namely: rule-based, corpus-level feedback and phrase-level feedback. The analysis shows that the use of the APEs improved the adequacy of the raw MT output 30-56% of the time (Parson et al. 2012, p. 117), but there was a trade-off between fluency and control over which errors to correct, indicating the need for further development of the error detection module of the APE system.

To illustrate the reasoning that is behind automated PE modules, Hutchins (1992) mentions that, with interactive PE, the user could be alerted to incorrectly translated units, and would have the option of automatically correcting similar errors in all remaining translation units.

One of the first authors to write about the automation of some of the PE tasks was Muriel Vasconcellos, who worked for many years in the Machine Translation Program for the Pan American Health Organisation (PAHO). Vasconcellos (1986) comments on the use of macros for quickly

moving portions of text or for replacing specific constructions while performing PE tasks. The macros can be useful, for instance, for changing Verb-Subject-Object constructions into Subject-Verb-Object constructions, when post-editing texts machine-translated from a Romance language into English. Although this article dates from over 20 years ago, it suggests techniques that could still be useful for post-editors nowadays, such as efficient use of the keyboard and of the search and replace functionality. Aymerich and Camelo (2009) discuss the tools and the environment involved in the use of MT at PAHO, commenting on several practices adopted to optimise the MT workflow, such as liaising with the post-editors to gather useful feedback in order to improve the MT engine, using databases for terminology search and using a translation tracking system to control all aspects of the workflow.

Allen (2001) has worked on the development of a translation tool to automate part of post-editors' work by means of interactive PE. The tool allows a translator to choose an alternative translation with a single click, for instance, instead of having to retype it. The author calculates the time spent to translate different texts with and without the tool and concludes that it can indeed help increase productivity. In addition, the dictionaries created with the tool may be used to ensure consistency in future projects.

There are studies underway for the development of statistical post-editing. Simard et al. (2007) review the use of a statistical MT system as a second step in the MT cycle, for automatically post-editing the output generated by a rule-based MT system. Due to the repetitive nature of errors

found in machine-translated texts, this strategy may help reduce the actual post-editing effort, when the texts are subsequently sent to human translators for manual PE. The statistical MT engine is trained to automatically correct frequent errors found in the machine-translated input (which, in this case, the SMT system considers to be the source language), with the use of human translations as reference sentences. The authors offer proof of the improvements that may be achieved with this method by demonstrating that the automatically post-edited output presents a high BLEU score in comparison to the texts generated by the rule-based MT system. More specifically, the authors report:

> (...) a reduction in post-editing effort of up to a third when compared to the input rule-based translation, and as much as 5 BLEU points improvement over the direct SMT approach. (Simard et al., 2007, p. 1)

BLEU scores are a method for automatically comparing a number of MT systems. According to Papineni et al. (2002, p. 1), "[t]he closer a machine translation is to a professional human translation, the better it is." This notion was the basis for the development of the BLEU metric, in which an algorithm is used to calculate the score of the segments taking into account the whole corpus of reference human translations. One of the criticisms that could be made regarding the BLEU score (and others of a similar nature) is that it does not provide an indication of the actual usefulness of the analysed texts. Instead, it calculates the number of correct segments according to the reference texts used, and this may not always be a precise indication of acceptable translations (Isahara et al., 2008, p. 3071; Ananthakrishnan et al.,

2007). Nevertheless, this still remains a widely used score, which can be helpful for assessing the quality of machine-translated texts, if its limitations are taken into account.

The reuse of PE corrections in order to improve existing MT systems is considered by Font Llitjós (2007), who proposes an automated method to gather and adopt corrections made by non-expert users. The corrections can be incorporated to fix frequent errors, and to improve the MT system itself. The users can input corrections using an online interface. The method can be applied to different language pairs, and it is helpful for languages lacking large amounts of parallel data, such as minority languages from South America. A rule refiner module extracts the errors identified by the users, and employs the data to trace the errors back to the rules and to refine them. This is accomplished by means of an automatic rule refinement algorithm. By testing this approach with different language pairs, the author verifies that there is an improvement in the translation quality of the output, as indicated by MT evaluation metrics (such as BLEU scores).

Another approach for improving translation quality is proposed by Itagaki et al. (2007), by means of an automatic consistency validation method. The authors extract a list of compound nouns from their source corpus in English, then extract a bilingual phrase table from the corresponding corpus in Japanese, and use an automated classifier to find translations for the source terms in the list from the target language. The final

step consists of measuring terminology consistency, also by using an automated consistency index.

Guzmán (2007, 2008) has published articles on the use of regular expressions to automate part of the PE task, specifically for rule-based MT output. Since the errors produced by rule-based MT systems tend to be repetitive, this method can be useful for reducing the subsequent manual PE effort (although statistics are not provided to indicate by how much the effort can be reduced). The author suggests the provision of post-editing context to disambiguate mistranslations (ibid 2008, p. 1-2), so that mistranslated segments can be automatically corrected by regular expressions. This can be done with the creation of a translation memory in which source and target segments are kept separate. The tags contained in the translation memory are used to further specify the scope of the regular expressions. As explained by Guzmán, translation units have tags separating the source and target segments. These tags can be used as anchor points to indicate where the source text ends and the target text starts in regular expressions (Guzmán, 2008, p. 2). This can help fix many issues, such as mistranslations of subordinate clauses or of verbs with several meanings, as exemplified by the author. The automated step with the use of regular expressions can immediately precede the human PE step, helping reduce the overall effort.

Companies who aim to implement MT systems should strive to train and inform their translators of techniques such as the ones reviewed in this section, and on the functionality, the limitations and the characteristics of MT

technology, so that current and future post-editors are well prepared for dealing with PE projects.

2.10. Profile of a good post-editor

As well as developing new tools to make PE more efficient and elaborating training programmes that can help prepare linguists for this type of activity, it is also important to try to determine who might be the best-suited candidates to potentially become proficient post-editors.

As part of a study on the use of domain-specific machine translation by a language service provider, Offersgaard et al. (2008: 153) tried to determine the profile of the "ideal post-editor". According to the authors, many of the characteristics that are commonly attributed to good translators could also be considered as inherent to good post-editors. However, based on a PE assignment carried out by the language service provider that took part in their research project, the authors concede that good translators might not necessarily be good post-editors. This might be linked to an important PE skill identified by the authors of this study: the capacity to decide quickly whether a machine-translated segment can be useful or whether it would be better to ignore it altogether and to translate the source segment from scratch. In the researcher's experience, some translators find it difficult to make this decision with the required speed, as imposed by the constraints of the PE task (such as very short deadlines, and an expected high daily productivity). This would make them less suited to work as post-editors - although they could

presumably learn this skill over time (this aspect will be revisited in Chapter 5). Furthermore, this potentially important ability also has many other ramifications: whether or not PE can be taught and developed as a skill, what the best way is to do so, determining whether or not previous experience as a translator can be helpful for a future post-editor, and how the demand for higher productivity affects the work of a post-editor.

2.11. MT and considerations on throughput

As localisation companies seek to achieve higher savings, it is inevitable that machine translation will be increasingly used, with the expectation of dealing with higher volumes of words at a lower cost. This may place more pressure on linguists, who may be expected to produce more words per day for lower rates of pay, as PE is supposed to be a means of automating repetitive tasks, streamlining actions such as corrections of frequent errors and standardisation of terminology. In addition, PE may entail a lower volume of corrections than that of a full revision, depending on many factors, such as the level of final quality expected for a PE project and the target audience. In this case, the volume of post-edited words per day would be expected to be proportionally higher than the volume of revised words per day.

Van der Meer (2003) comments that "documented cases show that post-editing can be done two to three times faster than translation", which may lower translation costs to US$0.095 per word for companies investing in

MT, representing a cost reduction of 40% in the total investment. Ryan (1993) observes that PE productivity may range from 2 to 6 pages per hour, the equivalent of "2 to 6 times faster than human translation". Allen (2004) adds that while a translator can achieve a productivity level of about 2400 words per day, taking from 3 to 4 days to finish a project of over 8000 words, the same project could be done in about one day of work with the use of MT and PE, which corresponds to a productivity increase of 25%-30%. Offersgaard et al. (2008: 158) report "a productivity gain of 67% saved time in post-editing in the test of SMT". While there are discrepancies in the numbers reported above, all of them seem to confirm that it is possible indeed to increase productivity (in terms of volume of words) with the adoption of PE.

In their study carried out in 2011, Carl et al. analyse aspects such as the time taken for completing a PE task, the effort involved and the quality of the post-edited segments. The study involves manually translated segments and automatically translated segments that were post-edited, with English as the source language and Danish as the target language, Google Translate as the MT engine and Translog as the PE environment. The translation and PE tasks were performed by seven translators. The quality of each translation was assessed by seven native speakers. The analysis indicates that the evaluators found the post-edited translations are slightly better than the manual translations. Regarding time, the findings of this study indicate that, on average, the PE task was performed more quickly by the participants than the translation task, although the results do not show a significant difference

of speed between manual translation and PE. The reasons for this may be related to the fact that, although all the participants had experience with translation, none of them had experience with PE or with CAT tools. The experiment also involved the analysis of gaze data, and the results point out that, when manually translating, the participants would read the source text first, then work on the target text, while they would do the opposite when post-editing.

Tatsumi (2009) conducts a very thorough analysis of automatic metric scores with a view to identifying an efficient method for gauging PE effort for Japanese. She conducts an experiment with three participants, who post-edit segments machine-translated from English into Japanese, using TER, GTM, NIST and BLEU as metrics and giving consideration to segment structure and length, among other aspects. The results of her study suggest a significant correlation between GTM scores and PE speed, but she also identifies factors that can have a strong effect on PE speed, such as sentence length and structure and errors in the raw MT output. Tatsumi and Roturier (2010) further analyse the factors that can have an impact on PE effort, including the complexity and the ambiguity of the text to be post-edited. Their results suggest that although these two aspects (textual ambiguity and complexity) play a role in the cognitive effort, they seem to have only a moderate correlation with PE time.

Specia et al. (2009) investigate how to predict the quality of MT output without the use of human references, by using regression estimation to

analyse different MT systems and language combinations. The results of this study show that, with this approach, it would be feasible to analyse different automatically-generated translations and select the ones that would be of higher quality, saving time and effort for the post-editors, who would not need to do the selection themselves. Specia (2011) also conducts an experiment in which translators are requested to score the quality of post-edited segments from 1 to 4, in which the lowest score means that the segment would need to be completely redone, and the highest score means that no corrections are required in the segment. The participants also post-edit the segments, and Specia analyses the results also including PE time and edit distance (recorded during the PE task). Her findings indicate that confidence estimation methods that take into account the annotations on translation quality provided by the post-editors provide a reliable classification of PE effort.

2.12. Emerging trends in PE

In the fast-paced world of localisation, companies are constantly seeking new strategies and solutions to cope with the growing demand to produce higher volumes of translated words. This motivates the adoption of new technologies and the investment in methodologies that can streamline the localisation cycle as a whole, contributing to an increase in productivity and a lowering of costs. In recent years, a new alternative devised to bring additional savings to the localisation industry has gradually been gaining

popularity: the practice of *crowdsourcing*. This method, which can be summarised as recruiting non-expert users to translate projects for free (or for small payments), has attracted companies such as Facebook, which, beginning in 2007, has relied on its users for the translation of its interface into more than 50 languages (García, 2009). The use of crowdsourcing precedes Facebook's adoption, and it has been the method of choice for many types of projects in fields other than the localisation industry, with tasks ranging "from labelling images with keywords to judging the relevance of search results to transcribing podcasts" (Kittur, 2010). Research on crowdsourced translation has been in consistent development in recent years, with many articles having been published (O'Hagan, 2009, Zaidan and Callison-Burch, 2011, Koehn, 2011, Tatsumi et al., 2012, to mention just a few) exploring the possibilities of this approach and investigating its combination with other methods, such as the generation of corpora to train SMT systems (Ambati et al., 2012) and the crowdsourcing of post-editing of raw MT output (Aikawa et at., 2012). Although crowdsourcing is outside the scope of the present research, it is important to mention that it might have an impact on the MT scene, especially if more companies start to opt for crowdsourcing the PE phase of their localisation projects, for instance. The wider adoption of crowdsourcing has the potential to affect the translation profession in different ways, such as a reduction of the volume of work and a lowering of prices. It is still too early to gauge exactly what the impact will be and what changes (if any) will be brought about, but it is important to

contemplate such potential developments. Another emerging trend (still in its infancy) is the recent research into monolingual PE (such as Koehn 2010), which investigates PE involving post-editors with no knowledge of the source language.

The reduction of costs, which is the motivating factor for companies to start considering crowdsourcing as an option, is also one of the main reasons why MT combined with PE has become a more popular alternative for localisation projects in recent years. Specifically in the case of MT and PE, it would be extremely important not to lose focus on the human aspect, however, taking into account the need to provide adequate training, tools and guidelines to the linguists who, by virtue of a suitable profile, may be selected to perform PE tasks. Measures to ensure job satisfaction and to decrease work pressure should be prioritised, so that translators do not feel demotivated to the point of contemplating leaving the profession. Ultimately, this would be one of the key issues to ensure the success of MT implementation.

2.13. Conclusion

This chapter presented an overview of several aspects of PE and provided relevant information about the literature available on each of them. A summarised chronology of the development of MT was provided, focusing on the main reasons that have motivated the interest in this technology over the years, and the different approaches that have been adopted for its

implementation. After providing definitions of PE, this chapter explored the typology that has been suggested to classify this activity according to the level of effort or time required. In addition, some of the main studies in the field in English, Portuguese and French were reviewed. The views held by translators regarding MT were subsequently discussed, as were a proposal for PE training, the need for PE guidelines and the usefulness of automated PE modules. The chapter concluded by discussing the emerging trends in PE.

As mentioned in our Introduction, the Literature Review, along with discussions with one of the research sponsors (VistaTEC) regarding issues related to PE (such as difficulties faced by potential post-editors, and the development of PE training and guidelines that could be successfully used across languages of the same language family with minimal adjustments, in order to streamline the PE workflow) served to identify research questions that were deemed worthy of exploration. Some of the themes that emerged as most interesting to us included the question regarding translator experience and its impact on the PE task, and the potential similarities within the PE task across languages from the same family. Both of these themes touch in their own way on the important topics of quality and training, which will influence the discussion throughout the following chapters. To recapitulate, then, the following research questions were identified:

I. Does the level of previous experience with translation influence the performance of translators when doing post-editing tasks?

I.b. If so, does the level of experience have a positive or a negative impact on the performance in terms of time spent and fitness for purpose of the final text product?

II. Are the same post-editing strategies employed across languages of the same family? (Test case: French and Brazilian Portuguese.)

The next chapter will describe the methodological approach adopted in order to seek answers to the research questions.

Chapter 3: Methodology

3.1. Introduction

This chapter covers several important items related to the preparation for the main PE experiment, including both theoretical and practical issues. Firstly, in section 3.2, the two pilot studies are discussed. Next, in section 3.3, important aspects are examined in detail, such as the research design, the variables used in this research, the corpus, the selection of participants and the validity of the data. The typology developed for this study is discussed in detail in section 3.4. The definition of a good post-editor formulated for the present research is introduced in section 3.5, followed by an explanation about outliers, in section 3.6. Section 3.7 discusses the PE environment used for the main PE experiment, the preparation steps carried out, the selection of participants and the PE instructions provided. The data analysis procedures are explained in section 3.8, followed by a discussion on the validity and trustworthiness of the findings, in section 3.9. This chapter concludes with a summary of the information presented here.

3.2. Pilot studies

Prior to the main PE project, two pilot studies were conducted in order to test the design, the setup and the data collection methods to be used.

3.2.1. First pilot project

The first pilot project was carried out taking advantage of resources available through a localisation project in the IT domain conducted by VistaTEC. This imposed a few limitations, but the main objective was to perform a small-scale initial project to test the methodology and the setup for the first time and to identify areas that required further improvement.

Setup:

The first pilot followed the same setup that was later employed for the main PE project: the participants received instructions for the PE task, the individual PE sessions were scheduled, the participants connected remotely to the researcher's computer in Dublin and post-edited segments that had been previously machine-translated, while the sessions were recorded with Camtasia Studio and InputLog. These two programs are explained in more detail in section 3.7.2.

Languages:

The source language of the localisation project was English, and its target languages were French and Spanish. Since the first pilot used resources from this project, it was not possible to have participants for Brazilian Portuguese. Still, this was a valuable opportunity to perform an initial test of the methodology (albeit with a limited scope). Therefore, French and Spanish were used as target languages for the first pilot (also taking into account that Spanish belongs to the same language family as

French and Brazilian Portuguese and was therefore also relevant for our second Research Question).

Participants:

The first pilot project included six participants, three per target language. This low number was due to the translators' limited availability, the localisation project's scope and the small scale planned for the first pilot.

For comparison purposes in the analysis, the translators selected for the first pilot had different levels of professional experience (in number of years). Some (but not all) of the participants had previous experience with PE. This was a deliberate choice, so that it would be possible to examine what differences, if any, might emerge among participants with different levels of experience when carrying out the PE task.

The researcher contacted the potential participants by e-mail. The initial e-mail contained summarised details about the experiment and an invitation to take part in it. After agreeing to participate in the pilot project, the participants signed a consent form, as determined by the DCU Ethics Committee. Full details (including Plain Language Statements and Consent Forms) and instructions about the pilot project were then sent to the participants.

For ethical reasons, the participants were informed in advance that their work would be recorded by InputLog and Camtasia Studio. The participants were also informed that their participation would be totally

anonymous, and that their personal identification details would not be disclosed.

The participants were asked to provide details about their educational background, as well as their level of translation and PE experience (in years and in number of PE projects, respectively).

Instructions:

The instructions sent to the participants about the PE task (which are explained in more detail in section 3.7.4 and provided in Appendices C and D) indicated the level of quality expected (intermediate). The instructions deliberately did not ask the participants to revise the file after post-editing it. This was done with the objective of allowing the participants to decide whether they would revise their work at the end or not. The participants were also informed that, while performing the PE task, they could refer to any online sources they might consider useful (for clarifying doubts about the meaning of specific words, for instance).

The participants were asked by e-mail to provide feedback on the task after completing it, pointing out the main difficulties found, in their opinion. The comments provided by the participants were not included in the analysis of the data; instead, they were used to help identify areas of the research design that might require further improvement.

Corpus and PE environment:

A file from the localisation project, totalling 350 words in English, was selected for the first pilot. The reason for selecting this specific short

number of words is that dealing with a higher word count might not be feasible within the constraints of this pilot experiment (i.e. translation environment, available files, and time limitations). Since the participants were professional translators who were working on the localisation project, they were likely to be very busy. Their availability was limited and they might not be able to take part in a PE experiment involving a larger word count.

The availability of the resources from the localisation project dictated the file format and the translation/PE environment (Idiom Workbench) to be used in the first pilot. These constraints also precluded the possibility of selecting specific segments from different files. Therefore, the sequence of segments from the source file selected was kept unchanged.

MT engine:

The MT engine used for the pilot project was Language Weaver. The reason for selecting it was that this was the statistical MT system employed for the localisation project. Due to time and budget constraints, it would not be feasible to choose a different MT engine and to train it from scratch, whereas Language Weaver had already been trained for the localisation project with approximately 3 million words from previous projects for the same client both for French and Spanish.

Analysis:

The PE items recorded during the individual sessions were quantified and classified according to the typology devised for the present research. The typology is explained in detail in section 3.4 of this chapter, but it can be summarised as a customised version of classification used by the LISA QA Model (which is explained in detail in section 3.4.1), with the addition of master categories (Essential Changes, Preferential Changes, Essential Changes Not Implemented and Introduced Errors, as further explained in section 3.4.4), main categories (following most of the categories used in the LISA model, in this case) and subcategories (with the inclusion of several subcategories from the GALE PE Guidelines, also explained in more detail in section 3.4.1, and a few subcategories added by the researcher). The results were cross-referenced with the level of translation and PE experience of the participants.

Findings:

A detailed comparison between the results of the first and the second projects and the main project is provided in Chapter 4. For the moment, it suffices to provide a synopsis of our findings and the lessons learned.

The two most experienced translators (in number of years) for both languages were also the fastest post-editors, as well as the two participants who made the highest number of what we termed Essential Changes (as explained in section 3.4 of this chapter).

The breakdown of PE edits implemented (or overlooked) by the participants showed many similarities across both target languages. For French, changes in the Language category accounted for a high percentage of the Essential Changes (49%). The Preferential Changes concentrated on the categories Lexical Choice (40%) and Language (37%). The post-editors overlooked only a small proportion of essential changes (mostly in the Language and Mistranslation categories, which together accounted for 66% of the essential changes not implemented). For Spanish, changes in the Language category accounted for nearly the majority of the essential changes (47%). There was a significant number of Preferential Changes (39 in total), mainly in the Language category (46%). The category Language accounted for 57% of the Essential Changes Not Implemented. By comparing the values recorded for French and Spanish, it is possible to see that there were many parallels between the two languages. The only differences were related to the category Essential Changes Not Implemented: for French, the subcategories with the highest values were Language and Mistranslation, while for Spanish the subcategory with the highest values was Language. No Introduced Errors were recorded for either of the target languages. In fact, as will be explained in section 3.2.2, the master category Introduced Errors was added to the typology when the analysis of the data from the second pilot project identified that the post-editors had made errors that were not originally present in the raw MT output.

Lessons learned from the first pilot:

Although the instructions mentioned that the participants were allowed to perform online searches to check for the meaning of words or to clarify other doubts, they did not follow the guidelines, and they performed searches on their own machines, thus making it impossible to record the searches as part of the PE sessions.

There were a couple of issues with the keyboard layout (two of the participants were using a keyboard layout on their remote computers that did not match the keyboard layout in the researcher's machine in Dublin), but this problem was promptly solved.

The issues identified in the pilot project were taken into account to make the instructions clearer. They also highlighted the need to check in advance if there were any difficulties or shortcomings that could be avoided with more careful planning.

3.2.2. Second pilot project

The second pilot project was carried out taking advantage of resources available through the EYECON project (O'Brien 2011). This was an independent project developed in DCU with the aim of analysing the correlation between PE effort and MT automatic evaluation scores. It consisted of individual PE sessions, in which eye tracking data were recorded during PE tasks. The fact that resources from the EYECON project were used imposed limitations on the second pilot project, as will be explained in more detail in the next sections. However, once again this

was a valuable opportunity to further test and refine the methodology before carrying out the main PE project.

Corpus:

The language combination used in the experiment was English (source language) and French (target language). The MT training corpus consisted of segments from Symantec material, totalling 55,000 sentence pairs.

Segments used:

A total of 10,000 sentence pairs from the corpus were reserved for the PE experiment. All the segments were classified either as Low (for sentences with a GTM score between 0 and 0.4), Medium (for sentences with a GTM score between 0.41 and 0.8) or High (for sentences with a GTM score between 0.81 and 1).

Twenty segments from each of the three GTM categories were randomly selected for the experiment, totalling 60 segments. In addition, three segments from each category were also randomly selected to be used in a warm-up task preceding the actual PE task, totalling nine segments. The segments for the warm-up task and the PE task were used to create two files in Alchemy Catalyst format (Catalyst was the translation environment used for the experiment).

A glossary was prepared with key terminology from the translated segments in the corpus.

MT engine:

The MT engine used for the experiment was the MaTrEx SMT system (Du et al., 2009). MaTrEx was trained with 45,000 sentence pairs from the corpus provided by Symantec. The segments used for training the system had already been translated from English into French. The 10,000 sentence pairs reserved for the PE experiment were then machine-translated from English into French.

Participants:

A total of seven participants were invited to take part in the PE experiment. All the participants were professional translators and native speakers of French, and had previous experience using Alchemy Catalyst (with only one exception). After agreeing to participate in the EYECON project, the participants signed a consent form, as determined by the DCU Ethics Committee.

For ethical reasons, the participants were informed in advance that their work would be recorded using Tobii Studio, so that eye tracking data could be gathered for each PE session. The participants were also informed in advance that their participation would be totally anonymous, and that their personal identification details would not be disclosed. The participants were asked to provide details about their level of translation experience using Catalyst and their PE experience. These details were necessary for the analysis, in order to determine if there was a correlation between the level of experience and the PE performance, and also to

make sure that all the participants met the requirements for taking part in the experiment.

Setup:

The setup for the second pilot presented differences in relation to the first pilot and to the main project. The differences were imposed by the specific setup used for the EYECON project.

A participant was selected to take part in the pilot phase, which was used to verify if any adjustments were required in the steps, tasks and instructions included in the EYECON project. The pilot phase included the same instructions and files that were used in the subsequent sessions.

During each individual session, the remaining participants received instructions about the tasks to be performed. The instructions covered the warm-up task and the PE task. An explanation was also provided regarding the level of PE required: the types of issues to be corrected were grammar errors, inaccuracies and mistranslations; on the other hand, any preferential or stylistic changes should be avoided (for instance, replacing a correct term with a synonym).

A time restriction was not specified for the participants. However, they were asked not to interrupt their work until the PE task was completed. The participants were also informed that, while performing the task, they should post-edit each segment only once, and they could use the glossary provided as a reference. The reason for the requirement of working on each segment only once was that the sessions were recorded with an eye-tracker, and the data referring to each segment had to be

recorded individually. In addition, the objective of the EYECON project was to measure PE effort, and going back to previously post-edited segments in order to make further changes would be outside the scope of the project, which focused on first-pass edits only.

The warm-up task was carried out first (nine segments), followed by the PE task (60 segments). The purpose of the warm-up task was to help the participants get acquainted with the instructions and the material, It also gave them the opportunity to ask questions and clarify any doubts before the PE task.

While each participant post-edited the text, their session was recorded with the eye tracker, including fixations and eye movements, as well as the actions performed onscreen. After the warm-up task and the PE task, the participants were asked to answer a short questionnaire, divided into three parts: how they would classify the quality of the raw MT output for 12 random segments from the corpus; what kind of indicator they would like to have in the interface of the PE environment regarding the quality of raw MT segments; and whether they would like to make any additional comments. The purpose of the questionnaire was to gather data about the participants' preferences regarding different methods to indicate the quality (expressed by GTM scores) of raw MT segments in the user interface of the editing environment (Catalyst). The indicators can help post-editors differentiate segments (in case there are MT segments and TM matches, for instance), and to have an idea of whether or not a segment will require significant PE effort before actually working on it. The questionnaire was also a means to gauge if the perceptions of the

participants regarding PE effort for different segments coincided with the GTM scores for the segments in question.

Analysis and lessons learned from the second pilot:

The eye tracking data recorded in the PE sessions was analysed, taking into account the GTM scores and the processing speed for the PE tasks. This step was separate from the analysis conducted for the present research, and the results are outside the scope of this study (see O'Brien 2011).

For the most part, the typology used for the second pilot project was the same as the one used in the first pilot (the typology is explained in detail in section 3.4 of this chapter). However, after an initial verification of the data from the second pilot, it became clear that the typology would need to be complemented with additional categories. The master category Introduced Errors was added, and the subcategory Untranslated text was added to the master category Accuracy.

The main reason for adding a master category and a subcategory is that there were differences in the data from the first and the second pilot projects, and the typology used for the first pilot would not cover all the items that required classification in the second pilot. This might suggest that the typology would need to be revised every time new participants were used. However, it must be taken into account that the adjustments made in the typology for the second pilot project were small in comparison to the full number of categories, and were required in order to cover items that did not occur in the first pilot project (the different setting and the

higher word count in the second pilot project being the most likely reasons for this). It was anticipated that, in its updated form, the typology would not require further adjustments for the main PE experiment, and this proved to be true (see Chapter 4).

The master category Introduced Errors was added because, unlike what was observed in the first pilot project, the participants introduced a few errors that were not present in the raw MT output. The subcategory Untranslated text was added to the category Accuracy because, also unlike what was observed in the first pilot project, in the second pilot, in a couple of cases, parts of segments were left untranslated by the MT engine, and this had to be corrected by the participants.

Findings:

As previously mentioned, a comparison of the data from the first and the second pilots and the main PE project is provided in Chapter 4. However, a summary of the findings of the second pilot is provided here.

Items in the Language category accounted for the majority of the Essential Changes (57.26%). The combined total of Essential Changes Not Implemented and Introduced Errors was much lower than the total of Essential Changes, accounting for 19.94% of all the items recorded. The total of the Preferential Changes was also much lower than the total of Essential Changes, accounting for 16.5% of all the items recorded. The category with the highest number of Essential Changes Not Implemented was Language (86.66%), and the category with the highest number of Introduced Errors was Mistranslation (50%).

There did not seem to be an evident correlation between translation experience and PE performance in the second pilot. The Pearson Product Moment Correlation was calculated in order to determine the level of correlation between PE time and translation experience. The correlation value found, 0.64, indicated a weak positive correlation: the PE time would increase as the translation experience would increase. Nevertheless, the p value returned was 0.11, which did not suggest a significant relationship.

Correlations were calculated with GTM scores (reference translations were available for the segments used in the EYECON project). As mentioned, the 60 segments used in the PE task were classified as High, Medium or Low, according to their GTM scores. The segments with medium GTM scores were the ones with the highest values for all categories from the typology, except Essential Changes Not Implemented. Apart from this category, the segments with a low GTM score were the ones with the second highest value, and the segments with a high GTM score were the ones with the lowest number of changes. For the most part, this would corroborate the results found in the analysis of the eye tracking data from the EYECON project, in which segments with a medium GTM score were the ones with the longest fixation lengths and fixation counts, indicating a higher level of cognitive effort (O'Brien 2011, p. 17). In addition, this would also be consistent with the findings from Krings' research (Krings 2001): segments with medium quality would require more cognitive effort to be post-edited. We can speculate that this is linked to segment length, in the case of the different results found for segments with low and medium GTM scores, as in the corpus selected for this experiment, low

GTM segments tended to be shorter than medium GTM ones. Tatsumi (2010) also found evidence that sentence length might be one of the factors affecting cognitive effort. It is worth commenting that the data also suggest the hypothesis that segments with medium GTM scores might lead to more errors being introduced during post-editing. More investigation would be required to confirm these suppositions, but this would be outside the scope of the present research.

Limitations of the second pilot project:

The fact that the second pilot project took advantage of resources from the EYECON project imposed a few constraints: the translation environment to be used, the profile of the participants, and the need to have the participants working *in loco* in DCU, as their sessions were to be recorded with the eye tracking equipment.

The translation environment used was Alchemy Catalyst, since Alchemy was actively involved in the EYECON project. There was a small number of participants for the experiment (one for the pilot phase and six for the actual PE sessions), as they had to fulfil certain requirements, namely, being professional translators, being native speakers of French, having previous experience with Alchemy Catalyst, and having availability to spend a few hours in DCU for the individual PE sessions to be recorded. An attempt was made to reproduce real working conditions in the experiment. However, due to the practical limitations, there were differences that could not be avoided. In order to record the sessions using the eye tracker, it was necessary to ask the participants to work on

each segment only once, and not to go back to the segments they had already post-edited; in addition, they were supposed to continuously look at the screen for the duration of the task, so that the recording would not be interrupted. These constraints, combined with the fact that the participants were aware that all their actions on-screen were being recorded, may have affected the performance of some of them.

Under the conditions employed for the experiment, the translation environment used (Alchemy Catalyst) may also have influenced the results. Although the participants selected were familiar with it (with only one exception), they could not change the screen layout or the display settings, in order not to affect the recording, and also to ensure that all the participants worked with the same settings. Again, testing to what extent the translation environment had an impact on the results would be outside the scope of our research, but it is possible to speculate that, under normal working conditions, some of the participants might have felt more comfortable using different display settings and might have produced different results.

Nonetheless, both pilots helped us set some baseline expectations and to refine the methodology for the main experiment, which we describe in detail in the following sections.

3.3. Research design

3.3.1. Approach

The nature of the present research can be described as empirical, since one of the main methods of data collection in the methodology was the observation of the PE process in action. A mixed-methods approach was adopted, and the research design was characterised as a concurrent triangulation strategy, for the cross-validation of findings from qualitative and quantitative methods (Creswell 2003: 217). The different methods utilised in the research (questionnaires, screen recordings of the PE sessions, written post-edited output produced by the participants) were utilised with a view to minimising the possible bias associated with the use of a single method. The combination of these different options helped ensure the robustness of the results.

3.3.2. Variables

The following independent variables were used for the analysis: level of experience with translation (months or years); level of experience with PE (months or years); bias regarding MT and PE.

The dependent variables were: types and number of PE edits (classified according to the typology); total time spent on the task; keyboard and mouse usage; switches between keyboard and mouse; online terminology research (if performed by any given participant); revision of the segments at the end of the PE activity (if performed by any given participant).

3.3.3. How the main PE experiment was conducted

The data for the main PE experiment was collected from the following sources:

A series of machine-translated segments were post-edited by the participants in the target languages used in the present research. The segments used were obtained from a corpus of segments provided by Autodesk, a leading software company in the fields of 3D design for engineering, manufacturing and other fields. Autodesk, with which VistaTEC has collaborated in localisation projects over the years, became interested in contributing to the present study, due to their own research initiatives into post-editing productivity. In consultation with the researcher and her two supervisors, Autodesk therefore agreed to provide the corpus used for the main PE experiment, and also allowed the use of their online PE environment (as discussed in section 3.7).

A short survey was filled out by the participants during the recruiting phase, detailing their previous experience (in years/months) with PE and translation, their academic background and their keyboard layout (so that the environment for the PE task could be configured accordingly).

A short questionnaire was filled out by the participants before each PE session, with the objective of providing information regarding the translators' attitudes, beliefs and ideas on MT and PE. Quoting Selltiz et al. (1962: 246), Silverman (2011: 171) mentions that:

> "(...) it is always important to check first whether the respondent has any beliefs about the topic in question, otherwise the researcher may put words into his/her mouth".

Therefore, the participants were asked to fill out the questionnaire before the PE task so that it would be possible to gather information about their opinions on PE without the risk of having their answers influenced by the task to be performed. This was particularly important in the case of participants who had little or no previous experience with PE.

The short questionnaire (provided in Appendix A) was worded in a clear and concise way, in order to facilitate the data collection and the subsequent analysis. Three items were included in order to elicit the participants' opinions about the relevance, applicability and utility of PE. The participants were asked to choose ratings based on their opinions on each of the three items (referring to the quality of MT output, to the usefulness of MT for translators and to the level of interest of the participants in doing PE). The ratings ranged from one to five, with one being the lowest rating and five the highest. This approach was used in accordance with the definition of summated rating scales, as clarified by Spector (1991: 1-2):

> First, a scale must contain multiple items. (...) Second, each individual item must measure something that has an underlying, quantitative measurement continuum. In other words, it measures a property of something that can vary quantitatively rather than qualitatively. An attitude, for example, can vary from being very favorable to being very unfavorable. Third, each item has no "right" answer, which makes the summated rating scale different from a multiple-choice test. Thus summated rating scales cannot be used to test for knowledge or ability. Finally, each item in a scale is a statement, and respondents are asked to give ratings about each statement. This involves asking subjects to indicate which of the several response choices best reflects their response to the item.

The questionnaires were sent by e-mail, since the participants were in diverse geographical locations. (We discuss participant recruitment and profile in section 3.7.3.)

3.3.4. Corpus for the PE experiment

Word count:

A corpus with 2119 words from software documentation (therefore, from the IT domain) was provided by Autodesk for the main experiment. A total of 74 segments containing 1008 words in English were selected from this corpus to be post-edited by the participants.

The reason for selecting this specific number of words was that the sessions had a time limit of two hours, for logistical reasons. Based on the results found in the pilot projects, it was possible to extrapolate that it was feasible for the participants to post-edit this volume of words within the time limit determined. In the first pilot project, the average productivity for post-editing 350 words was 12.5 words/minute. By applying this value to the word count used in the main experiment, it would take 80.64 minutes to post-edit 1008 words. In the second pilot project, the average productivity for post-editing 804 words was 16.4 words/minute, which would yield 61.46 minutes to post-edit 1008 words. Both extrapolated times would fall below the maximum limit of 2 hours.

In addition, it was expected that the participants would not be willing to process a higher word count within the constraints of this experiment. They would likely be busy with their own academic and/or professional commitments and, therefore, would not be able to work on a more time-

consuming task, with a higher word count. Nonetheless, the translators were paid for their participation in the PE experiment.

It is important to mention that the participants were informed in advance that, if all the segments had not been post-edited when the time limit was reached, they would be required to stop. While it was predicted that most participants would be able to post-edit all the segments, the aim was to reproduce real work conditions, in which time pressure and deadlines are daily realities. As such, it was necessary to register if some of the participants would be unable to finish all the segments. This information was then correlated with the other measurements recorded.

Selection of the segments to be post-edited:

Independent segments of text were selected for the PE task to be performed by the participants. While using naturally occurring and continuous text would be advantageous for providing a logical flow, again, the objective was to try to reproduce real work conditions as much as possible. In the researcher's own extensive experience in localisation, translators are very often asked to work on files consisting of independent, isolated segments, instead of continuous text. For this reason, the same type of material was chosen for the experiment. The segments selected are listed in Appendix B.

The segments were randomly selected, and it then was checked if they adhered to the following criterion: they should not contain Autodesk's

specific UI terminology, otherwise, this could be a problem for participants not familiar with the company's projects. This was also important because product glossaries were not provided to the participants, as this would be outside the scope of the experiment and might pose confidentiality issues in relation to Autodesk's glossaries. Segments that contained specific UI terminology were discarded and replaced by other randomly selected segments, which were also checked according to the same criterion (i.e. avoidance of UI terminology).

Domain of the segments:

The segments were from the IT domain. The reason for selecting this domain (for the pilot projects and for the main project) was that PE is becoming increasingly common for localisation projects in this field, more so than for other fields (as observed by Allen, 2003: 300, for instance). Moreover, this domain was of most interest to the industrial sponsor.

3.4. Typology for the classification of PE data

A typology was devised for the present research, in order to classify the data collected from the PE sessions. The typology was tested and refined in the pilot projects, before being used for the main PE project.

The following sections provide details about different typologies used to classify MT errors and PE edits. In addition, the two classifications (the LISA QA Model and the GALE Post-editing Guidelines) that were

combined in order to create the PE typology for the present research are presented, followed by the customisation of the LISA QA Model and the Gale PE Guidelines.

3.4.1. Typologies for the analysis of post-editing activity

At present, there is not an internationally adopted and recognised model for classifying changes implemented during post-editing, which we refer to from now on as 'PE changes' for brevity. Therefore, the first step for preparing the linguistic analysis was to prepare a typology to classify the different types of changes made by the participants of the two languages. The typology should be sufficiently broad in order to cover the main categories of changes made, thus providing a good understanding of the edits that the participants made to the raw MT output. On the other hand, if the typology were to include too many categories, this might make the analysis less clear, so a balance had to be achieved.

Pym (1992, pp. 282) suggests a binary and non-binary classification of translation errors. Binary errors involve opposing "a wrong answer to the right answer", while non-binary errors are those involving a selection "from a potential TT series of more than one viable term" (in other words, there could be more than one right option, in this case). This classification could probably be applied to the issues found in the raw MT output, as there would be instances which could be described as having either a "wrong answer" or a "right answer" (binary errors), such as agreement errors, for example, while there would be other cases in which there could be more than one "right answer" (non-binary errors), such as a

mistranslated term in the raw output that can be corrected by using different options of correct terms. However, it can be argued that since this classification focuses on terms, it would not be sufficiently broad for the purposes of the present research. A more detailed classification would be required, including not only PE changes related to terms, but also corrections related to word order, sentence structure, addition of missing parts of the text or removal of parts of the text that were not present in the source files, to mention a few examples. Therefore, classifying PE changes only as binary and non-binary would not be flexible enough to cover all the different PE corrections that the present research aims to identify and analyse. Also, it would not facilitate a broader view of the difficulties encountered by the post-editors so that, subsequently, alternatives can be suggested to minimise the PE effort and to improve PE performance.

Flanagan corroborates the idea that a more complex classification would be needed for PE changes (Flanagan 1994, p. 65)

> Defining the boundaries of errors in MT output is often difficult. Errors sometimes involve only single words, but more often involve phrases, discontinuous expressions, word order or relationships across sentence boundaries. Therefore, simply counting the number of wrong words in the translation is not meaningful.

She proposes a classification of errors found in MT output by suggesting several categories, such as Spelling, Capitalisation, Elision, Verb Inflection and others, totalling 19 items (ibid, p. 67). The categories in this classification were devised taking into account the most frequent errors identified in the machine translation of a Hewlett-Packard test suite, an extensive set of sentences in English, compiled by Flickinger et al.

(1987). The classification was tested by CompuServe as part of an evaluation of three MT systems (the author does not provide a detailed description of the evaluation performed by CompuServe). The categories refer to issues found in English-to-French MT. Flanagan's classification provides a reasonable level of detail. However, ultimately its purpose is to help improve MT systems by classifying the errors found and pinpointing corrections to be implemented in the MT system itself, while the objective of the present research is to analyse PE practices and behaviours followed by linguists when carrying out PE tasks. This is an important difference: although Flanagan proposes 19 categories, which would seem like a very detailed classification, they were not devised for the same objective as the focus of the present project, which makes some of the classifications unsuitable for its purposes. For instance, these classifications do not offer a clear distinction between issues related to style and terminology, which would be an important aspect to analyse in order to differentiate essential and preferential post-editing changes. Additionally, Flanagan's classification was used in conjunction with RBMT systems, whereas the intention here would be to use a typology that would not be linked to a specific type of MT engine. Not linking the typology to a single MT paradigm would help to future-proof it: since MT technologies are constantly evolving, hybrid systems are becoming more widely used, and new paradigms may be developed which greatly differ from the systems currently in use.

The classification proposed by Marrafa and Ribeiro (Marrafa and Ribeiro: 2001) includes many more detailed categories than the examples

previously mentioned. This classification was also proposed as a framework to evaluate MT quality. The typology is used to calculate errors that could occur in a sentence, as well as the actual errors identified, and these numbers are then used to generate a score. The typology is divided into two main categories, Lexicon (with five subcategories) and Syntax (comprising eight subcategories, which, in turn, also include their own subcategories). While it can be very valuable to work with such a detailed categorisation, the focus of this classification is also the improvement of MT systems, not the analysis of PE practices. It could be argued that both are linked, since PE deals with the raw output generated by an MT system. However, there is an important distinction between these two perspectives: classifications geared towards improving MT systems examine the raw MT output, the errors found and the possible ways of changing the system so that these errors do not reoccur; on the other hand, the goal of the present research is to analyse how post-editors deal with the MT output, so the typology should enable this. Additionally, taking into account that one of the objectives here is to analyse the data gathered from several participants from two different languages, it might prove unfeasible to deal with so many subcategories.

Krings (2001, pp. 264-267) also proposes a classification of MT errors, comprising 11 categories:

- Lexical: Part of speech recognition error
- Lexical: other
- Morphology: Word formation
- Morphology: Other

- Syntax: Word order
- Syntax: Other
- Stylistic usage norms
- Punctuation
- Textual coherence
- Textual pragmatics
- Literal transfer from ST.

Kring's classification was not followed because, again, some of these categories would not be specific enough for the purposes of the present research. For instance, the category Lexical: Part of speech recognition error could be applied to several different types of errors. As pointed out by Krings himself (p. 266):

> It should be expressly pointed out that this error typology is not to be understood as a general typology for the classification of MT errors. It is exclusively related to the linguistic materials used in the present study. In other machine translation corpora, errors could appear that might require another classification and/or additional categories.

Loffler-Laurian (1996, pp. 96-97) proposes a typology of MT errors comprising 10 categories, such as Vocabulary and terminology, Determiners, Verb tenses, Modality and Negation, among others. Her typology was devised based on the analysis of several post-edited texts that had been machine-translated by Systran from English into French (the author does not specify the total number of words analysed). As with Krings' typology, her categories were appropriate for the PE analysis that she carried out, but they do not fully correspond to the objectives of the present research. Instead, the idea would be to employ a classification encompassing essential and preferential changes, as well as essential

changes not implemented, and that could be considered flexible enough to be employed across different languages. In addition, Loffler-Laurian's typology was based on a specific RBMT system, whereas, as already stated, the present analysis is not to be dependent on one MT paradigm.

Therefore, after much deliberation, taking into account the advantages and disadvantages of the alternatives previously mentioned, the decision was made to develop a typology specifically for the purposes of the present research, combining some of the categories from the LISA QA Model (The Localization Industry Standards Association 2009) and some from the GALE Post-editing guidelines (Post Editing Guidelines For GALE Machine Translation Evaluation 2007). It is important to point out that, with the setting used in the research (i.e. analysing PE work done on machine-translated IT texts), it makes sense to use the LISA QA Model as one of the starting points for the typology, since this model is widely used in the localisation industry.

The GALE guidelines were originally devised by the National Institute of Standards and Technology – Information Access Division/Speech Group and Linguistic Data Consortium, in the United States, for the post-editing of machine-translated texts from Chinese and Arabic into English (Post Editing Guidelines For GALE Machine Translation Evaluation, 2007, p. 3). Rather than being an error typology, it is in fact a set of PE guidelines, with examples and explanations about how to proceed in relation to each of the types of issues that they describe (such as Phrasal Ordering, Parts of Speech and Capitalisation).

The categories that are covered by the GALE guidelines (presented here in summarised form, for brevity) are:

- Capturing the Meaning of the Reference Sentence:
  - Synonyms;
  - Acronyms;
  - Symbols (including the subcategory Emoticons);
  - Numbers (including the subcategory Roman Numerals);
  - Abbreviations;
  - Contractions;
  - Phrasal Ordering;
  - Parts of Speech (including the subcategories Verb Tense, Prepositions, Adverbs and Adjectives, Determiners and Pronouns);
  - Extra Information in MT Output;
  - Information Missing from MT Output.
- Making the MT Understandable:
  - Reference Ambiguity.
- Minimizing the Number of Edits:
  - Spelling;
  - Proper Names (including the subcategories Proper names in Arabic source data and Proper names in Chinese source data);
  - Systematic Duplicates (including the subcategory Arabic salutations);
  - Dates;
  - Decimal Points.
- Punctuation and Capitalisation:

    - Sentence Ending Punctuation;
    - Quotation Marks;
    - Change/Insert/Delete Punctuation;
    - Capitalisation (including the subcategories Inserting words and punctuation and Moving words in MT output).
- Odd but Understandable English.
- Speech Data:
    - Disfluencies (including the subcategories Filled Pauses, Discourse Markers, Repetitions and Repairs);
    - Special Markup (including the subcategories Unintelligible Speech, Partial words and Foreign speech).
- Web data.
- Completely Incoherent MT System Translations.

(ibid, pp. 5-19).

The LISA QA Model, as indicated by its name, was developed for assessing the quality of translation in the localisation domain, not for assessing post-editing work. However, its categories are sufficiently broad to cover the main changes that could be implemented in post-editing tasks, its definitions are clear and, additionally, this QA model is widely adopted in the localisation industry worldwide. According to the Localization Industry Standards Association (homepage, 2009), "Approximately 20% of all companies involved in localized product testing use the LISA QA Model in some way, making it the most widely-used QA metric in the localization industry today." The LISA QA Model is heavily used in the IT sector, and it is adopted by the industrial sponsor of the current research, which was

one of the motivations for choosing it as the starting point for our typology. Another advantage of this model is that it can be employed for different languages, and it is flexible enough to be customised with additional subcategories, if so desired (some localisation companies customise it to some degree, according to their individual needs).

The main categories of the LISA QA Model are: Mistranslation, Accuracy, Terminology, Language, Style, Country and Consistency. A summary of the definitions provided in the model for each of these categories is provided in Table 3.1:

Table 3.1: LISA QA Model categories and their explanations

LISA QA Model category	Explanation
Mistranslation	Incorrect translation of the source text
Accuracy	Missing or extra information in the translated output, not reflecting the source text (but not a mistranslation of the original meaning)
Terminology	Inadequate terminology/lexicon for the context
Language	Issues related to grammar, semantics, spelling and punctuation.
Style	Non-compliance with the project's style guide
Country	Incorrect country standards, such as currency and decimal separators
Consistency	Non-standardised terminology used in the text

The categories used in the LISA QA Model do not overlap, serving distinct purposes instead. "Terminology" refers to the use of appropriate terms to a given context, whereas "Consistency" refers to the consistent use of terms in a document, in a software program or in other contexts (for instance, if there is a command called "Management", the same translation should be consistently used for it throughout the text; if some of the occurrences are translated as "Gerenciamento" and some others as "Gestão", this is considered as an inconsistency). As pointed out by Dunne

(2009, p. 200): "Consistency minimizes the risk of confusing users. Moreover, in some cases, proper functionality may depend on consistent translation." "Accuracy" is used to indicate the accurate rendering of the information contained in the source text (in other words, the target text should not contain information that is not present in the source text, and no information from the source text should be missing from the target text).

As well as being assigned categories, the issues identified can also be assigned severity levels, such as minor and major. For QA purposes, in addition to classifying the types of errors found, it is important to indicate their level of severity, so that an overall assessment of the work done can be calculated. The application of severity levels can also be considered relevant for evaluating the quality of post-edited MT output: in this specific case, the highest level of quality would not always be the intended result (although it can be beneficial), and the focus is on conveying the correct message rather than improving the style of the text. If a weighting system such as the one included in the LISA QA Model were used in the classification of PE changes, this would add a degree of detail that is not necessary for the purposes of the present research. All the corrections made to the raw MT output would have to be weighted as preferential, minor, major or critical, according to the model (i.e. one classification for preferential changes and three classifications for essential changes). While this weighting system can be very useful for assessing the quality of a translated text, it was outside the scope of the present analysis, as the intention here was to analyse and classify the *types* of changes made by post-editors in order to find out more about the strategies adopted by them,

in a **descriptive** manner, rather than evaluating their work. Therefore, the classification of levels of severity was not included in the typology. The only exception to this was the severity level "Preferential". In the LISA QA Model, the label "Preferential" is used as one of the levels of severity. However, as previously explained, in the typology used in the present research, we opted not to include a classification of the degree of **severity of errors**. Instead, we used "Preferential" in our typology as the name of a major **category of corrections** (the use of this specific category is further explained in section 3.4.3). The usefulness of this approach was tested in the pilot projects, and it proved to be appropriate for the overall analysis. Although severity levels are outside the scope of the present research, it is important to mention that if the typology used here (or a simplified version of it) is adopted in the future by a localisation company (for assessing a sample of PE work, for instance), severity levels could be employed, depending on the type of assessment desired.

In order to provide an additional degree of detail, the categories of the LISA QA Model were complemented with subcategories based on the GALE Post-editing guidelines. Although the language pairs used to develop the GALE post-editing guidelines do not correspond to the languages included in the present research, the guidelines provide detailed information about several subcategories that were expected to be suited to the present analysis. The subcategories from the GALE guidelines (such as Adjectives, Capitalisation and Determiners from the Category “Language") were added with a view to rendering the classification more detailed.

The elaboration of the typology was carried out prior to the analysis of the data from the two pilot projects. Subsequently, during the analysis of the data from the pilot projects, the applicability of the typology was tested and the modifications implemented were reviewed.

Not all of the subcategories from the GALE guidelines were included in the typology; only those that were anticipated to be relevant for the pilot data sets and for the analysis of the PE task were included. One of the main reasons for not including all of the categories from the GALE guidelines was that some of them might be too specific for the language combinations for which the guidelines were prepared (Chinese/Arabic-English), and less relevant for the language combinations on which the present research concentrates (English-Romance languages). For instance, this would be the case for categories such as Proper Names in Arabic Source Data and Proper Names in Chinese Source Data. Additionally, some of the GALE classifications (such as Symbols) would be more useful and relevant for the specific types of texts for which they were prepared: "newswire, blogs and newsgroups, and broadcast news and conversation recordings" (ibid, p. 3). The classification "Emoticons", for instance, makes much more sense in the context of the GALE guidelines than in the scope of the present research. The validity of this decision was tested when analysing the data from the pilot projects, and it was confirmed.

The subcategories added to the LISA QA Model (based on the GALE classifications) are presented in Table 3.2.

Table 3.2: subcategories from GALE added under the category Accuracy

Subcategory	Explanation
Extra information in MT output	This category involves removing extra information that is present in the MT output, but not in the source text.
Information missing from MT output	This category involves adding information that is missing from the MT output, but is present in the source text.

Table 3.3: Subcategories from GALE added under the category Language

Subcategory	Explanation
Adjectives	This category involves the correction of adjectives that are grammatically or otherwise incorrect or inaccurate in the target text.
Adverbs	This category involves the correction of adverbs that are grammatically or otherwise incorrect or inaccurate in the target text.
Capitalisation	This category involves the correction of capitalisation that is grammatically or otherwise incorrect or inaccurate in the target text.
Determiners	This category involves the correction of determiners that are grammatically or otherwise incorrect or inaccurate in the target text.
Phrasal ordering	This category involves the correction of phrasal ordering that is grammatically or otherwise incorrect or inaccurate in the target text.
Prepositions	This category involves the correction of prepositions that are grammatically or otherwise incorrect or inaccurate in the target text.
Pronouns	This category involves the correction of pronouns that are grammatically or otherwise incorrect or inaccurate in the target text.
Proper names	This category involves the correction of proper names that are grammatically or otherwise incorrect or inaccurate in the target text.
Punctuation	This category involves the correction of punctuation that is grammatically or otherwise incorrect or inaccurate in the target text.

Subcategory	Explanation
Spelling	This category involves the correction of spelling that is grammatically or otherwise incorrect or inaccurate in the target text.
Verb tense	This category involves the correction of verb tenses that are grammatically or otherwise incorrect or inaccurate in the target text.

Table 3.4: subcategories from GALE added under the category Country

Subcategory	Explanation
Decimal points	This category involves the correction of decimal points that are grammatically or otherwise incorrect or inaccurate in the target text.
Quotation marks	This category involves the correction of quotation marks that are grammatically or otherwise incorrect or inaccurate in the target text.

It is important to mention that six categories that **were not** part of the LISA QA Model nor of the GALE guidelines were also added: the master category Introduced Errors, the main category Format, and the subcategories "Gender" and "Number" (added under the main category Language), Untranslated text (added to the category Accuracy) and Date/time format (added to the category Country Standards). It was anticipated that many corrections implemented by the post-editors would be related to gender or number agreement, taking into account that, in Romance languages, words are normally inflected for gender and number, and this can cause errors in the MT output. It was also anticipated that there would be corrections related to formatting issues (since IT texts may include formatting tags). The LISA QA Model and the GALE Guidelines did not include specific subcategories for Number Agreement, Gender

Agreement and Format, so they were added to the typology. This choice of categories was tested in the pilot projects, and it proved to be justified.

The master category Introduced Errors was added so that it would also be possible to classify and quantify errors introduced by the participants that were not originally present in the raw MT output. It might be argued that this is different from the other categories, which are used in relation to corrections made. However, if it is considered that the categories are used to classify **PE changes**, the master category Introduced Errors is also covered by the same definition as the others. An error not present before, but introduced by the post-editor can be considered as a **change** made to the output.

The subcategory Untranslated Text was added to the category Accuracy to cover any untranslated items in the raw MT output that were corrected by the participants. The assumptions about these subcategories were then tested in the pilot projects, and they proved useful for the analysis.

Table 3.5: Categories and subcategories added to the typology that were not present in the LISA QA Model nor in the GALE Guidelines

Category or subcategory	Explanation
Introduced errors	Applicable when any errors (such as grammatical, semantic or related to formatting, for instance) that were not present in the raw MT output are introduced by the post-editors.
Format	This category involves the correction of tags, characters etc. that are grammatically or otherwise incorrect or inaccurate in the target text.
Gender	This category involves the correction of gender agreement that is grammatically or otherwise incorrect or inaccurate in the target text.
Number	This category involves the correction of number agreement that is grammatically or otherwise incorrect or inaccurate in the target text .
Untranslated text	This category involves the correction of items that are left untranslated by the MT engine in the raw MT output.
Date/time format	This category involves the correction of the date and/or time format in the target text if it is incorrect for the target language.

Another change introduced was the renaming of the category "Terminology" from the LISA QA Model to "Lexical Choice" in the typology. This was done to more accurately describe the PE changes that this category would cater for in the typology (i.e. changes related to the lexicon, in a broader sense, and not only terminology, which could be understood as project-specific terminology, for instance).

It was expected that the same typology could be applied not only to essential changes, but also to preferential changes, as well as essential changes not implemented and Introduced errors. Therefore, the typology includes four master categories: Essential Changes, Preferential Changes, Essential Changes Not Implemented and Introduced Errors. Under each of them, there is the same set of subcategories. For example, Language -

Gender may be a subcategory under any of our four master categories.

The complete typology can be seen in Table 3.6.

3.4.2. Typology for the linguistic analysis of the data

Table 3.6 presents the typology used to classify the PE data.

Table 3.6: Typology for classifying post-editing changes

Master categories			
Essential changes	**Preferential changes**	**Essential changes not implemented**	**Introduced errors**
Subclassification for the four master categories			
Main categories		**Subcategories** (if applicable)	
Accuracy (completeness)		Extra information in MT output Information missing from MT output Untranslated text	
Consistency		N/A	
Country		Decimal points Quotation marks Currency symbol Date/time format	
Format		N/A	
Language		Adjectives Adverbs Capitalisation Conjunctions Determiners Gender Nouns Number Phrasal ordering Prepositions Pronouns Punctuation Spelling Verb tense	
Mistranslation		N/A	
Style		N/A	
Lexical Choice		N/A	

Colour coding:
Black - categories from the LISA QA Model
Blue - categories from the GALE PE Guidelines
Magenta - categories devised by the researcher

It was assumed that further subcategories would not be needed, such as specifying if determiners are definite or indefinite, or if gender agreement issues referred to the masculine or feminine forms, as the subcategories added to the main categories would provide the necessary level of detail without making the typology excessively complex to employ.

3.4.3. Observations regarding the use of the typology

In this section, further details are provided about the use of the typology devised for the analysis of the data. Firstly, for the purposes of the present analysis, it is necessary to establish a working definition for PE changes. Elliott et al. (2004, p. 66) propose the following definition for errors in the MT output:

> (...) a unit of language that surprises the reader because its usage does not seem natural in the context in which it appears.

This definition is used as a starting point to formulate our definition of a PE change. Since an error in the MT output is considered as a unit of language, a PE change is also defined here as a unit of language, which may comprise a single word or two or more words. This definition is used to quantify the PE changes implemented by the post-editors. For further clarification, Table 3.7 provides examples of changes involving one word and several words:

Table 3.7: Examples of changes involving one word and more than one word

Source text	Raw MT output	Post-edited text	Comments
The installation media **contain** all the necessary packages and meta information that is necessary to install a &productname;.	A mídia de instalação **contêm** todos os pacotes necessários e metainformações que é necessária para instalar um.	A mídia de instalação **contém** todos os pacotes e metainformações necessários para instalar um &productname;.	Language - Number: the verb was changed from the third person plural ("contêm") to the third person singular ("contém") to correct the number agreement with the subject "A mídia de instalação" (which is in the singular). **This change involved only one word.** (For this example, only the change in question was highlighted, but not the other PE changes made.)
Networked Installation	**Réseau Installation**	**Installation réseau**	Language - Phrasal ordering The segment was changed from "Réseu Installation" to "Installation réseau" to correct the phrasal ordering. **This change involved two words.**

In some cases, a specific issue in the raw machine translation output may be corrected in a number of different ways in the post-edited text, depending on the judgement of each post-editor, and different solutions may be equally acceptable. In such cases, different changes adopted for the same issue by different post-editors may entail different classifications according to the typology, depending on the items changed. Table 3.8 provides an example of this.

Table 3.8: Examples of different ways of post-editing the same segment from the raw MT output

Source text	Raw MT output	Post-edited text 1	Post-edited text 2	Post-edited text 3
In some cases this happens because the computer is connected to the Internet.	Dans certains cas **produire cela** parce que l'ordinateur est connecté à Internet.	Dans certains cas**, ceci survient** parce que l'ordinateur est connecté à Internet.	Dans certains cas, **cela se produit** parce que l'ordinateur est connecté à Internet.	Dans certains cas, **cela arrive** parce que l'ordinateur est connecté à Internet.

In the previous example, the text "this happens" was mistranslated as "produire cela" in the raw MT output. The three post-editors corrected the mistranslation in different ways in order to convey the same meaning as the source text: "ceci survient", "cela se produit" and "cela arrive". The three different options were correct. This illustrates the different ways in which a segment from the raw MT output can be post-edited.

A specific change may require more than one classification; for instance, a term may be changed in order to have the correct number and gender in the output, agreeing with other items in a sentence. It was necessary to determine whether this should be counted as one or two

changes. In his analysis of PE, Krings (2001, p. 266) adopted the following strategy to classify overlapping MT issues:

> Several machine translation errors can overlap. Insofar as such errors can be assigned to different categories, each category assignment is counted as a discrete error.

While the focus here is on the corrections made by the post-editors, not the MT errors, Krings's solution was adopted for the present research, so that all items would be included in the classification. In the case of overlapping categories, each of them was separately classified and counted. For example:

Table 3.9: Example of overlapping PE corrections

Source text	Raw MT output	Post-edited text	Classification
All combinations of media for booting and installing may be used.	Todas as combinações de mídia para inicializar e instalar podem ser **utilizado**.	Todas as combinações de mídia para inicialização e instalação podem ser **utilizadas**.	Language - number: changed "utilizado" (singular) to "utilizada**s**" (plural) for number agreement with "combinações" Language - gender: changed "utilizad**o**" (masculine) to "utilizad**a**s" (feminine) for gender agreement with "combinações".

(This example focuses only on the overlapping corrections, which are specifically highlighted, and the other corrections made in the segment are not included).

3.4.4. Observations regarding the master categories of the typology

As previously mentioned, the typology includes four master categories: Essential Changes, Preferential Changes, Essential Changes Not Implemented and Introduced Errors. This section provides further details about each of them.

In the present analysis, a change is considered as essential when, **if the change 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, or it contains extra information that is not present in the source text).

Conversely, a change 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 change in question was not implemented**. In order to differentiate essential and preferential changes, these two definitions were strictly followed.

As well as accounting for the corrections made, it was also important to keep track of any essential changes not implemented by the participants. Issues in the raw MT output that were not corrected by the participants were identified during the analysis, when applicable, and their total number was also included in the tables for the participants.

A change is classified here as an essential change not implemented when, **due to the fact that it was *not* implemented**, the sentence (or part of it) is either:

c) Grammatically incorrect (i.e. it obviously breaches a grammatical rule specified in accepted grammar books), or

d) 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, or it contains extra information that is not present in the source text).

When an essential correction was not implemented by a given participant, it was counted and indicated in the corresponding table in the column "Essential changes not made". If several essential changes in one sentence were not implemented, they were also all counted as discrete occurrences.

Finally, a PE change is considered as an introduced error if:

a) **The error was *not* present** in the raw MT output, and **it was introduced by the post-editor** while editing a sentence;

b) **Because of it, the sentence (or part of it) is grammatically incorrect and/or inaccurate**.

In short, the category Introduced Errors caters for errors introduced by the post-editors (as opposed to errors that were present in the raw MT output). Examples might include (but are not limited to) typos and misspellings.

The PE changes were classified by the researcher after they had been made by the participants during their individual PE sessions. As

previously mentioned, sometimes PE issues can be corrected in different ways, and a PE change may involve one or more words; for this reason, in some cases, the results found for the present analysis (presented in Appendix F) contain different sums for essential changes and essential changes not implemented.

Due to the complexity of the classification with the typology, and also due to the fact that applying it was a very time-consuming process, it was not possible to include participants who could act as validators for the classification. Potential validators would need to have a similar profile to the researcher (i.e. extensive experience with translation, MT and PE and strong familiarity with the LISA QA Model), as well as sufficient availability to perform a task that would require 10 days or more (including some time to get acquainted with the typology and how to use it), and, of course, willingness and interest to work on such a lengthy and complex task. The number of language professionals for both target languages who could fulfil all of these requirements was more limited than the number of potential participants for the PE experiments (which did not entail the same requirements). In addition, logistic and budgetary limitations precluded the hiring of such validators for the extended periods of time that would be necessary. On the other hand, the researcher met all the requirements, and since she devised the typology herself, she did not need training in order to use it.

It could be useful to have the input of validators, but it might be argued that several disagreements regarding the classifications were to be expected, particularly in dealing with such a high number of items to be

classified. For instance, in her study, Guerberof (2012: 153) reported that the reviewers disagreed on 26.93% of the segments, and she also mentioned other studies in which disagreements were reported (Carl et al. 2011 and García 2010, 2011).

It is also important to highlight that, unlike other research projects on MT or PE that make use of annotators, in the present research the segments were not classified as "correct vs. incorrect" or "understandable vs. unintelligible". The typology was used descriptively, to classify all the changes (essential or preferential) made by the post-editors, as well as the changes that they failed to implement, and errors that did not exist in the original and that they inadvertently introduced.

In order to ensure the precision of the classification, the researcher performed three verifications of all the items classified, with an interval of time of no less than a month between each verification pass to ensure adequate distancing and a "new awareness" (Horning and Becker 2006: 168). In addition, after the three verifications had been carried out, two language professionals with experience similar to the researcher's, and who were native speakers of French and Brazilian Portuguese, respectively, validated a sampling of the classifications done by the researcher with the typology. While the sampling was much more limited in scope than a validation of the full classification would have been (comprising approximately 5% of the segments), it was useful to confirm that two very experienced language professionals who had not taken part in any of the experiments agreed with the classifications selected by the researcher for different items, particularly in relation to the main categories

of the typology. (It must be added that, had the validation included the complete classification, it is possible that the two validators would have chosen different classifications in some cases, as previously mentioned.)

3.5. Definition of a good post-editor

Specifically for the first research question, it was necessary to have an operational definition of what the requirements are for being a good post-editor. Different definitions could be proposed, such as "producing a final text with the best possible quality" or "being able to perform a PE task strictly according to the guidelines". These two definitions could be considered suitable for specific situations (in the first case, if the text in question were to be published for a wide audience, for instance, and in the second case, if the text were targeted at a smaller audience and for informative purposes only). However, although valid, these definitions are related to specific scenarios only, and do not indicate the general qualities that a good post-editor should possess. Bearing this in mind, the following definition of a good post-editor was formulated: a good post-editor is able to fully adapt to different PE guidelines and constraints, producing a final text according to the level of PE requested for a specific job, within the timeframe required.

This definition takes into account the concept of "fitness for purpose". This concept is widely employed in different industries to delimit the applicability of products and services. In fact, useful explanations that

clarify the notion of fitness for purpose can be found in reference works from different sectors, such as the explanation provided by Mukherjee (2006, 33):

> The product or service should serve the purpose for which it is used. If the product is used efficiently, but it does not serve the purpose for which it is intended to be used, it is not supposed to be having good quality.

Clarke (1996, 67) further clarifies this concept:

> (...) fitness for purpose, i.e. simply asking that the product be capable of doing the job required of it, **no more and no less**. *(researcher's emphasis)*

Taking into account the explanations provided by the authors above and extrapolating this concept for the activity of PE, it would be possible to say that this reiterates the definition of a good post-editor proposed for the present research. In other words, a good post-editor would meet the requirements of the specific PE task at hand, making changes and corrections only according to the guidelines provided, and delivering a final text with the required level of quality in the time-frame specified.

Using the suggested definition of a good post-editor, it is possible to test the post-editors' ability to adhere to PE guidelines according to the level of quality expected of the end product (as required for specific audiences, projects, clients and use-case scenarios).

3.6. Outliers

When examining the texts post-edited by the participants, it was also necessary to take into account the possibility that some of them might qualify as outliers. As defined by Han et al. (2012: 544), "(...) an outlier is a data object that deviates significantly from the rest of the objects, as if it were generated by a different mechanism". Specifically in the scope of the present research, therefore, an outlier would be a post-editor who would finish the PE task significantly faster than the average time of other participants and, at the same time, would make corrections so as to produce a final text with a very high level of quality. Conversely, an outlier might also be a participant who would require a significantly longer period of time to complete the task, and would produce a final text with a level of quality much lower than could be accepted. Thus, an outlier may not necessarily conform to the operational definition of a good post-editor (as previously outlined), since they might make a higher or lower number of corrections than the actual number and scope indicated by the task's guidelines, and they might not be able to complete the task within the required time limit.

3.7. PE environment

The environment used for the PE task was Autodesk's Post-editing Workbench. The Workbench was developed by Autodesk for previous

internal projects, but it is also very suitable for the experiment. It consists of an interface presented as a webpage (see Figure 3.1).

All the participants received the same instructions (provided in Appendices C and D) on how to carry out the task using this environment. However, it is important to mention that the Workbench is very easy to use and it does not require previous knowledge or experience. This helped ensure that all the participants were levelled out (i.e. the level of experience with the environment was not a variable for the purposes of the experiment). In addition, this avoided the need to introduce a new tool that might require installation and prior training, as could be the case for a translation or PE tool. The workbench is not a tool: it is a website with fields for editing the segments.

Using a Web-based environment was also helpful for the setup devised, as the Workbench was opened in a browser on the researcher's computer, and then it was remotely accessed by each participant. The researcher could still see all the PE actions performed on-screen by the participants and record the PE session.

The Workbench presents the segments in a format that is similar to that used by SDL Trados: each segment in English is presented in a separate field, in blue, and is **locked** (i.e. cannot be edited), and the corresponding raw MT segment is presented in the field beneath it, in yellow, **unlocked** for editing.

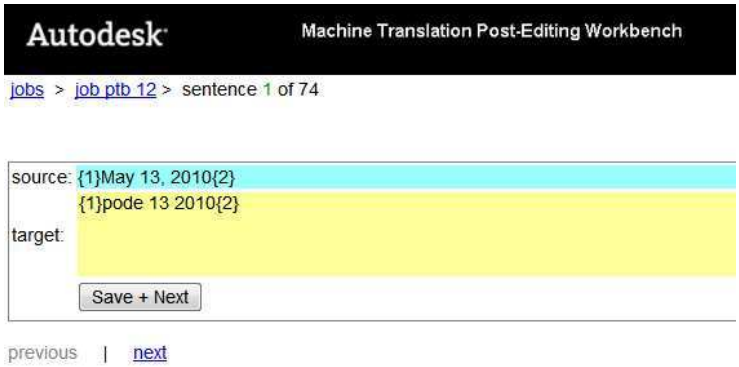


Figure 3.1: Screenshot of the Post-Editing Workbench

In the segment displayed in Figure 3.1, "{1}" and "{2}" are representations of formatting tags. Post-editors need to take them into account during PE tasks, since there may be missing or extra tags in the raw MT output, and tags may not be in the right order.

3.7.1. Preparation of the segments for the main experiment

The selected segments were machine-translated from English into Brazilian Portuguese and from English into French in advance by Autodesk, using their own encoded terminology. The MT system used was Moses, which was trained with previous projects from Autodesk for the language pairs used in the main PE experiment. Autodesk informed us that there was not a significant difference between the total word counts of the corpora used to train the MT engine for each language pair.

3.7.2. Configuration

Keylogging:

A software program called InputLog (www.inputlog.net) was used to record all the keyboard and mouse actions performed by each participant while post-editing the segments. InputLog was developed by Luuk van Waes as a keylogger that records mouse and keyboard actions, including the time spent on them and any switches between them. If, instead of a keylogger, a translation or PE environment with screen and/or keyboard recording capabilities had been used, it might have been unfamiliar for some the participants. As a result, training and additional instructions might be necessary. The use of a keylogger avoided this issue. The keylogger is not intrusive, as it works in the background, regardless of the PE environment used. Therefore, while the participants used the Post-Editing Workbench, the keylogger recorded their actions and did not interfere with their work.

Screen recording:

In addition to keylogging, screen recording software, Camtasia Studio, was used to record the screen. The PE sessions were thus recorded in real time, and this did not interfere with the participants' work in any way either. Camtasia Studio and InputLog were not perceived to slow down the computer used for hosting the PE sessions.

Connection:

LogMeIn (https://secure.logmein.com) was used as the method for remote connection. The researcher registered an e-mail account and password in LogMeIn's website and downloaded the client version of the software to the computer to be used for the experiment. Once the computer was registered in the researcher's account, it could be remotely accessed by the participants by means of a straightforward process. The researcher sent the participants LogMeIn's URL, the username and password. These were the only items required for the participants to remotely connect to the researcher's computer in Dublin, after LogMeIn was activated in it. Once connected, the participants were able to view the remote computer in a resizable window in their own computer screens, and in that window they had control of the remote computer as if they were using it in person. At the same time, the researcher still had control of the remote computer, and it was possible to observe all the actions performed by the participants on the screen.

Preparation of the computer for the experiment:

The researcher asked each participant in advance what the keyboard layout and the language settings were that he or she normally used, so that the remote computer could be configured with the same settings. Before starting each session, a quick test was done with the

participants to make sure that the keyboard layouts of the participant's and the researcher's computers did indeed match.

The researcher's computer was prepared prior to each PE session as follows: the Post-Editing Workbench was opened in a browser window, ready for the post-editor to start working; the keyboard layout settings were changed according to the remote participant's settings (as informed in advance by each participant); LogMeIn was activated in order to allow the remote connection; Camtasia Studio and InputLog were activated.

During the entire PE session, the researcher remained connected on a second computer, to take notes about the session.

Time limit:

A time restriction of two hours was specified for the participants. This was due to logistical reasons, such as the budget allocated to pay the post-editors, as well as the availability and willingness of potential participants to work on the project.

Additionally, the time restriction ensured that the maximum time for the experiment would be a constant. The participants were informed in advance that, in case the time limit was reached, their PE sessions would need to be ended, but they would be able to work at their own pace, and it would not be a problem if they were not able to finish post-editing all segments by the end of the two hours. This allowed the researcher to

measure and compare the time required for each participant to perform the PE task.

The participants were allowed to go back at any moment to segments they had already post-edited, if they wanted to clarify any doubts or to make further edits. No restrictions were imposed regarding this, also due to the objective of trying to reproduce real working conditions as much as possible.

3.7.3. Participants

It was very important to choose the appropriate participants for the experiment, since the data would be derived from the output produced by them. The participants had to have the right qualities according to the scope of the present research project. To reflect the translation community at large, and to address one of our research questions, they had to have different levels of professional translation experience, as well as different levels of PE experience.

Number of participants:

Arnold et al. (1993) suggest that the higher the number of participants, the more reliable the results, and Hatch points out that "(...) the fewer the number of participants, the more important it is to include multiple data sources" to achieve a balance in the study (Hatch 2002: 50). Together with logistic and budgetary limitations, these considerations were

taken into account for designing the project, and the goal was to include a total of 20 participants, 10 per target language, who would embody a range of levels of experience with translation and PE.

Selection criteria:

The selection of the participants was guided by the following criteria:

Ten participants had to be native speakers of Brazilian Portuguese, and the other ten had to be native speakers of French. The first reason for selecting these two languages was that both Brazilian Portuguese and French are among the languages with the highest demand in the localisation industry, due to market size. This is confirmed by data gathered by Common Sense Advisory, which classifies both French and Portuguese as among the ten languages that "account for 76% per cent of the people on the web" (DePalma et al., 2007: 9). Also, Brazilian Portuguese is the native language of the researcher, and French is one of the languages with which she works.

The participants for each language had to be either professional translators or translation students. The reasoning behind this was to have participants with a range of experience (in number of years/months), including those who were still beginners in the translation profession (the students).

The translators selected for the experiment had to have different levels of PE experience (in number of years/months). Some of the participants had to have previous experience with PE, and some others had to have no previous experience with it. The reasoning for having

participants with varied levels of experience was linked to the very nature of the first research question: it would help investigate to what extent the post-editors' previous translation experience might influence their PE performance.

Selection process:

VistaTEC compiled a list of French and Brazilian translators from its database of contractors. An initial contact by e-mail was made by the researcher with several possible candidates (bearing in mind the number of participants envisaged for the project) to ask them if they would be interested in taking part in a research project. The initial e-mail contained summarised details about the experiment and an invitation to take part in it. Some of the potential participants contacted in the first phase were not available or did not express interest in taking part in the experiment; for this reason, additional participants from the list were contacted by the researcher in order to fulfil the number of post-editors required for the main PE project.

Those who expressed interest in taking part in the research after the initial contact were sent a survey inquiring about their level of experience with translation and with PE (in years/months), whether they had an academic background in translation (although this was not a requirement in order to take part in the project), whether they had fast Internet access (this was a requirement so that they could connect remotely to the researcher's computer to perform the PE task), and

whether they were native speakers of one of the two target languages included in the main PE project. The participants were also asked to indicate the keyboard layout and the language settings that they used on their own computers. This was required for the setup of the experiment. Not all the potential participants were familiar with the concepts of keyboard layouts and language settings (presumably due to using the default settings on their computers, and not having had to check them or change them), so the researcher provided additional details about these items as required, in order to help the participants identify the necessary settings.

The translators were also informed that they would be paid a standard hourly rate for their participation, and that all their identification details would be kept anonymous.

After the surveys were returned by the potential participants, the researcher followed up with each of them individually to set dates and times for their PE sessions, according to their availability. (They all signed consent forms and read the Ethics Form, which are available in Appendix E.)

One of the French participants (FR03) did not manage to complete the PE task due to connectivity issues. In addition, the file with the session recorded for another participant (FR06) was corrupted. Therefore, it was not possible to include the data from these two French participants in the analysis.

3.7.4. Instructions

Once the individual PE sessions were scheduled, the researcher sent instructions about the PE task to the participants. The instructions for connection were in English, and they included details on the logon process to the remote computer using Logmein. Additionally, the researcher sent PE instructions to the participants. The instructions were in English, and two different versions were prepared: both contained the same set of PE instructions, but one had examples in French, and the other one had examples in Brazilian Portuguese. The versions sent corresponded to the native language of the participants. (The instructions are available in Appendices C and D.)

The instructions explained the task to be carried out (i.e. post-editing the segments in the environment used in the main PE project), the meaning of PE, the level of quality expected, the types of corrections that could be considered valid and those that could be considered preferential. The instructions were concise and, insofar as possible, reflected the instructions that are sent to translators for professional PE projects (according to the researcher's own professional experience with PE and our literature review on this topic). The instructions specified that the participants should correct grammatical errors or issues that would make the text difficult to understand, but they should not make preferential corrections, such as replacing correct words with synonyms. They also explained that PE is not the same as revision.

The instructions specified that the level of quality expected for the post-edited segments was intermediate (for instance, as it would be the case for documents to be distributed internally, but not for publication and/or distribution to a wide audience), as opposed to full PE. This followed the concept of minimal PE, as discussed by Allen (2003): an intermediate level between rapid PE, which entails the lowest level of corrections and is indicated for texts used for gisting purposes only, and full PE, which aims to produce post-edited texts with the same quality as human translation.

The reason for selecting an intermediate level of PE for the experiment was that fast PE may present additional difficulties to translators who are not used to this type of work. As described by Allen, fast PE should be used only to remove the most serious errors. Therefore, some of the issues identified (such as agreement errors, for instance) would not be corrected if they did not compromise the general understanding of the text or the rendering of the meaning intended in the source text. For some translators, it might take additional time and practice to adapt to this level of PE, since leaving issues uncorrected is not part of standard practices that they would have been used to following for translation projects. On the other hand, full PE aims to produce texts with a degree of quality equivalent to human translation; therefore, it would be much more time-consuming, which might potentially require more time than the maximum allocated per session for the main experiment. Moreover, opting for full PE (which would require the implementation of all the corrections deemed necessary, without restrictions, to achieve the

maximum level of quality) might defeat one the objectives of the main project, which was to test our proposed definition of a 'good post-editor'. This definition includes a stipulation that the post-editor can adhere to specific guidelines.

The participants were informed in advance that, while performing the task, if necessary, they would be free to refer to any online sources they might consider useful, if they would like to clarify any doubts on terminology. Terminology searches were allowed as long as they were performed using the remote computer's screen, so that they could be recorded as part of each session. The keylogger and Camtasia recorded any visits made by the participants to Web sites to check the meaning of words or to look for other information, and the researcher also observed this onscreen. The participants were asked not to consult hard-copy resources, since the researcher would not be able to record this remotely.

The instructions to be provided to the participants deliberately did not include guidelines about revising the segments at the end. The reason for not including this was that it would be relevant to investigate whether revising the segments at the end of the task would be a strategy spontaneously chosen by the participants, and whether this choice would have any link with their level of professional experience. This was recorded as part of the strategies employed in each session, and then compared with the variables in the analysis.

3.8. Data analysis procedures

The items below were recorded for both target languages, in order to be compared with the participants' level of translation and PE experience and their personal bias regarding MT (as indicated by the information gathered in the questionnaire). The objective was to investigate whether the level of previous experience with translation might have an influence on PE performance (the scope of Research question 1), specifically examining:

a. Number of corrections made by each participant.

b. Types of corrections made by each participant (classified according to the typology).

c. Total time spent on the task.

e. Number of switches between keyboard and mouse and total time using each input method.

f. Number of online terminology searches conducted (if any).

g Types of items searched online (if any).

h Whether or not each participant revised the post-edited segments at the end.

The data gathered for Brazilian Portuguese and for French were compared with the objective of determining if similar post-editing strategies would be employed across languages of the same family (the scope of the research question 2).

The findings from the pilot projects were compared to the findings of the main project, to identify any possible links.

Reference human translations were not available for all the segments used in the main PE project, so scores such as BLEU or TER were not used.

3.9. Validity and trustworthiness of the findings

As much as possible, the ecological validity of the research was ensured by reproducing real work conditions for the participants. This was done by using a PE environment that resembles a commonly used CAT tool, allowing the participants to work at their own pace, though with a reasonable deadline, by providing PE instructions that were similar to the instructions provided in real projects, by allowing the participants to complete the task from their own home or office, by letting the participants decide whether they wanted to revise the post-edited segments at the end or not, and whether they wanted to search for terminology online or not.

As explained by Frey et al. (1991), there are different threats to validity. The "Researcher Personal Attribute Effect", which is of particular interest to the present research, can occur if the research task proposed is ambiguous and the participants therefore rely on the researcher to obtain indications on how to perform it, or if the participants feel the need to perform it according to characteristics that they perceive in the researcher. One of the ways of avoiding this threat is to adopt standard procedures and to provide the same environment for all participants. This was implemented in the present research by providing exactly the same instructions to the participants, by using the same setup and the same PE

environment for all the sessions. Further threats to the validity and the reliability of the procedures, as described by Frey et al., were avoided by always following the procedures in the same order (i.e. initial contact, survey, instructions, questionnaire and PE session), and by ensuring that the data from different participants were analysed using consistent procedures, and in the most accurate way.

3.10. Summary

This chapter began by providing detailed information about the pilot studies conducted prior to the main PE project. This was followed by a section covering the research design, which included details on the approach adopted, the variables, a description of how the study was conducted, the corpus used for the main PE experiment, the typology employed for the classification of the data, the definition of a good post-editor employed for the analysis, the PE environment used, the configuration, the selection of participants and the instructions provided. The final section discussed the data analysis procedures and the validity of the findings.

Chapter 4: Analysis

4.1. Overview of this chapter

This chapter presents the analysis of the data from the main PE project for Brazilian Portuguese and French. In order to be more succinct, the results are presented here in summarised format, followed by a discussion about them. The detailed data on which the summaries are based are presented in Appendix F, Presentation of the data.

This chapter begins with an investigation of comparisons of translation experience, PE experience, total PE time, number of changes implemented or not and introduced errors. The comparisons are presented both in tables and figures, including the results of the Pearson Product Moment Correlation. Next, this chapter presents the analysis of the results of the questionnaire answered by the participants, and how they compare with translation and PE experience. This is followed by comparisons with the results of the pilot projects. The analysis of the data related to keyboard and mouse usage follows, in which possible links between translation and PE experience, keyboard and mouse use and switches between keyboard and mouse are discussed. The next sections deal with productivity and quality, respectively. A discussion of the common trends and strategies observed between the two target languages is then presented. The chapter concludes with a summary of the results.

4.2. Comparisons between total PE time and translation and PE experience

In this section, first the values for each target language are presented separately, including the total time spent for the PE task and the translation and PE experience of the participants. The values are then presented with the two target languages combined. By presenting the data in this way, we get an appreciation of the PE performance for each individual target language, followed by an analysis that combines both target languages. This method gives us a larger group of participants from which to draw conclusions about experience and PE activity.

4.2.1. Comparisons between total PE time and translation and PE experience - French

Table 4.1 provides a summary of translation and PE experience and the total time taken for the PE task for each participant for French. In order to use the same unit for all the measurements, the values in years were converted to months. However, for convenience, both values (in years, where applicable, and in months) are presented in this and in all the subsequent tables in this section.

Table 4.1: Comparisons between total translation experience and total PE time - French

Participant	Translation experience (months)	PE experience (months)	Total PE time (minutes)
FR01	84 months (7 years)	36 months (3 years)	49
FR02	36 months (3 years)	0	75
FR04	0	0	105
FR05	48 months (4 years)	6 months	64
FR07	36 months (3 years)	3 months	122
FR08	144 months (12 years)	36 months (3 years)	65
FR09	216 months (18 years)	12 months (1 year)	120
FR10	252 months (21 years)	5 months	75
Arithmetic mean	102 (8.5 years)	12.25 (1.02 year)	84.375
Median	66 (5.5 years)	5.5 (0.45 year)	75
Standard deviation	92.28	15.14	27.58

The next figures show comparisons between translation experience and PE experience and the total time taken for the PE task.

Figure 4.1: Comparison between translation experience and total PE time - French

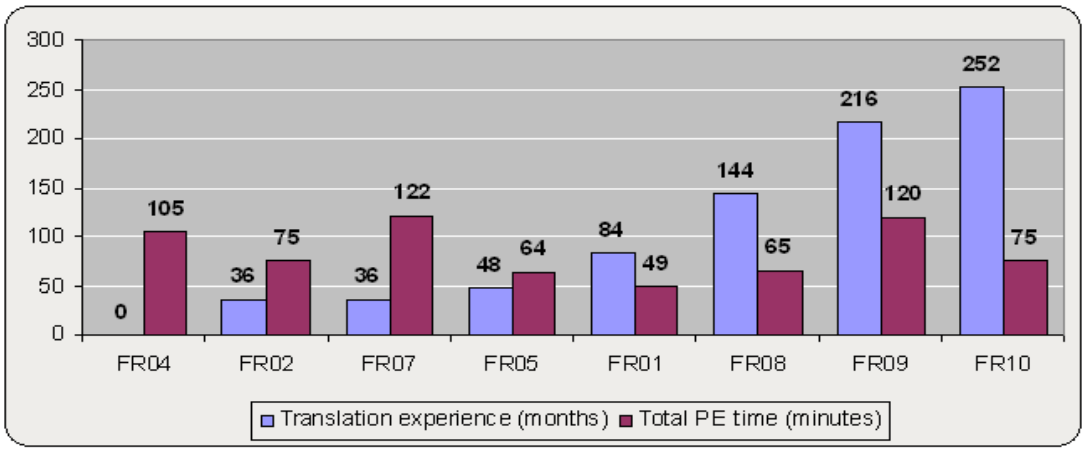


Figure 4.2: Comparison between PE experience and total PE time - French

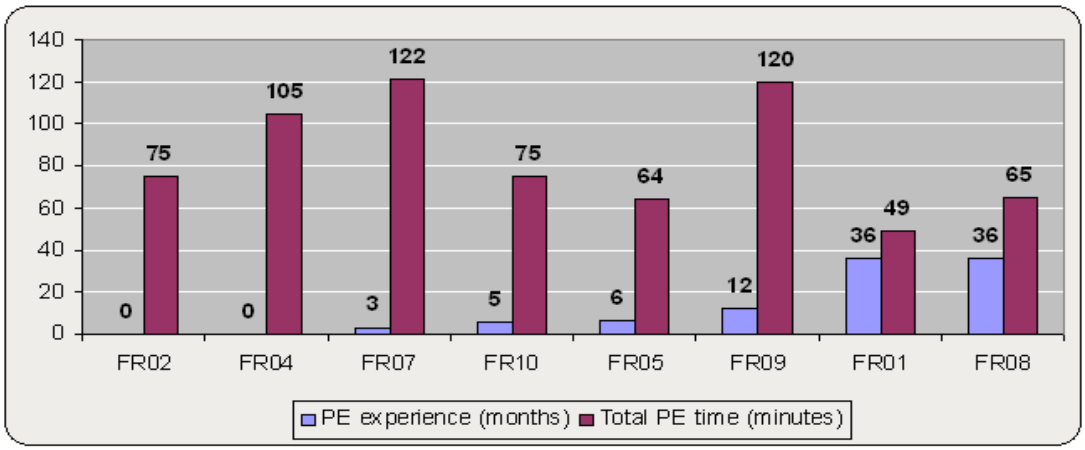
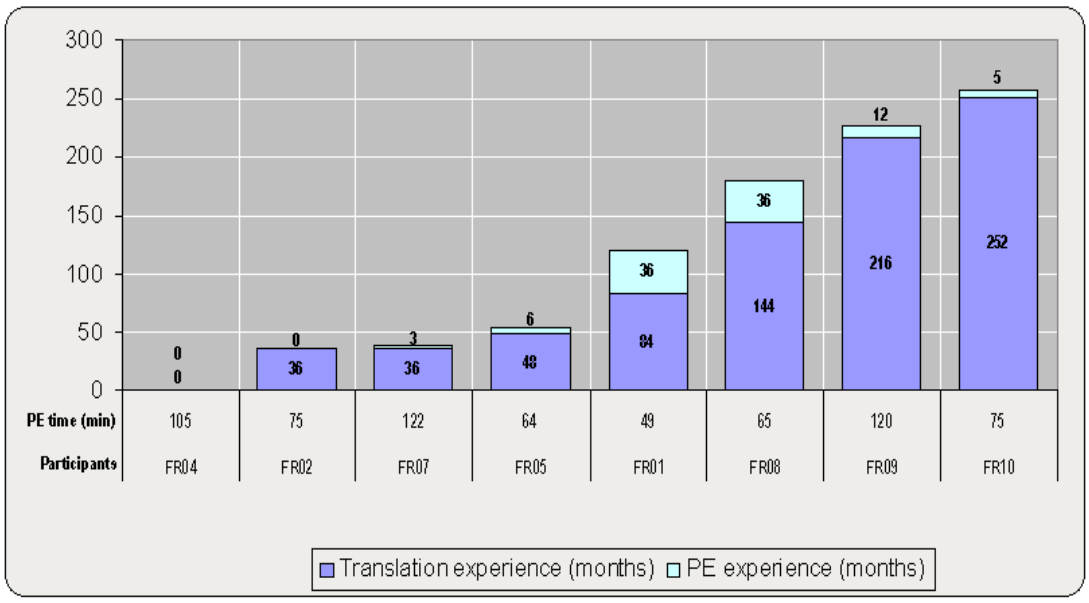


Figure 4.3: Comparison between translation and PE experience (combined) and PE time - French



The Pearson Product Moment Correlation was calculated, and Tables 4.2 and 4.3 display the results.

Table 4.2: Pearson Product Moment Correlation: translation experience and PE time - French

Statistic	Variable X: translation experience (months)	Variable Y: PE time (minutes)
R-value	-0.021	
Degrees of Freedom	6	
Number of Observations	8	

The correlation coefficient of -0.021 indicates that, based on the values presented here, there does not seem to be an obvious correlation between translation experience and PE time for French.

Table 4.3: Pearson Product Moment Correlation: PE experience and PE time - French

Statistic	Variable X: PE experience (months)	Variable Y: PE time (minutes)
R-value	-0.546	
Degrees of Freedom	6	
Number of Observations	8	

The correlation coefficient of -0.546 indicates a moderate negative correlation: the PE time would decrease as the level of PE experience increases.

4.2.2. Comparisons between total PE time and translation and PE experience - Brazilian Portuguese

Table 4.4 provides a summary of translation and PE experience and the total time taken for the PE task for each participant for Brazilian Portuguese.

Table 4.4: Comparisons between total PE time and translation and PE experience - Brazilian Portuguese

Participant	Translation experience (months)	PE experience (months)	Total PE time (minutes)
BR01	192 months (16 years)	12 months (1 year)	93
BR02	48 months (4 years)	12 months (1 year)	60
BR03	132 months (11 years)	0	95
BR04	240 months (20 years)	36 months (3 years)	61
BR05	156 months (13 years)	36 months (3 years)	53
BR06	192 months (16 years)	36 months (3 years)	65
BR07	72 months (6 years)	0	75
BR08	240 months (20 years)	24 months (2 years)	99
BR09	360 months (30 years)	0	46
BR10	60 months (5 years)	0	84
Arithmetic mean	169.2 (14.1 years)	15.6 (1.3 year)	73.1
Median	174 (14.5 years)	12 (1 year)	70
Standard deviation	97.23	16.04	18.82

The next figures show comparisons between translation experience and PE experience and the total time taken for the PE task.

Figure 4.4: Comparison between translation experience and total PE time - Brazilian Portuguese

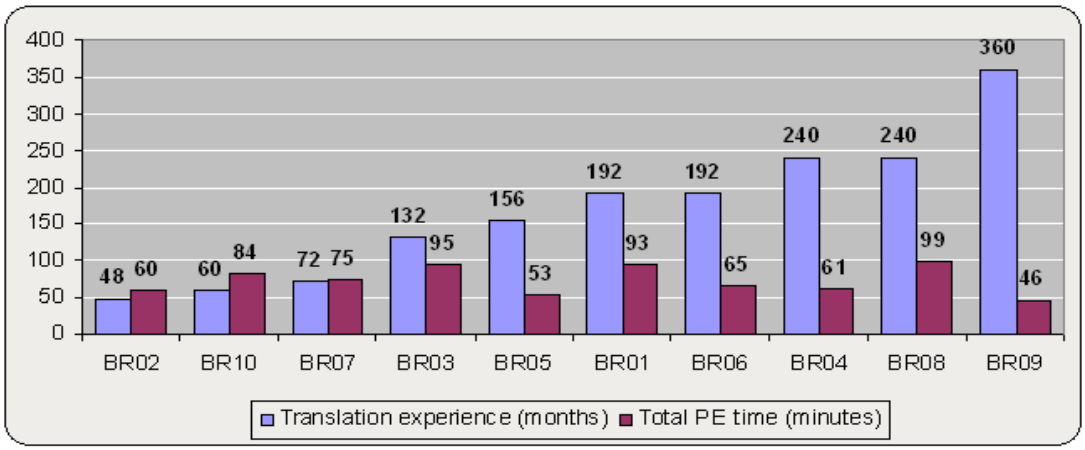


Figure 4.5: Comparison between PE experience and total PE time - Brazilian Portuguese

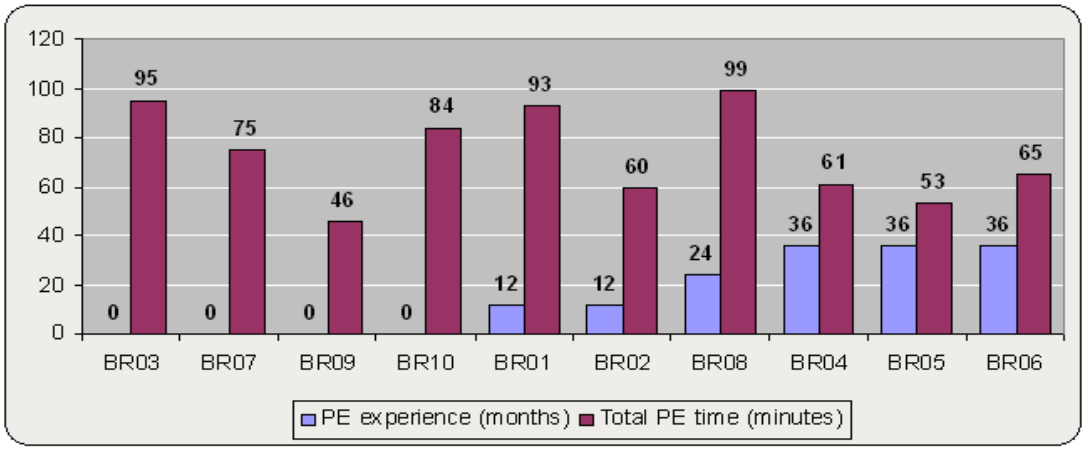
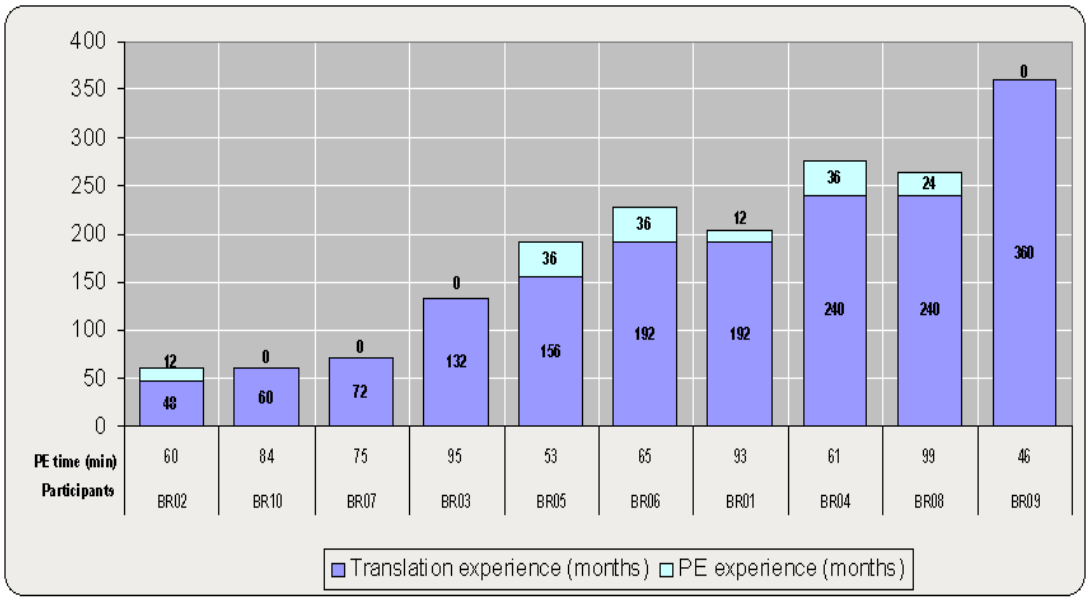


Figure 4.6: Comparison between translation and PE experience (combined) and PE time - Brazilian Portuguese



Similarly to what was done for French, the Pearson Product Moment Correlation was calculated, and Tables 4.5 and 4.6 display the results.

Table 4.5: Pearson Product Moment Correlation: translation experience and PE time - Brazilian Portuguese

Statistic	Variable X: translation experience (months)	Variable Y: PE time (minutes)
R-value	-0.258	
Degrees of Freedom	8	
Number of Observations	10	

The correlation coefficient found, -0.258, indicates a weak negative correlation: the PE time would decrease as the level of translation experience increases.

Table 4.6: Pearson Product Moment Correlation: PE experience and PE time - Brazilian Portuguese

Statistic	Variable X: PE experience (months)	Variable Y: PE time (minutes)
R-value	-0.274	
Degrees of Freedom	8	
Number of Observations	10	

The correlation coefficient of -0.274 indicates a weak negative correlation: the PE time would decrease as the level of PE experience increases.

4.2.3. Comparisons between total PE time and translation experience - French and Brazilian Portuguese combined

Table 4.7 provides the values for translation and PE experience and the total time taken for the PE task for the participants of the two target languages.

Table 4.7: Comparisons between total PE time and translation experience - French and Brazilian Portuguese combined

Participant	Translation experience (months)	PE experience (months)	Total PE time (minutes)
FR01	84 months (7 years)	36 months (3 years)	49
FR02	36 months (3 years)	0	75
FR04	0	0	105
FR05	48 months (4 years)	6 months	64
FR07	36 months (3 years)	3 months	122
FR08	144 months (12 years)	36 months (3 years)	65
FR09	216 months (18 years)	12 months (1 year)	120
FR10	252 months (21 years)	5 months	75
BR01	192 months (16 years)	12 months (1 year)	93
BR02	48 months (4 years)	12 months (1 year)	60
BR03	132 months (11 years)	0	95
BR04	240 months (20 years)	36 months (3 years)	61
BR05	156 months (13 years)	36 months (3 years)	53
BR06	192 months (16 years)	36 months (3 years)	65
BR07	72 months (6 years)	0	75
BR08	240 months (20 years)	24 months (2 years)	99
BR09	360 months (30 years)	0	46
BR10	60 months (5 years)	0	84
Arithmetic mean	**139.33 (11.61 years)**	**14.11 (1.17 year)**	**78.11**
Median	**138 (11.5 years)**	**9 (0.75 year)**	**75**
Standard deviation	**98.45**	**15.28**	**23.11**

Figures 4.7 and 4.8 show comparisons between translation experience and PE experience and the total time taken for the PE task, with the values for both target languages combined.

Figure 4.7: Comparison between translation experience and total PE time - French and Brazilian Portuguese combined

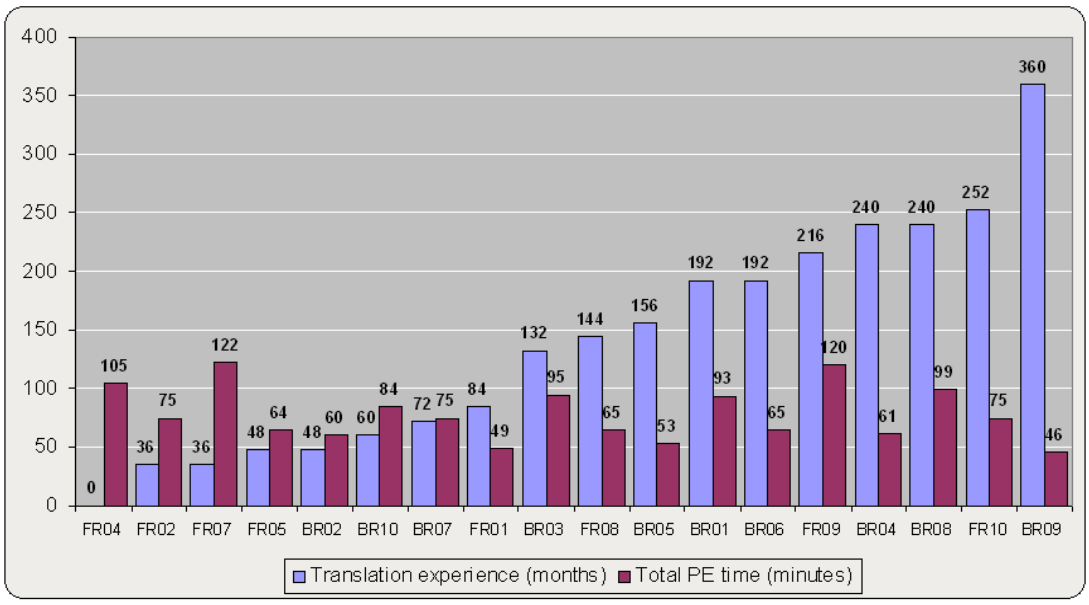


Figure 4.8: Comparison between PE experience and PE time - French and Brazilian Portuguese combined

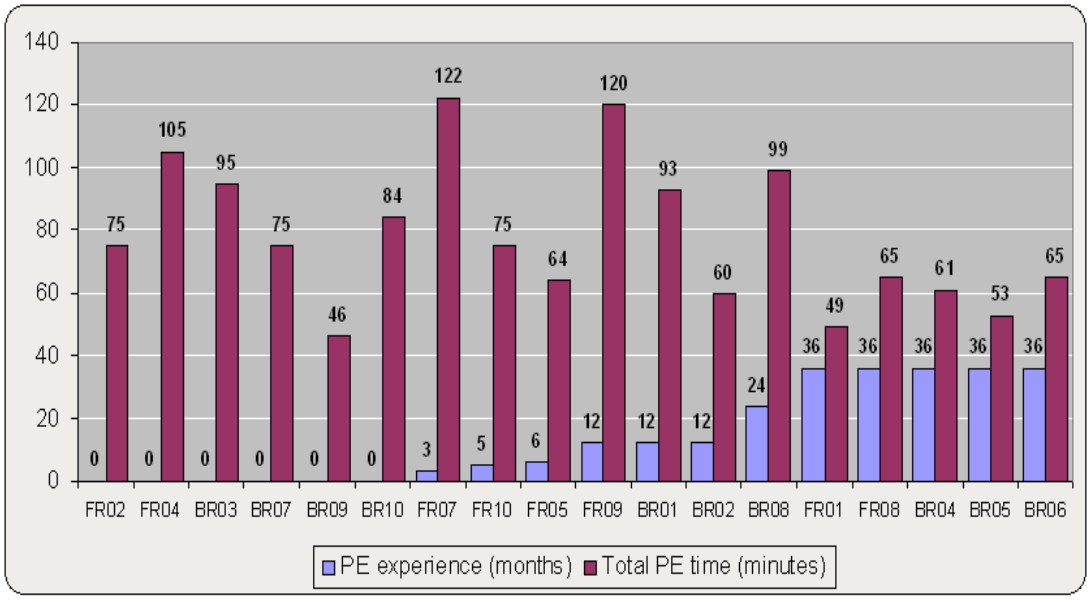
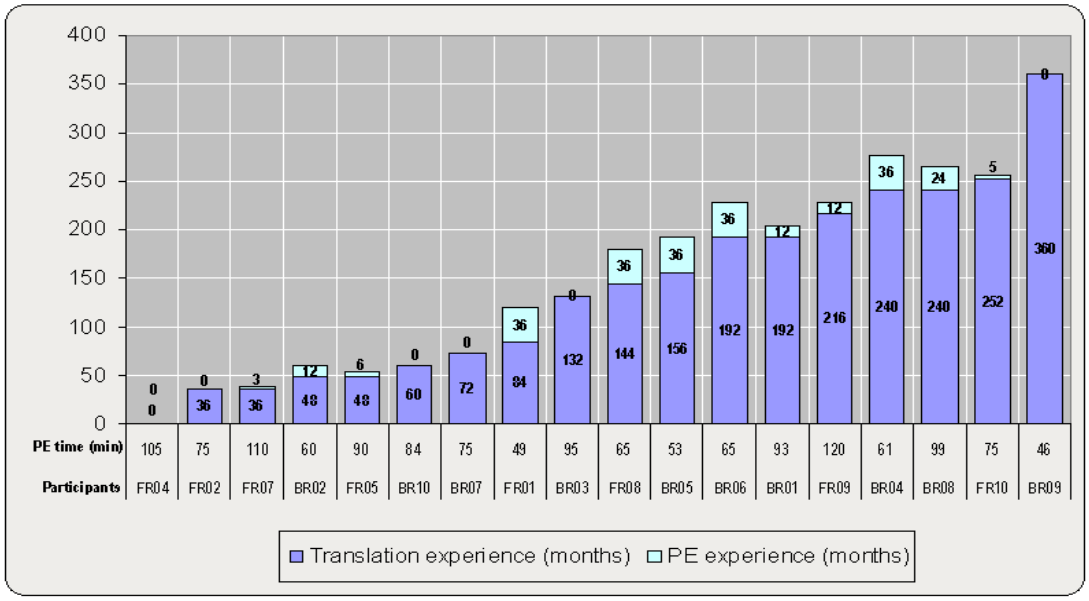


Figure 4.9: Comparison between translation and PE experience (combined) and PE time - French and Brazilian Portuguese combined



The Pearson Product Moment Correlation was calculated to see if there were any significant correlations between translation experience and PE time and PE experience and PE time for both target languages combined. Tables 4.8 and 4.9 display the results.

Table 4.8: Pearson Product Moment Correlation: translation experience and PE time - French and Brazilian Portuguese combined

Statistic	Variable X: translation experience (months)	Variable Y: PE time (minutes)
R-value	-0.207	
Degrees of Freedom	16	
Number of Observations	18	

The correlation coefficient of -0.207 indicates a weak negative correlation: the PE time would decrease as the level of translation experience increases.

Table 4.9: Pearson Product Moment Correlation: PE experience and PE time - French and Brazilian Portuguese combined

Statistic	Variable X: PE experience (months)	Variable Y: PE time (minutes)
R-value	-0.418	
Degrees of Freedom	16	
Number of Observations	18	

The correlation coefficient of -0.418 indicates a moderate negative correlation: the PE time would decrease as the level of PE experience increases.

The results presented in this section, which suggest that there is no correlation between translation experience and total PE time for the two target languages investigated, are in line with the findings of Guerberof (2012). As part of her study, she clustered the participants of her experiment into four groups according to various criteria, such as level of experience with localisation and with PE, typing speed, and experience with tools. A statistical analysis was carried out, and the findings indicated the following:

> "the incidence of experience on the processing speed is not significantly different (...). Translators with more experience performed similarly to other very novice translators. Translators with less or no experience in post-editing were the slowest group but again the differences were not significant."
> (Guerberof 2012: 216)

Guerberof points out that a higher number of participants can have an effect on the results, which could also apply to the findings of the main PE project for the present research.

4.2.4. Comparisons between number and types of changes, total PE time and translation experience - French and Brazilian Portuguese

This section examines the results found for the main categories of the typology for both target languages, comparing them with the levels of translation and PE experience.

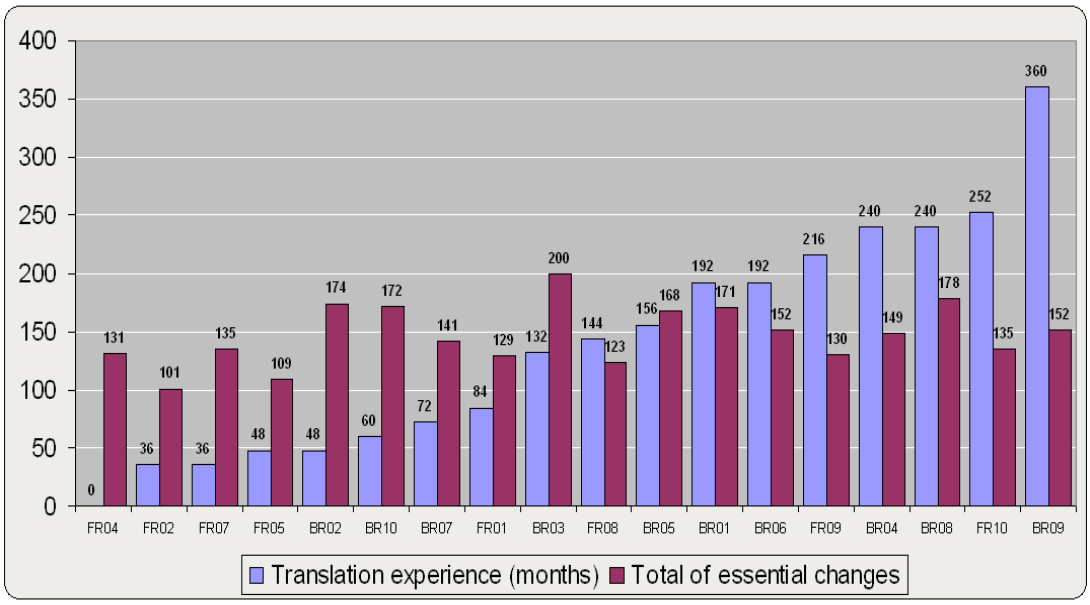
Table 4.10 provides a summary of translation experience and PE experience, total PE time, the total number of Essential and Preferential Changes, the total number of Essential Changes Not Implemented and the total number of Introduced Errors for each participant for French and Brazilian Portuguese. We further break down this data in the figures that follow (Figures 4.10 to 4.17) and our observations on the results follow these figures.

Table 4.10: Comparisons between number and types of changes, total PE time and translation and PE experience

Participant	Translation experience (months)	PE experience (months)	Total PE time (minutes)	Total essential changes	Total preferential changes	Total essential changes not impl.	Total introduced errors
FR01	84 months (7 years)	36 months (3 years)	49	129	26	18	10
FR02	36 months (3 years)	N/A	75	101	54	49	12
FR04	N/A	N/A	105	131	46	22	7
FR05	48 months (4 years)	6 months	64	109	95	24	14
FR07	36 months (3 years)	3 months	122	135	47	11	4
FR08	144 months (12 years)	36 months (3 years)	65	123	38	24	6
FR09	216 months (18 years)	12 months (1 year)	120	130	51	21	8
FR10	252 months (21 years)	5 months	75	135	52	9	9
BR01	192 months (16 years)	12 months (1 year)	93	171	58	21	10
BR02	48 months (4 years)	12 months (1 year)	60	174	28	31	10
BR03	132 months (11 years)	N/A	95	200	64	14	22
BR04	240 months (20 years)	36 months (3 years)	61	149	58	53	28
BR05	156 months (13 years)	36 months (3 years)	53	168	50	41	10
BR06	192 months (16 years)	36 months (3 years)	65	152	9	60	7
BR07	72 months (6 years)	N/A	75	141	37	54	19
BR08	240 months (20 years)	24 months (2 years)	99	178	35	29	8
BR09	360 months (30 years)	N/A	46	152	27	58	8
BR10	60 months (5 years)	N/A	84	172	38	31	15
Arithmetic mean	**139.33 (11.61 years)**	**14.11 (1.17 year)**	**78.11**	**147.22**	**45.16**	**31.66**	**11.5**
Median	**138 (11.5 years)**	**9**	**75**	**145**	**46.5**	**26.5**	**10**
Standard deviation	**98.45**	**233.75**	**534.33**	**26.03**	**18.67**	**16.66**	**6.11**

The data are illustrated in more detail in the next figures.

Figure 4.10: Comparison between translation experience and total of Essential Changes



Although the chart suggests that there is little correlation between translation experience and the total of Essential Changes, it indicates that participants with a high level of translation experience seem to have less variation in the total of Essential Changes. A similar trend is suggested by Figure 4.11: participants with a high level of PE experience seem to have less variation in the total of Essential Changes.

Figure 4.11: Comparison between PE experience and total of Essential Changes

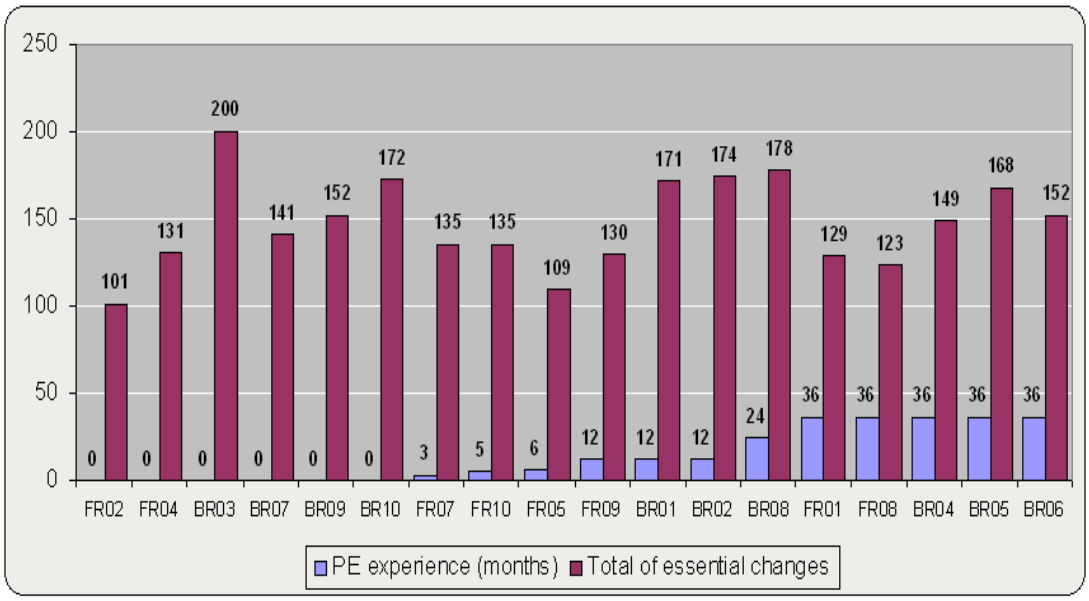
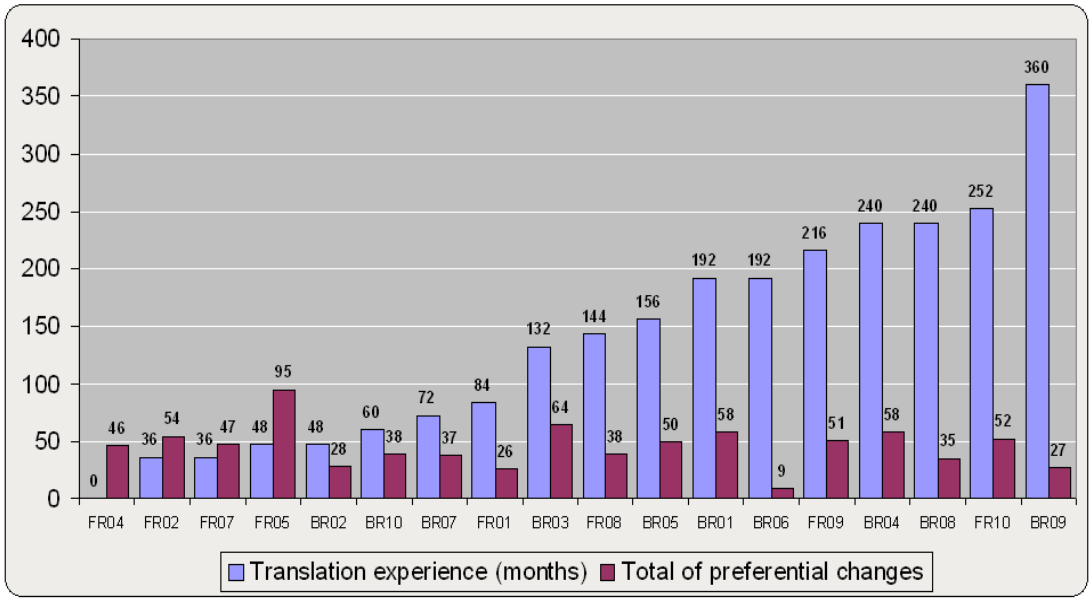


Figure 4.12: Comparison between translation experience and total of Preferential Changes



It is interesting to observe that the two participants who made the highest number of preferential changes (FR05, with 95 changes, and BR03, with 64 changes) were not among the most experienced translators. Both of these participants expressed either negative or neutral views

regarding MT and PE (as explained in Section 4.3). More data and more participants would be necessary in order to confirm these observations, but it might be possible to speculate that the level of translation experience and the views on MT and PE of these two participants might have had an effect on the number of preferential changes implemented by them.

Figure 4.13: Comparison between PE experience and total of Preferential Changes

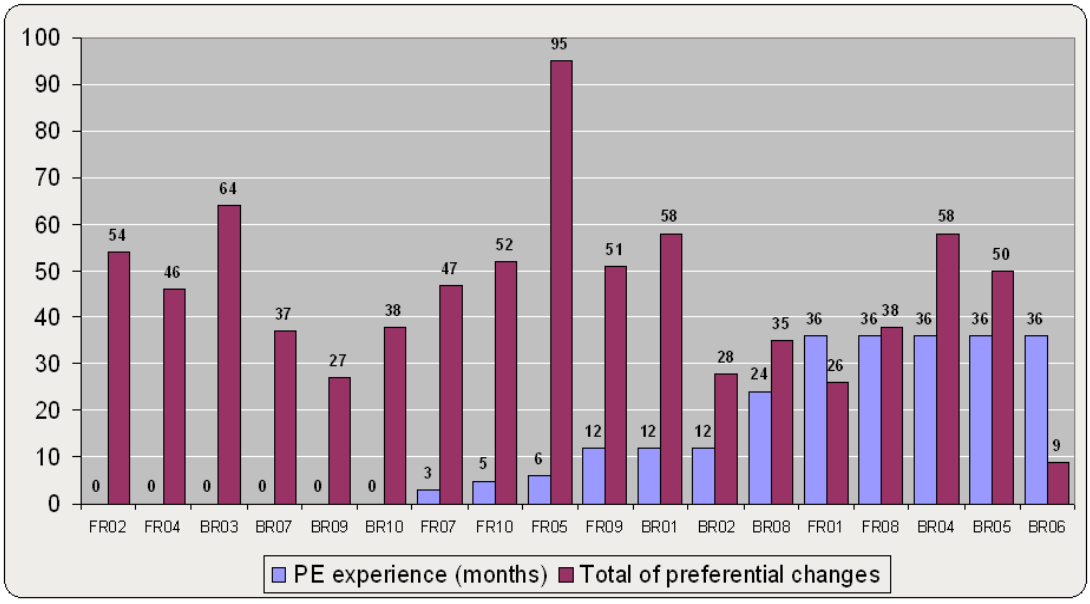


Figure 4.14: Comparison between translation experience and total of Essential Changes Not Implemented

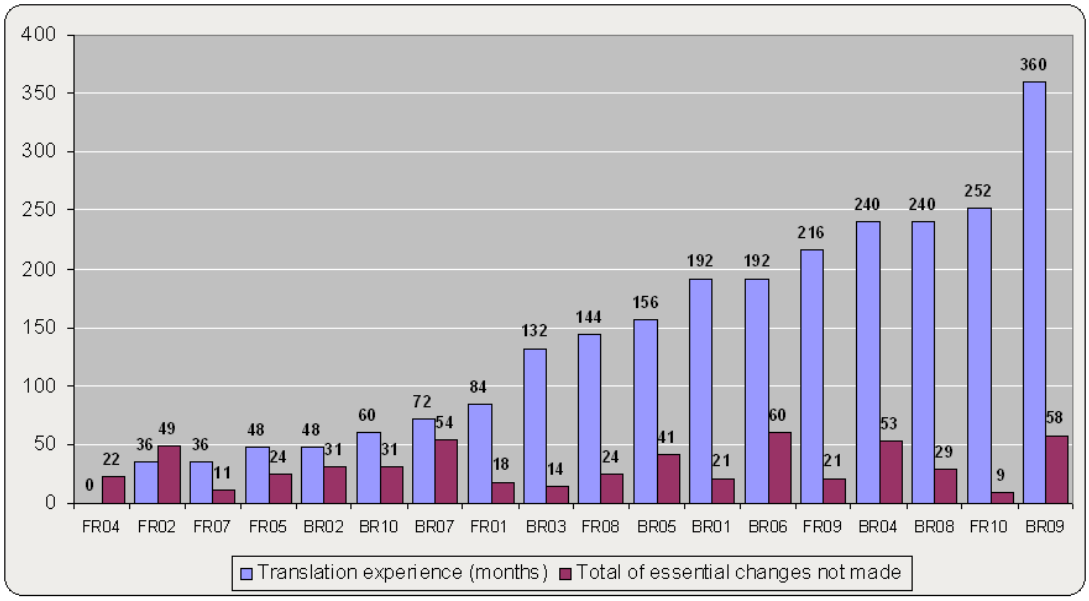


Figure 4.15: Comparison between PE experience and total of Essential Changes Not Implemented

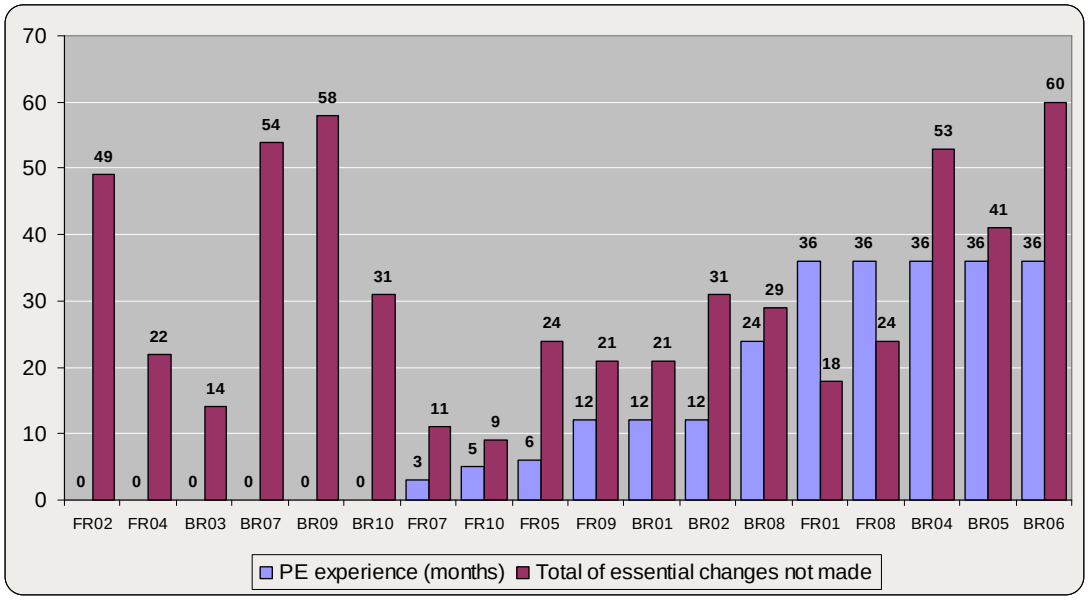


Figure 4.16: Comparison between translation experience and total of introduced errors

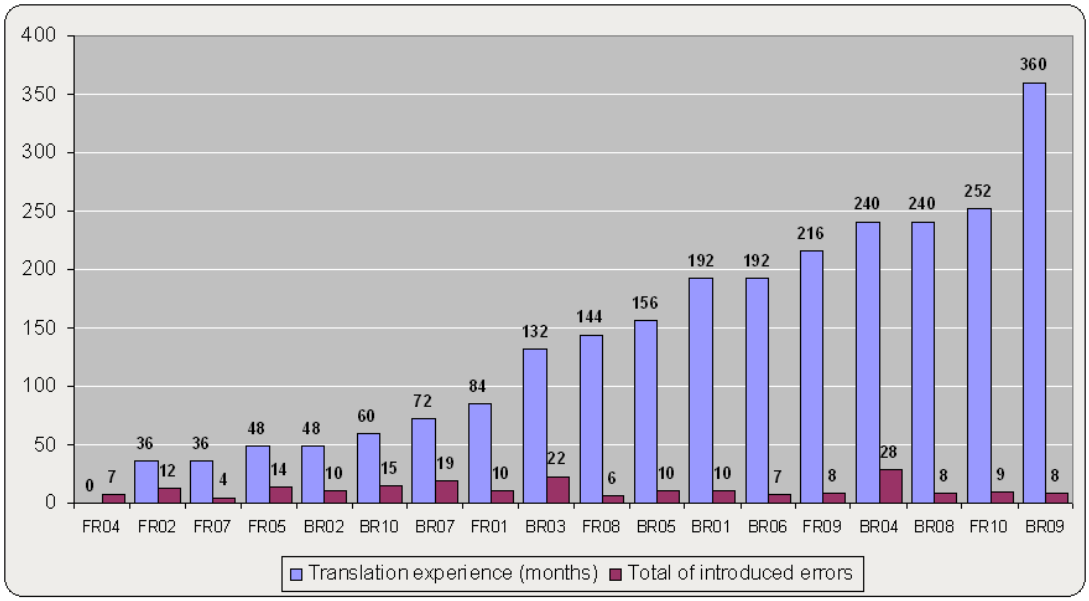
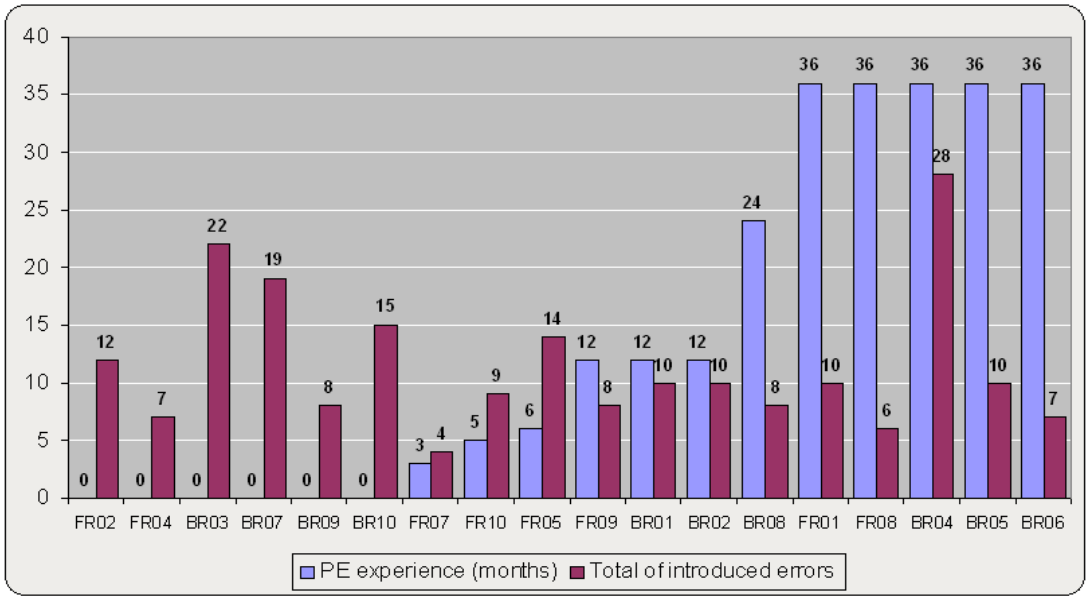


Figure 4.17: Comparison between PE experience and total of introduced errors



The previous figures seem to suggest that participants with an intermediate level of PE experience (between six and twelve months) seem to introduce fewer errors (e.g. Figure 4.17), have fewer Essential

Changes Not Implemented than those with lower or higher levels of PE experience (Figure 4.15), but implement more Preferential Changes (Figure 4.13).

The Pearson Product Moment Correlation was calculated, and Table 4.11 shows the results obtained.

Table 4.11: Results of Pearson Product Moment Correlation: main categories from the typology and translation and PE time - French and Brazilian Portuguese combined

Values correlated	R-value
Translation experience and total of Essential Changes	0.270
PE experience and total of Essential Changes	0.039
Translation experience and total of Preferential Changes	-0.187
PE experience and total of Preferential Changes	-0.309
Translation experience and total of Essential Changes Not Implemented	0.260
PE experience and total of Essential Changes Not Implemented	0.186
Translation experience and total of Introduced Errors	0.009
PE experience and total of Introduced Errors	-0.035

The correlation coefficients returned indicate the following:

- A weak positive correlation between translation experience and the total of Essential Changes (the number of Essential Changes increases as the translation experience increases).
- No correlation between PE experience and the total of Essential Changes.

- A weak negative correlation between translation experience and total of Preferential Changes (the number of Preferential Changes decreases as the level of translation experience increases).
- A moderate negative correlation between PE experience and total of Preferential Changes (the number of Preferential Changes decreases as the level of translation experience increases).
- A weak positive correlation between translation experience and the total of Essential Changes Not Implemented (the number of Essential Changes Not Implemented increases as the translation experience increases).
- A weak positive correlation between PE experience and the total of Essential Changes Not Implemented (the number of Essential Changes Not Implemented increases as the translation experience increases).
- No correlation between translation experience and the total of introduced errors.
- No correlation between PE experience and the total of introduced errors.

4.2.5. General observations

The first research question of the present project is:

Does the level of previous experience with translation influence the performance of translators when post-editing?

The results found up to this section suggest that the level of translation and PE experience does not influence the performance of post-editors. It was possible to see that clusters of participants presented common trends, namely, participants with a high level of translation experience seemed to have less variation in the total of Essential Changes, and participants with an intermediate level of PE experience (between six and twelve months) seemed to introduce fewer errors, have fewer Essential Changes Not Implemented, but implement more Preferential Changes.

Additional observations can be drawn if we divide the participants in two groups, according to their level of translation experience. For the first group, with a level of translation experience of up to 132 months (therefore below the mean value of 139.33), the mean value calculated for the number of Introduced Errors is 12.55, and the median is 12. For the second group, with a level of translation experience of 144 months or more (therefore above the mean value of 139.33), the mean value obtained for the number of Introduced Errors is 10.44 and the median is 8. The mean and median values for Introduced Errors found for the group of more experienced translators are below the overall mean and median values for this category (which were 11.5 and 10, respectively). On the other hand, the mean and median values for Introduced errors found for the group of less experienced translators are above the overall mean and median values. This would suggest that the level of previous translation did play a role in the number of errors introduced by the participants of the present study, at least to a certain extent.

Although the values found for the Pearson Product Moment Correlation for both languages suggest no correlations, several additional elements need to be taken into consideration. The next sections explore in more detail other aspects of the PE sessions carried out by the participants, in order to derive further insight from the results.

4.3. Translators' bias regarding PE

As explained in Chapter 3, a three-item questionnaire was answered by the participants before carrying out the individual PE sessions. The objective of the questionnaire was to gauge the participants' views regarding MT and PE in order to see if specific views might be linked with specific PE activity.

The results of the questionnaire are presented in this section. Each question is presented separately, followed by the results and the analysis.

Question 1

Taking into account your previous experience with machine translation (using online MT engines, using MT as part of translation/localisation projects or in any other context, as a translator or as a user), please highlight one of the options below to indicate your opinion regarding the general quality of translated texts produced by machine translation only (without post-editing).

Table 4.12: Answers to Question 1 of the questionnaire - French

Participant	Answers				
	1 - Very bad quality	2 - Bad quality	3 - Average quality	4 - Good quality	5 - Very good quality
FR01			●		
FR02	●				
FR04			●		
FR05		●			
FR07		●			
FR08				●	
FR09			●		
FR10			●		

Table 4.13: Answers to Question 1 of the questionnaire - Brazilian Portuguese

Participant	Answers				
	1 - Very bad quality	2 - Bad quality	3 - Average quality	4 - Good quality	5 - Very good quality
BR01			●		
BR02				●	
BR03			●		
BR04		●			
BR05			●		
BR06				●	
BR07		●			
BR08	●				
BR09	●				
BR10			●		

Question 2

Please highlight one of the options below to indicate your opinion about the following statement:

Machine translation can be helpful for translators (as a productivity tool, for instance).

Table 4.14: Answers to Question 2 of the questionnaire - French

Participant	Answers				
	1 - Strongly disagree	2 - Disagree	3 - Neither agree nor disagree	4 - Agree	5 - Strongly agree
FR01				●	
FR02					●
FR04				●	
FR05		●			
FR07				●	
FR08				●	
FR09			●		
FR10				●	

Table 4.15: Answers to Question 2 of the questionnaire - Brazilian Portuguese

Participant	Answers				
	1 - Strongly disagree	2 - Disagree	3 - Neither agree nor disagree	4 - Agree	5 - Strongly agree
BR01			●		
BR02				●	
BR03			●		
BR04				●	
BR05			●		
BR06					●
BR07				●	
BR08				●	
BR09			●		
BR10				●	

Question 3

Please highlight one of the options below to indicate your opinion about the following statement:

Post-editing texts produced by machine translation is an activity that interests me as a translator, as it can provide me with new sources of work and new professional skills.

Table 4.16: Answers to Question 3 of the questionnaire - French

Participant	Answers				
	1 - Strongly disagree	2 - Disagree	3 - Neither agree nor disagree	4 - Agree	5 - Strongly agree
FR01				●	
FR02		●			
FR04				●	
FR05			●		
FR07			●		
FR08					●
FR09				●	
FR10				●	

Table 4.17: Answers to Question 3 of the questionnaire - Brazilian Portuguese

Participant	Answers				
	1 - Strongly disagree	2 - Disagree	3 - Neither agree nor disagree	4 - Agree	5 - Strongly agree
BR01			●		
BR02				●	
BR03		●			
BR04		●			
BR05			●		
BR06				●	
BR07				●	
BR08					●
BR09				●	
BR10				●	

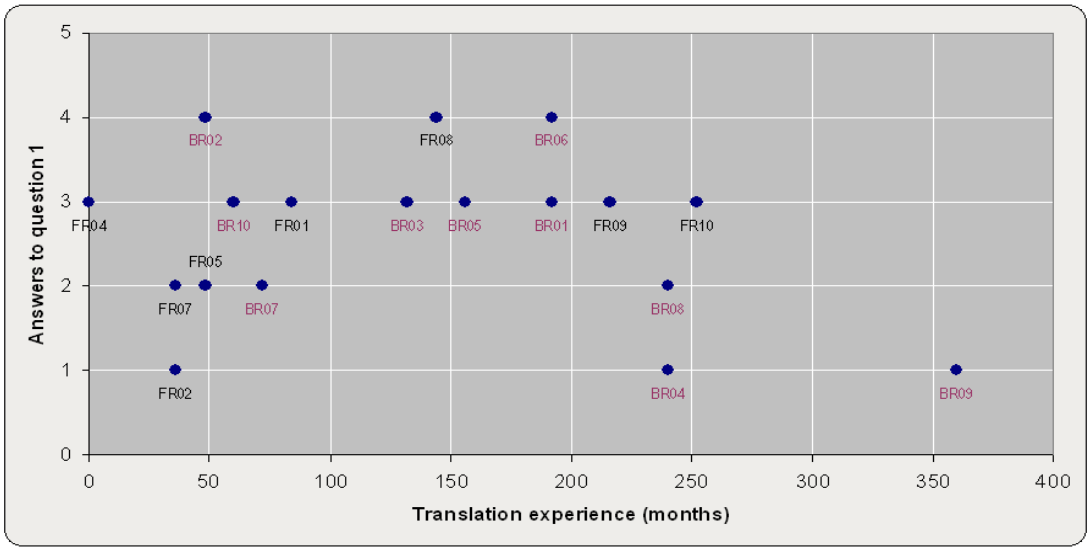
4.3.1. Comparison between translation/PE experience and answers to the questionnaire

This section investigates the answers to the questionnaire provided by the participants, compared to their level of translation and PE experience. The questions and the scale of answers are repeated here for convenience.

Question 1

Taking into account your previous experience with machine translation (using online MT engines, using MT as part of translation/localisation projects or in any other context, as a translator or as a user), please highlight one of the options below to indicate your opinion regarding the general quality of translated texts produced by machine translation only (without post-editing).

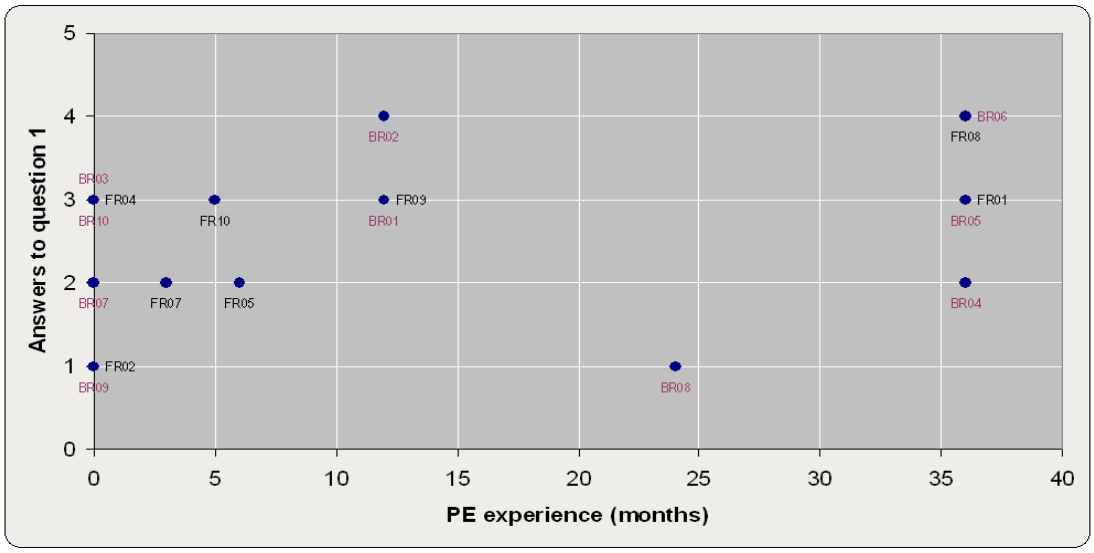
Figure 4.18: Comparison between translation experience and answers to Question 1 - French and Brazilian Portuguese combined



Scale of answers:
1 - Very bad quality
2 - Bad quality
3 - Average quality
4 - Good quality

5 - Very good quality

Figure 4.19: Comparison between PE experience and answers to Question 1 - French and Brazilian Portuguese combined



Scale of answers:
1 - Very bad quality
2 - Bad quality
3 - Average quality
4 - Good quality
5 - Very good quality

Regarding the results shown in Figures 4.18 and 4.19, of the three post-editors who chose 1 ("Very bad quality") as the answer to the first question (regarding the level of quality of raw MT output), two of them (BR04 and BR09) were very experienced translators (with 20 and 30 years of experience, respectively), and also two of the fastest of all the participants (with total PE times of 61 and 46 minutes, respectively). FR02 had considerably less translation experience (three years), but also had a relatively short PE time (75 minutes). Neither BR09 nor FR02 had PE experience, but BR04 had three years of experience as a post-editor.

It is possible to speculate that BR09's and FR02's views may have been somewhat influenced by the fact that they did not have PE experience, and may have formed their opinions due to other types of

exposure to MT (presumably with low-quality output). On the other hand, despite having experience with PE, BR04 may have worked in projects that also had low-quality output.

Three post-editors chose 4 ("Good quality") as the answer to question 2: BR02, BR06 and FR08. BR02 had three years of experience with translation and one year of experience with PE. BR06 and FR08 were experienced translators (with 16 and 12 years of experience, respectively) and post-editors (both with three years of PE experience). In this case, it is possible to speculate that their experiences with PE may have involved a better level of quality, and, therefore, a better predisposition for MT.

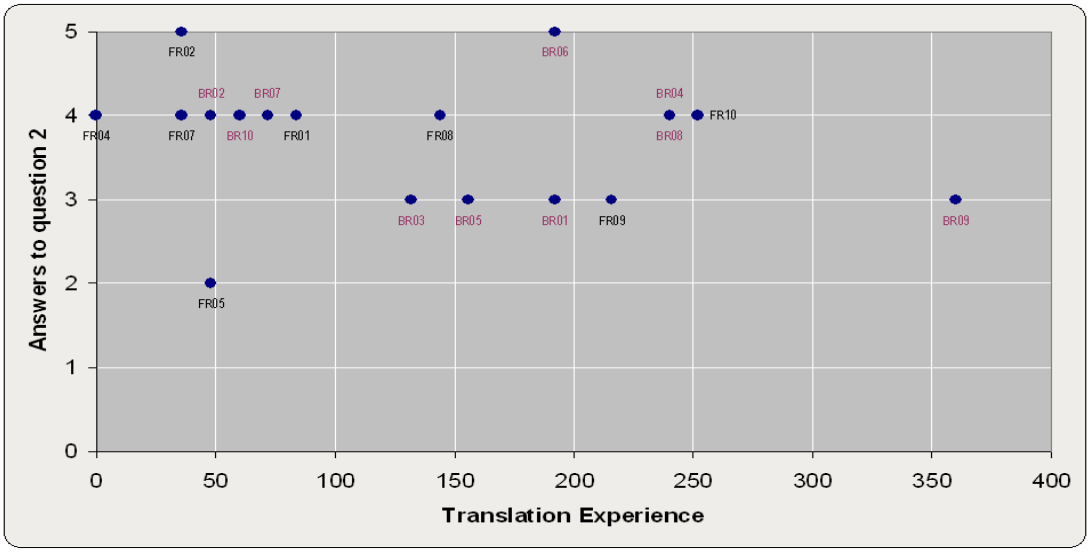
The trend that seems to emerge from the answers to question 1, though, is that the majority of the participants (10) expressed either neutral or moderately positive views on the quality of raw MT output.

Question 2

Please highlight one of the options below to indicate your opinion about the following statement:

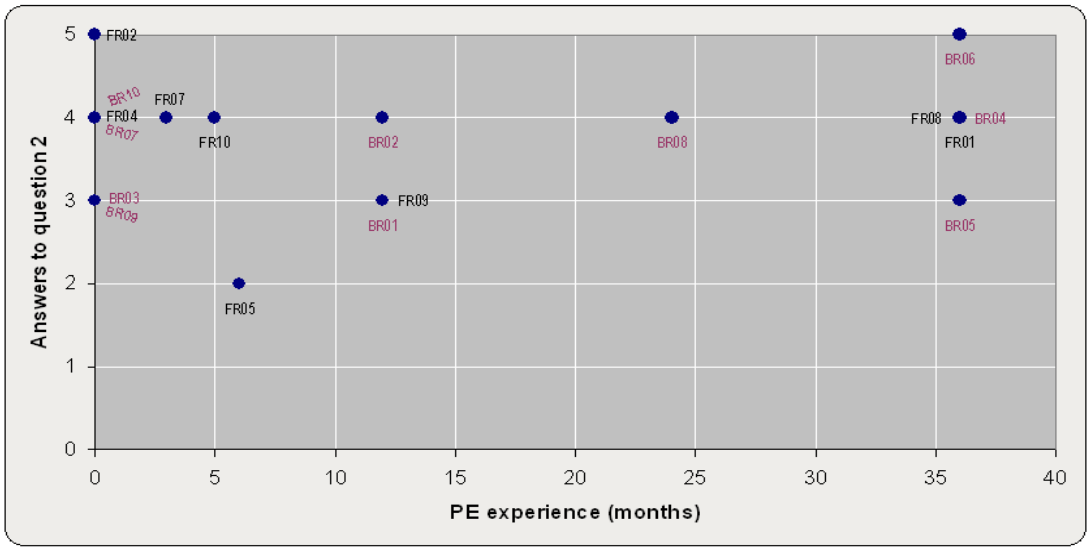
Machine translation can be helpful for translators (as a productivity tool, for instance).

Figure 4.20: Comparison between translation experience and answers to Question 2 - French and Brazilian Portuguese combined



Scale of answers:
1 - Strongly disagree
2 - Disagree
3 - Neither agree nor disagree
4 - Agree
5 - Strongly agree

Figure 4.21: Comparison between PE experience and answers to Question 2 - French and Brazilian Portuguese combined



Scale of answers:
1 - Strongly disagree
2 - Disagree
3 - Neither agree nor disagree
4 - Agree
5 - Strongly agree

The only participant who expressed a moderately negative view (by choosing answer 2, "Disagree") for Question 2 was FR05. This participant had four years of translation experience and six months of experience with PE, and had a total PE time of 64 minutes.

The trend observed in Figures 4.20 and 4.21 is that, except for FR05, all the other participants expressed neutral or positive views. Two participants (FR02 and BR06) selected answer 5 ("Strongly agree"). It is interesting to observe that, despite selecting answer 5 for question 2, and thus signalling strong agreement that MT can be useful for translators, FR02 selected answer 1 for question 1, rating the quality of machine-translated texts as "Very bad". This apparent contradiction may have different explanations, but one possible reason for it is that FR02

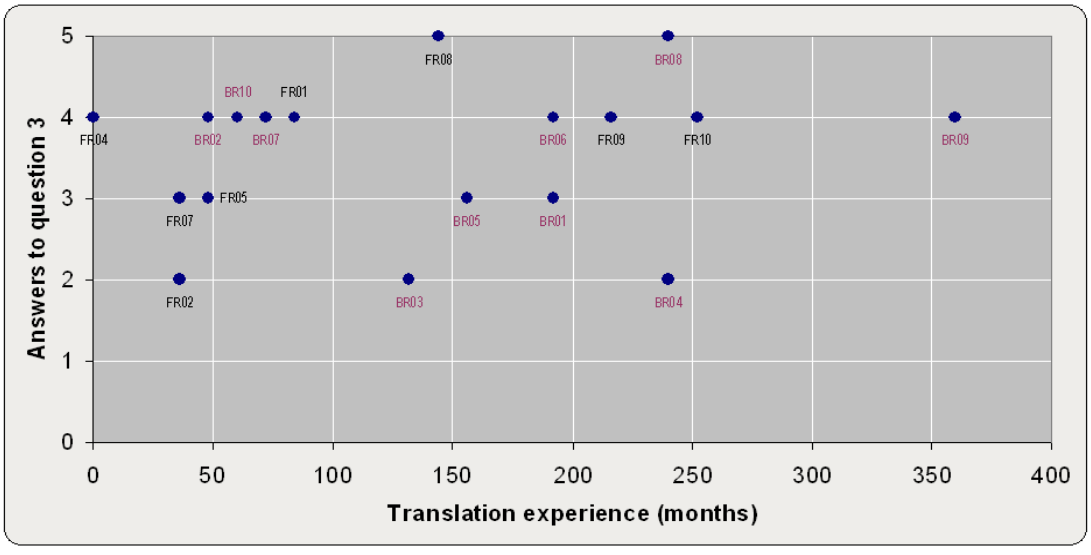
considers that, despite the bad quality of the output, MT can still be useful for translators as a productivity tool.

Question 3

Please highlight one of the options below to indicate your opinion about the following statement:

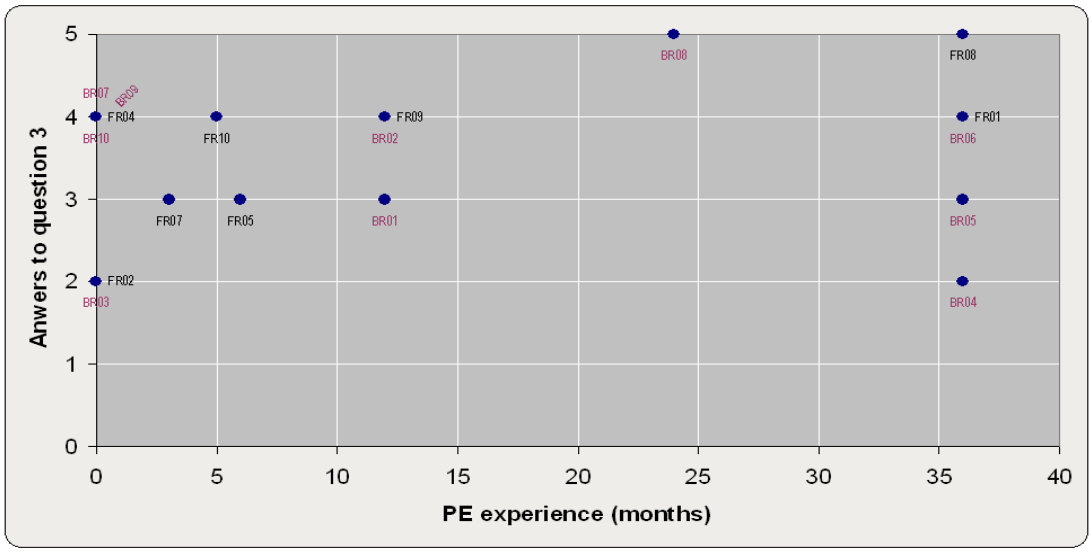
Post-editing texts produced by machine translation is an activity that interests me as a translator, as it can provide me with new sources of work and new professional skills.

Figure 4.22: Comparison between translation experience and answers to Question 3 - French and Brazilian Portuguese combined



Scale of answers:
1 - Strongly disagree
2 - Disagree
3 - Neither agree nor disagree
4 - Agree
5 - Strongly agree

Figure 4.23: Comparison between PE experience and answers to Question 3 - French and Brazilian Portuguese combined



Scale of answers:
1 - Strongly disagree
2 - Disagree
3 - Neither agree nor disagree
4 - Agree
5 - Strongly agree

BR03, BR04 and FR02 chose answer 2 ("Disagree") for question 3, signalling that PE is not an activity that interests them. Similarly to what was pointed out for question 1, it is possible to speculate that their views may have been somewhat influenced by lack of exposure to PE (in the case of BR03 and FR02) and/or by previous experiences with MT (not necessarily as post-editors) involving low-quality raw MT output. The remaining participants expressed neutral or positive views. BR08 and FR08 chose answer 5 ("Strongly agree"), expressing the highest level of agreement. Both were experienced translators (with 20 and 12 years of experience, respectively) and post-editors (with two and three years of experience, respectively).

4.3.2. General observations

It is interesting to note that most of the participants seemed to express positive or neutral opinions regarding MT and PE in their answers to the three questions. The exceptions were FR02, BR08 and BR09, who considered the quality of MT very bad (answer to Question 1: 1-Very bad quality), FR05, who did not consider MT as a useful tool for translators (answer to Question 2: 2-Disagree), and FR02, BR03 and BR04, who did not consider PE an activity that would interest them as translators (answer to Question 3: 2-Disagree).

It is relevant to verify the experience levels of the participants who expressed negative views, to determine if further clues can be obtained. Table 4.18 provides more details about this.

Table 4.18: Views of the participants

Participant	Translation experience (months)	PE experience (months)	Answer to Question 1	Answer to Question 2	Answer to Question 3
FR02	36	0	1-Very bad quality		2-Disagree
FR05	48	6		2-Disagree	
BR03	132	0			2-Disagree
BR04	240	36			2-Disagree
BR08	240	24	1-Very bad quality		
BR09	360	0	1-Very bad quality		

Half of the participants who expressed negative views had PE experience, and amongst this group, two of them had a level of PE experience above the mean value of 14.11. These two participants (BR04 and BR08) also had a level of translation experience above the mean value of 139.33 (both had 240 months, or 20 years of experience). Further

studies with a higher number of participants would be necessary to validate this hypothesis, but it is possible to speculate that, despite the fact that their level of PE experience was above the average, their level of translation experience might also have some influence on their views regarding PE. Also, it is possible that they may have dealt with PE projects that had an MT output of bad quality, which may have resulted in the formation of negative opinions about PE.

With the exception of FR05, who had six months of PE experience, the remaining participants who expressed negative opinions did not have any PE experience at all. In their case, the fact that they had little or no exposure to PE might have influenced the views that they expressed.

Finally, regarding the fact that the majority of the participants expressed positive opinions, it is interesting to observe that this corresponds to the findings of other recent research projects in the field of PE and MT. Guerberof (2012: 259) mentions that, overall, the approach of the participants in her study towards MT was "flexible and practical", and the overall attitude was positive. Tatsumi (2010: 198) observes that the participants in her study displayed a "flexible and down-to-earth attitude towards PE". García (2010) found a similar trend in a study involving English to Chinese SMT output. It is important to mention that, with the exception of FR04 (who was a postgraduate translation student), all the participants in the present research were technical translators with different levels of experience in the localisation industry, so this may also be a factor to be considered in relation to the majority of positive opinions expressed.

4.4. Comparisons between translation experience and PE experience and keyboard and mouse use

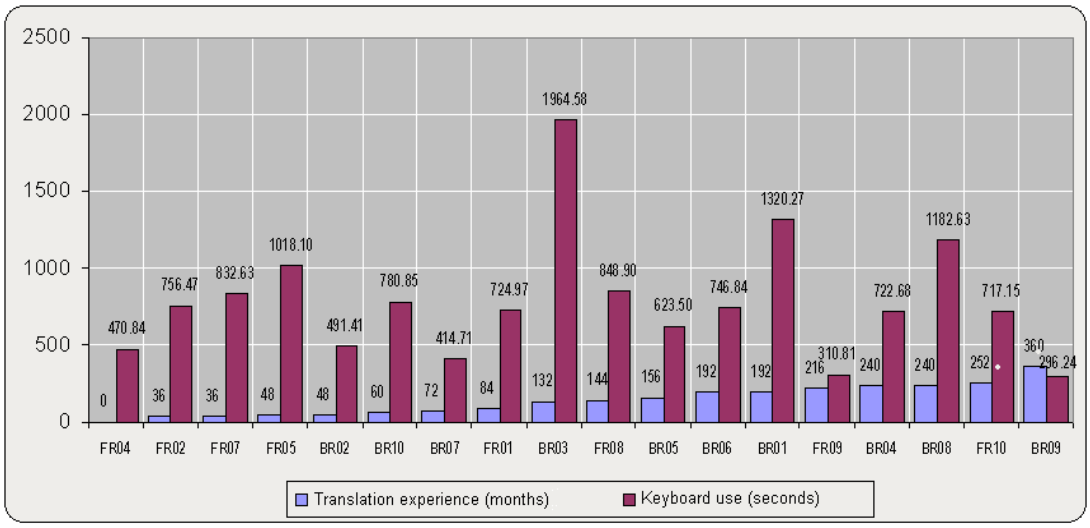
This section presents comparisons between the use of keyboard and mouse, and the levels of translation and PE experience of the participants. Comments are provided after the tables and figures. "Keyboard and mouse usage" here means any time keys are pressed or mouse movements and clicks are performed (as opposed to no keys being pressed or no mouse actions being performed).

Data on keyboard and mouse usage were included in the present research because of the possible relationship with productivity. The underlying hypothesis was that efficient use of input methods would be beneficial for the overall productivity, and would potentially contribute towards a good overall PE performance. In that respect, minimising the number of switches between the two input methods, for instance, could be a helpful strategy for optimising the use of keyboard and mouse. Favouring the use of the keyboard over the mouse could be considered another optimisation technique. Therefore, it was of interest to see how the data on the input methods would correlate with other elements being investigated here. Table 4.19 displays the results found, which are further illustrated by figures 4.24 to 4.30. The Pearson correlations for the values are also presented in this section.

Table 4.19: Comparisons of translation/PE experience, measures recorded and keyboard and mouse use

Participant	Trans. exper. (months)	PE exper. (months)	Total PE time (minutes)	Essent. changes	Pref. changes	Essent. changes not impl.	Introduced errors	Keyb. time (sec)	Mouse time (sec)	Switches
FR01	84	36	49	129	26	18	10	724.97	1483.76	106
FR02	36	0	75	101	54	49	12	756.47	3627.83	106
FR04	0	0	105	131	46	22	7	470.84	6070.06	279
FR05	48	6	64	109	95	24	14	1018.10	467.51	126
FR07	36	3	122	135	47	11	4	832.63	6184.58	321
FR08	144	36	65	123	38	24	6	848.90	2326.08	123
FR09	216	12	120	130	51	21	8	310.81	2777.77	108
FR10	252	5	75	135	52	9	9	717.15	1834.68	125
BR01	192	12	93	171	58	21	10	1320.27	1758.45	137
BR02	48	12	60	174	28	31	10	491.41	2817.73	385
BR03	132	0	95	200	64	14	22	1964.58	1597.42	239
BR04	240	36	61	149	58	53	28	722.68	1157.76	177
BR05	156	36	53	168	50	41	10	623.50	2249.40	155
BR06	192	36	65	152	9	60	7	746.84	2345.92	170
BR07	72	0	75	141	37	54	19	414.71	3485.32	191
BR08	240	24	99	178	35	29	8	1182.63	4116.37	299
BR09	360	0	46	152	27	58	8	296.24	2114.69	268
BR10	60	0	84	172	38	31	15	780.85	641.73	124
Arith. mean	139.33	14.11	80	147.22	45.16	31.66	11.5	790.19	2614.28	191.05
Median	138	9	82.5	145	46.5	26.5	10	735.90	2287.74	162.5
St. dev.	98.45	15.28	20.79	26.03	18.67	16.66	6.11	400.17	1601.67	86.11

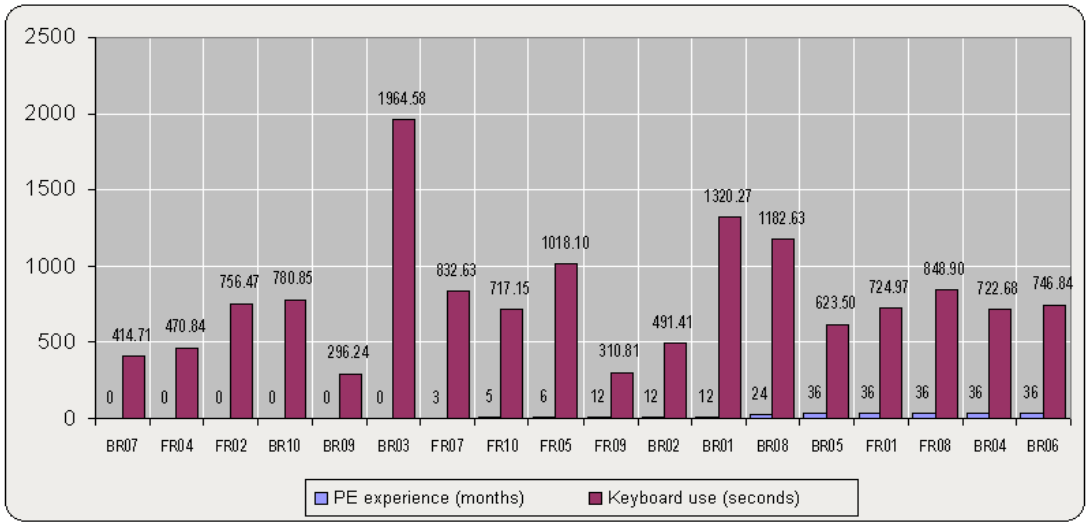
Figure 4.24: Comparisons between translation experience and keyboard use



No obvious trends seem to emerge in Figure 4.24. Some of the participants with the highest levels of translation experience had keyboard times slightly higher than the less experienced translators. There were a few exceptions, however: BR03, who has an intermediate level of translation experience (132 months, or 11 years) in comparison with the

other participants, had the highest keyboard time, and BR09, who has a higher level of translation experience (360 months, or 30 years), had the lowest keyboard time of all.

Figure 4.25: Comparisons between PE experience and keyboard use



No obvious trends seem to emerge in Figure 4.25 either. Some of the participants who had PE experience (BR05, FR01, FR08, BR04 and BR06) presented low keyboard times, but the lowest keyboard times were recorded among participants with no PE experience (such as BR07, FR04 and BR09). On the other hand, BR03, one of the participants who did not have PE experience either, had the highest keyboard time.

Figure 4.26: Comparisons between total PE time and keyboard use

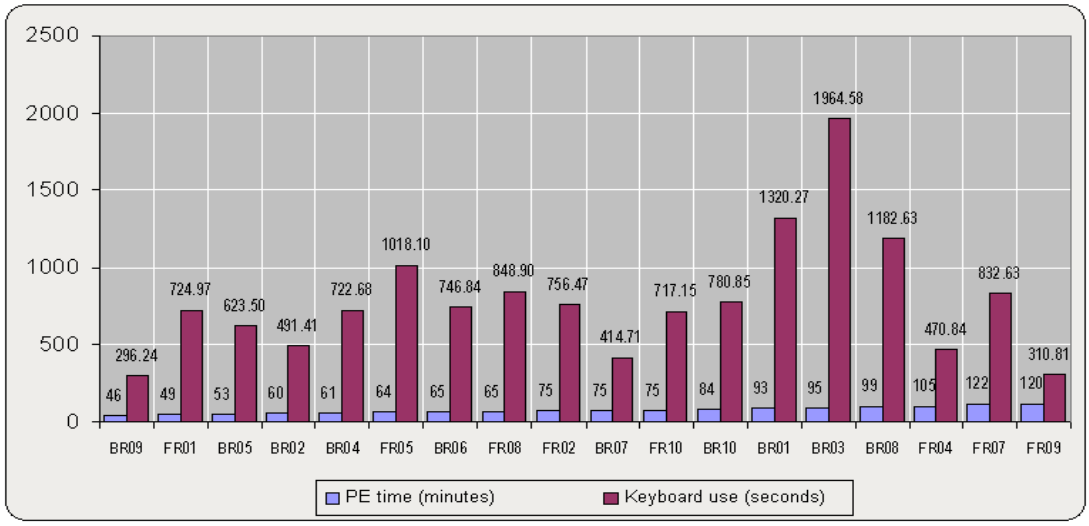


Figure 4.26 suggests that some of the participants with the highest PE times (BR01, BR03, BR08 and FR07) used the keyboard for longer periods than the post-editors with the lowest PE times. An exception would be FR05, who had a low PE time (64 minutes), yet a high value for keyboard use (1018.10 seconds).

Figure 4.27: Comparisons between translation experience and mouse use

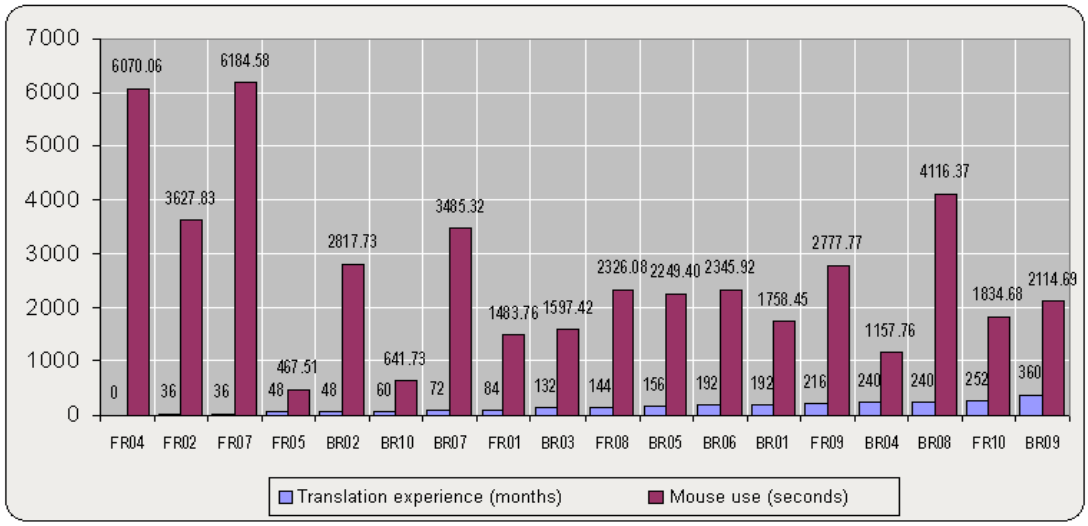
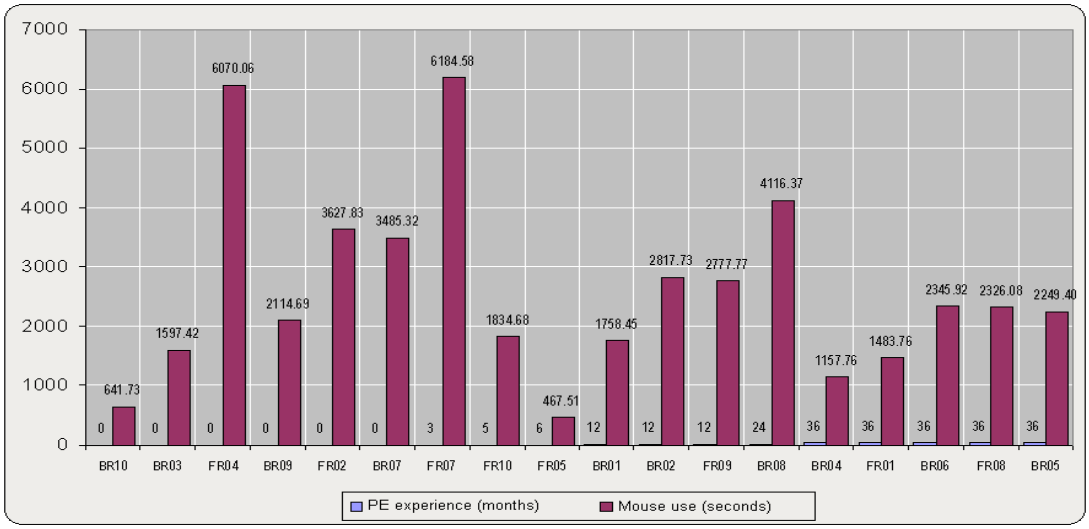


Figure 4.27 indicates a trend among participants with an intermediate level of translation experience, from FR01, with 84 months (7

years) to BR01, with 192 months (16 years). This group of post-editors presented intermediate mouse times, ranging from 1483.76 to 1758.45 seconds. Their mean value is 1958.50 (compared to the overall mean of 2614.28) and their median value is 2003.92 (compared to the overall median of 2287.74).

Figure 4.28: Comparisons between PE experience and mouse use



Again, an interesting trend can be seen Figure 4.28: the group of five participants with thirty six months of PE experience, or three years (BR04, FR01, BR06, FR08 and BR05) had intermediate values for mouse time, ranging from 1157.76 to 2249.40 seconds (with a mean value of 1912.58, compared to the overall mean of 2614.28, and a median value of 2249.4, compared to the overall median of 2287.74). It is also interesting to remark that two of the highest values for mouse use were recorded for two participants with little or no PE experience (FR04 and FR07).

Figure 4.29: Comparisons between total PE time and mouse use

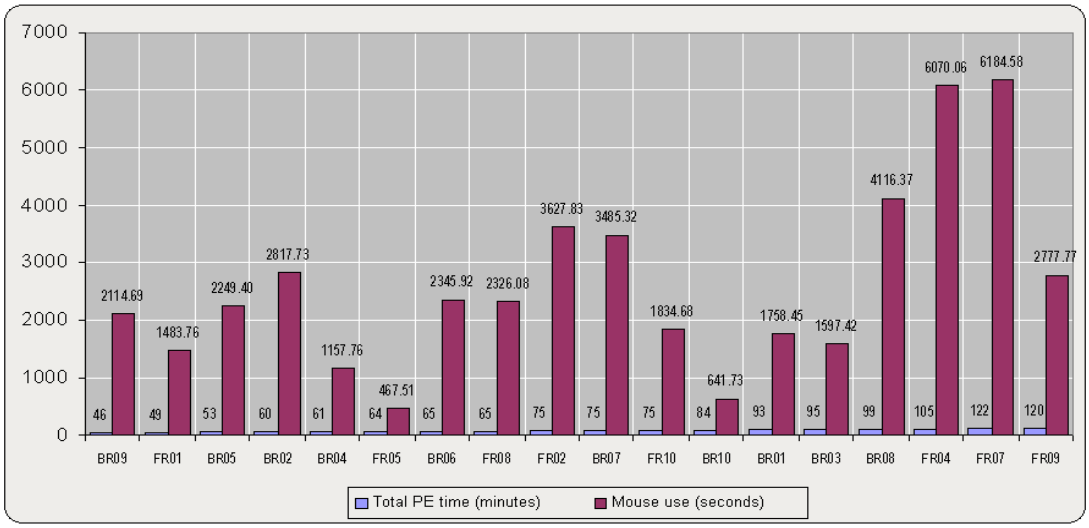


Figure 4.29 indicates that the last four participants with the highest PE times (BR08, FR04, FR07 and FR09) had some of the highest mouse times overall, ranging from 2777.77 to 6184.58 seconds, with a mean value of 4787.19, compared to the overall mean of 2614.28, and a median value of 5093.21, compared to the overall median of 2287.74.

Figure 4.30: Comparisons between translation experience and switches between keyboard and mouse

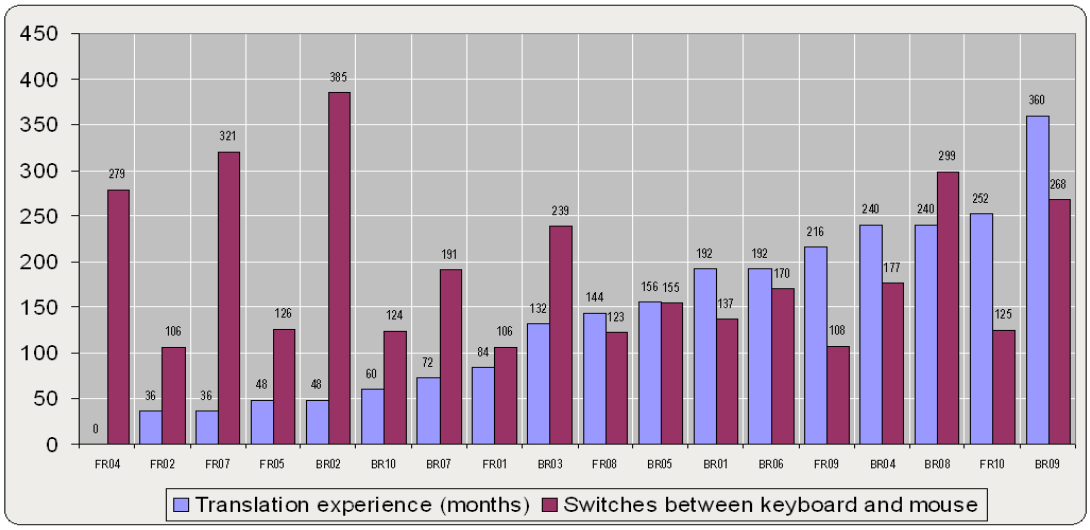


Figure 4.30 indicates that some of the participants with a higher level of translation experience seem to have lower numbers of switches,

such as FR09 and FR10. It is interesting to observe that the three highest numbers of switches were recorded for participants with a low level of translation experience: FR04, FR07 and BR02.

Figure 4.31: Comparisons between PE experience and switches between keyboard and mouse

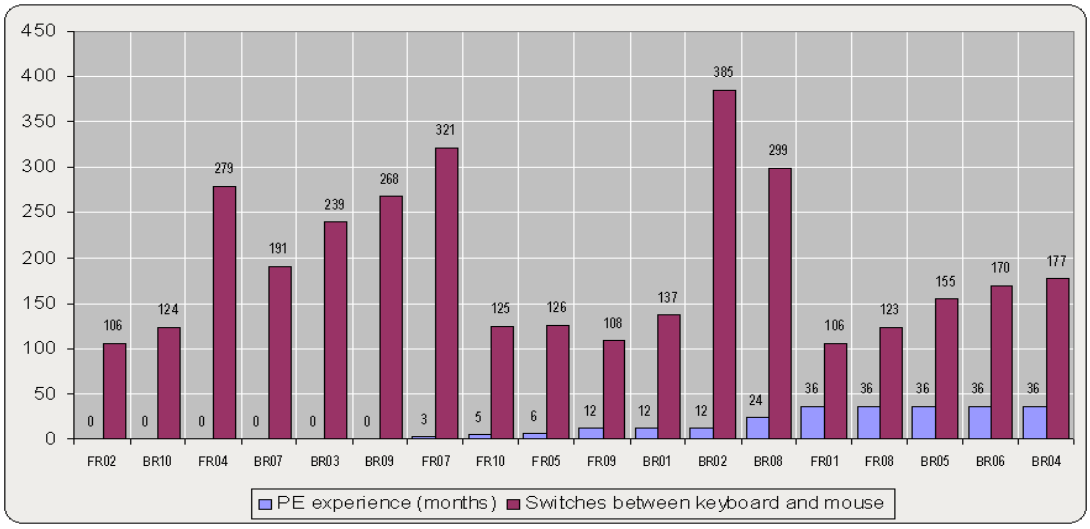


Figure 4.31 shows a trend amongst the participants with the highest level of PE experience (from FR01 upwards): they generally switched less often between mouse and keyboard when compared with the other participants. (with a mean value of 146.2, compared to the overall mean of 191.05, and a median value of 155, compared to the overall median of 162.5).

Figure 4.32: Comparisons between total PE time and switches between keyboard and mouse

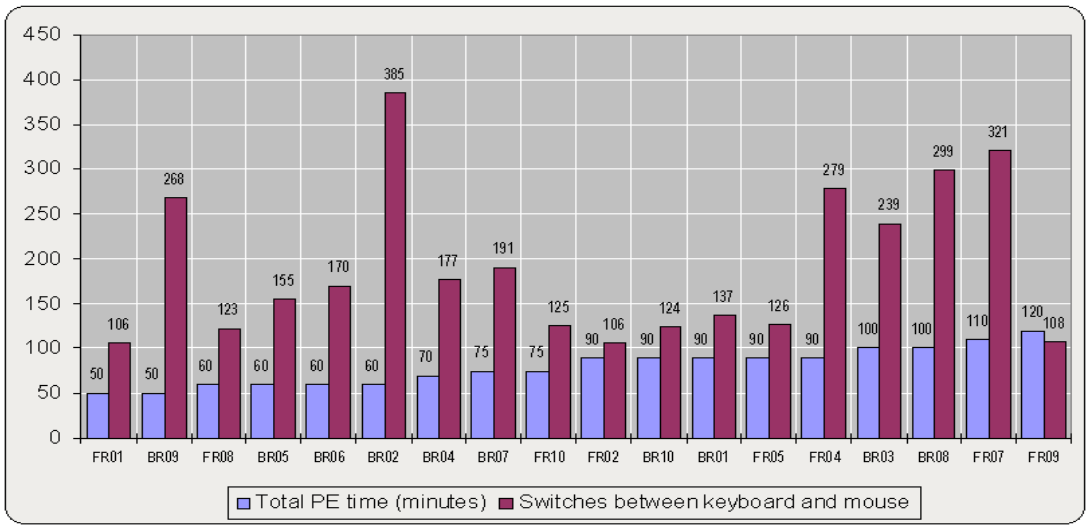


Figure 4.32 suggests a moderate trend amongst the participants with intermediate PE times, from FR10 to FR05. These participants had a low total number of switches (ranging from 106 to 137 switches), with a mean value of 123.6, compared to the overall mean of 191.05, and a median value of 125, compared to the overall median of 162.5.

Before proceeding to a more detailed investigation of the trends observed, the Pearson Product Moment Correlation was used to determine the level of correlation, if any, amongst all of the values presented up to now. The results are provided in table 4.20, followed by additional comments.

Table 4.20: Results of Pearson Product Moment Correlation: keyboard and mouse usage - French and Brazilian Portuguese combined

Values correlated	R-value
Translation experience and keyboard time	-0.040
PE experience and keyboard time	-0.025
Translation experience and mouse time	-0.319
PE experience and mouse time	-0.245
Translation experience and switches between keyboard and mouse	-0.077
PE experience and switches between keyboard and mouse	-0.229

The correlation coefficients returned indicate the following:

- No correlation between translation experience and keyboard time.
- No correlation between PE experience and keyboard time.
- A moderate negative correlation between translation experience and mouse time (the mouse time decreases as the level of translation experience increases).
- A weak negative correlation between PE experience and mouse time (the mouse time decreases as the PE experience increases).
- No correlation between translation experience and switches between keyboard and mouse.
- A moderate negative correlation between PE experience and switches between keyboard and mouse (the number of switches decreases as the PE experience increases).

4.4.1. General observations

A few trends were observed when looking at the results for specific groups of participants. In this section, such trends are examined in more detail, in order to obtain a better understanding of the results found.

It is possible to observe that participants with levels of translation experience up to 84 months (FR01, in this case), which are below the mean value of 139.33, had keyboard values ranging between 414.71 and 1018.10 seconds. BR03, who has 132 months of translation experience, seems to be an outlier amongst all the participants with regards to keyboard use, with a keyboard value of 1954.58 seconds. The remaining participants, with levels of translation experience ranging from 144 months (FR08) to 360 months (BR09), had keyboard values ranging between 296.24 and 848.90 seconds, except for BR01, whose keyboard time was 1320.27 seconds, and BR08, whose keyboard time was 1182 seconds. This suggests that, if BR01, BR03 and BR08 are excluded, the participants with levels of translation experience above the mean value of 139.33 had a lower average keyboard value than the participants with lower levels of translation experience, as indicated in tables 4.21 and 4.22.

Table 4.21: Comparison between translation experience (below the mean value) and keyboard use (excluding BR03)

Participant	Translation experience (months)	Total keyboard time (minutes)
FR04	0	470.84
FR02	36	756.47
FR07	36	832.63
FR05	48	1018.10
BR02	48	491.41
BR10	60	780.85
BR07	72	414.71
FR01	84	724.97
Arithmetic mean	**48**	**686.24**
Median	**48**	**740.72**
Standard deviation	**25.65**	**208.74**

Table 4.22: Comparison between translation experience (above the mean value) and keyboard use (excluding BR01 and BR08)

Participant	Translation experience (months)	Total keyboard time (minutes)
FR08	144	848.90
BR05	156	623.50
BR06	192	746.84
FR09	216	310.81
BR04	240	722.68
FR10	252	717.15
BR09	360	296.24
Arithmetic mean	**222.85**	**609.44**
Median	**216**	**717.15**
Standard deviation	**72.61**	**219.13**

As indicated in tables 4.21 and 4.22, if the outliers (BR01, BR03 and BR08) are excluded, there seems to be a trend: the participants who had a higher level of experience tended to make less use of the keyboard than less experienced participants.

Regarding the values recorded for mouse use, it is possible to observe that some of the participants with translation experience of up to 132 months (the closest value to the mean, which is 139.33) had high mouse values, ranging between 1483.76 and 6184.58 seconds, except for FR05, who had a mouse value of 467.51 seconds, and BR10, who had a mouse value of 641.73 seconds. The remaining participants, with translation experience ranging from 144 months (FR08) to 360 months (BR09), had mouse values ranging between 1157.76 and 277.77 seconds, except for BR08, who had a mouse value of 4116.47 seconds. This suggests that, if FR05, BR10 and BR08 are excluded, the participants with levels of translation experience above the mean value of 139.33 had a lower average mouse value than the participants with lower levels of translation experience, as indicated in tables 4.23 and 4.24.

Table 4.23: Comparison between translation experience (below the mean value) and mouse use (excluding FR05 and BR10)

Participant	Translation experience (months)	Total mouse time (seconds)
FR04	0	6070.06
FR02	36	3627.83
FR07	36	6184.58
BR02	48	2817.73
BR07	72	3485.52
FR01	84	1483.76
BR03	132	1597.42
Arithmetic mean	**58.28**	**3609.55**
Median	**48**	**3485.52**
Standard deviation	**42.38**	**1910.28**

Table 4.24: Comparison between translation experience (above the mean value) and mouse use (excluding BR08)

Participant	Translation experience (months)	Total mouse time (seconds)
FR08	144	2326.08
BR05	156	2249.40
BR06	192	2345.92
BR01	192	1758.45
FR09	216	2777.77
BR04	240	1157.76
FR10	252	1834.68
BR09	360	2114.69
Arithmetic mean	**219**	**2070.59**
Median	**204**	**2182.04**
Standard deviation	**68.10**	**487.18**

As indicated in tables 4.23 and 4.24, if the outliers (FR05, BR10 and BR08) are excluded, there seems to be a trend: the participants who had a higher level of experience tended to make less use of the mouse than less experienced participants.

Finally, the same type of verification is performed here regarding the number of switches between keyboard and mouse. The participants with translation experience of up to 132 months (the closest value to the mean, which is 139.33) had totals of switches ranging from 124 (BR10) to 385 (BR02), except for FR01 and FR02, both of whom had a total of 106 switches. The remaining participants, with translation experience ranging from 144 months (FR08) to 360 months (BR09), had totals of switches ranging from 108 (FR09) to 268, except for BR08, who had a total of 299 switches. This suggests that, if FR01, FR02 and BR08 are excluded, the participants with translation experience above the mean value of 139.33 had a lower average value of switches than the participants with

translation experience above the mean value, as indicated in tables 4.25 and 4.26.

Table 4.25: Comparison between translation experience (below the mean value) and switches between keyboard and mouse (excluding FR01 and FR02)

Participant	Translation experience (months)	Switches between keyboard and mouse
FR04	0	279
FR07	36	321
FR05	48	126
BR02	48	385
BR10	60	124
BR07	72	191
BR03	132	239
Arithmetic mean	**56.57**	**237.85**
Median	**48**	**239**
Standard deviation	**40.22**	**98.25**

Table 4.26: Comparison between translation experience (above the mean value) and switches between keyboard and mouse (excluding BR08)

Participant	Translation experience (months)	Switches between keyboard and mouse
FR08	144	123
BR05	156	155
BR01	192	137
BR06	192	170
FR09	216	108
BR04	240	177
FR10	252	125
BR09	360	268
Arithmetic mean	**219**	**157.87**
Median	**204**	**146**
Standard deviation	**68.10**	**50.52**

As indicated in tables 4.25 and 4.26, if the outliers (FR01, FR02 and BR08) are excluded, there seems to be a trend: the participants who had a higher level of experience tended to make fewer switches than less experienced participants.

4.5. Productivity

In this section, extrapolated productivity values are calculated for all the participants, and this is compared with their levels of translation and PE experience, as well as with the totals of Essential Changes, Preferential Changes, Essential Changes Not Implemented and Introduced Errors.

4.5.1. Productivity values for both target languages

The next tables present the productivity values calculated for French and for Brazilian Portuguese, and then for both target languages combined.

Table 4.27: Productivity values for French

Participant	Trans. exper. (months)	PE exper. (months)	Total PE time (min.)	Essent. changes	Pref. changes	Essent. changes not impl.	Intr. errors	Keyb. time (sec.)	Mouse time (sec.)	Switches	PE productivity: words/hour	Extrapolated PE productivity per day (words/8 hours)
FR01	84	36	49	129	26	18	10	724.97	1483.76	106	1234.28	9874.24
FR02	36	0	75	101	54	49	12	756.47	3627.83	106	806.40	6451.20
FR04	0	0	105	131	46	22	7	470.84	6070.06	279	576	4608
FR05	48	6	64	109	95	24	14	1018.10	467.51	126	945	7560
FR07	36	3	122	135	47	11	4	832.63	6184.58	321	495.73	3965.84
FR08	144	36	65	123	38	24	6	848.90	2326.08	123	930.46	7443.69
FR09	216	12	120	130	51	21	8	310.81	2777.77	108	504	4032
FR10	252	5	75	135	52	9	9	717.15	1834.68	125	806.40	6451.20
Arithmetic mean	102	12.25	84.37	124.12	51.12	22.25	8.75	709.98	3096.53	161.75	787.28	6298.27
Median	66	5.5	75	129.5	49	21.5	8.5	740.72	2551.92	124	806.4	6451.2
Standard deviation	92.28	15.14	27.58	12.57	19.93	12.20	3.24	223.04	2086.02	86.47	255.17	2041.41

Table 4.28: Productivity values for Brazilian Portuguese

Participant	Trans. exper. (months)	PE exper. (months)	Total PE time (min.)	Essent. changes	Pref. changes	Essent. changes not impl.	Intr. errors	Keyb. time (sec.)	Mouse time (sec.)	Switches	PE productivity: words/hour	Extrapolated PE productivity per day (8 hours)
BR01	192	12	93	171	58	21	10	1320.27	1758.45	137	650.32	5202.58
BR02	48	12	60	174	28	31	10	491.41	2817.73	385	1008	8064
BR03	132	0	95	200	64	14	22	1964.58	1597.42	239	636.63	5093.05
BR04	240	36	61	149	58	53	28	722.68	1157.76	177	991.47	7931.76
BR05	156	36	53	168	50	41	10	623.50	2249.40	155	1141.13	9129.05
BR06	192	36	65	152	9	60	7	746.84	2345.92	170	930.46	7443.69
BR07	72	0	75	141	37	54	19	414.71	3485.32	191	806.40	6451.20
BR08	240	24	99	178	35	29	8	1182.63	4116.37	299	610.90	4887.27
BR09	360	0	46	152	27	58	8	296.24	2114.69	268	1314.78	10518.26
BR10	60	0	84	172	38	31	15	780.85	641.73	124	720	5760
Arithmetic mean	169.2	15.6	73.10	165.7	40.4	39.2	13.7	854.37	2228.47	214.5	881	7048.08
Median	174	12	70	169.5	37.5	36	10	734.76	2182.04	184	868.43	6947.44
Standard deviation	97.23	16.04	18.82	17.42	17.10	16.34	7.10	503.49	1043.58	82.56	236.63	1893.10

The next figures illustrate the comparison of translation and PE experience with the values calculated for extrapolated daily PE productivity. For these two figures only, the values in months were converted to days, so that it is easier to see the values in the figures.

Figure 4.33: Comparison between translation experience and extrapolated daily PE productivity

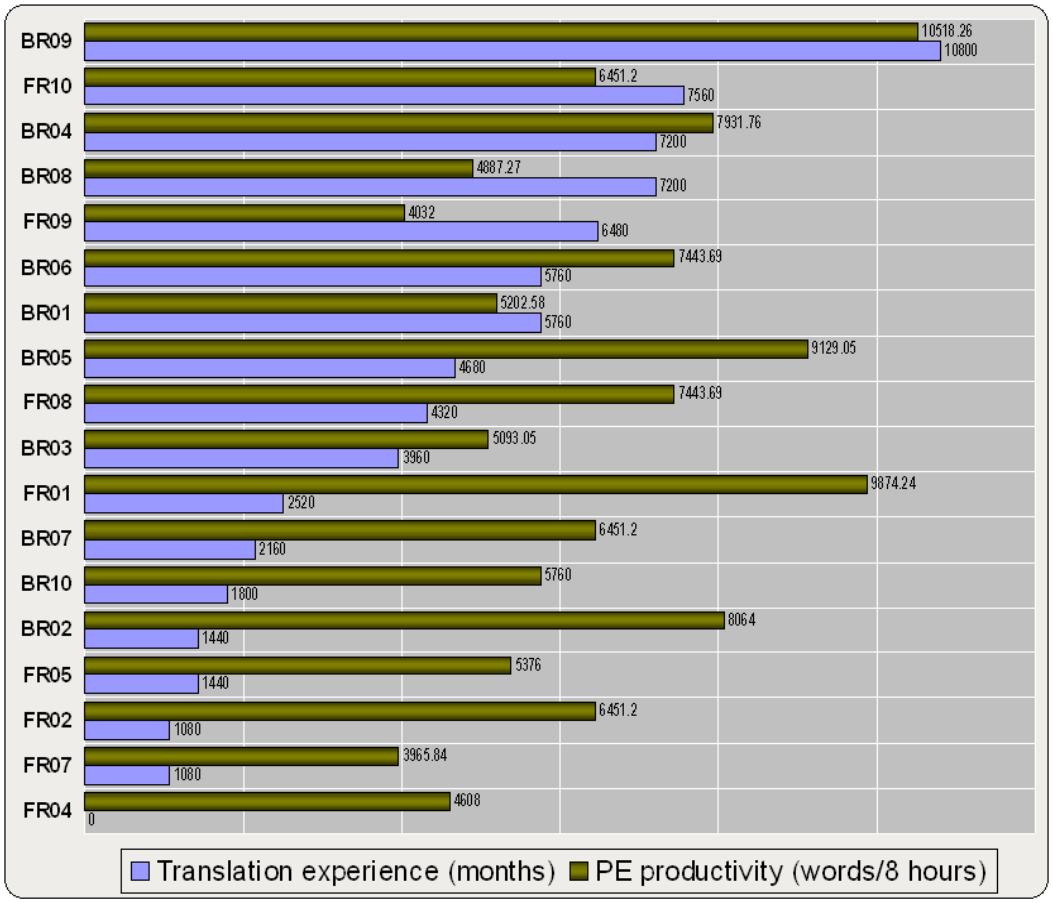
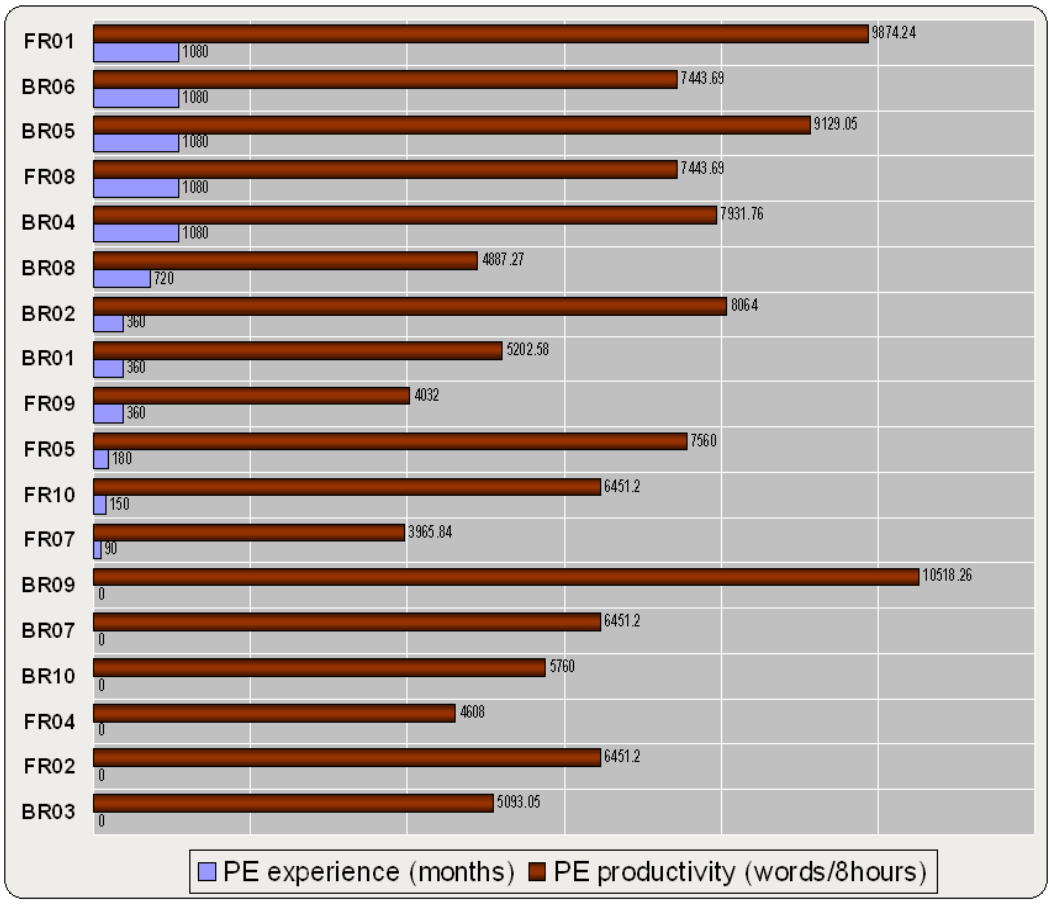


Figure 4.34: Comparison between PE experience and extrapolated daily PE productivity



It is interesting to observe that, as shown in Figure 4.33 (comparison between translation experience and PE productivity), BR09, the participant with the highest level of translation experience (30 years) is also one of the two participants with the highest extrapolated PE productivity (10,518.26 words/8 hours). FR01 also presents a high extrapolated productivity (9,874.24 words/8 hours), but a lower level of translation experience (7 years). However, as shown in Figure 4.34, FR01 is among the participants with the highest level of PE experience (3 years). It is possible to speculate that, in the case of BR09, the high level of translation experience influences the PE productivity, whereas for FR01

the combination of translation and PE experience influences the PE productivity.

The results presented in this section are consistent with the results presented in section 4.2.1, which indicated that there was no correlation between translation experience and PE time. In general, it is possible to see that increasing PE experience can lead to higher PE productivity. However, there are translators who can be highly productive post-editors even if they have not had any PE experience, such as BR09. A high level of translation experience can lead to high PE productivity, but does not necessarily do so, as can be seen with FR09, for example. Low levels of translation experience may also be associated with relatively lower PE productivity in some cases, as observed for FR04 and FR05.

4.6. Revision and Internet searches

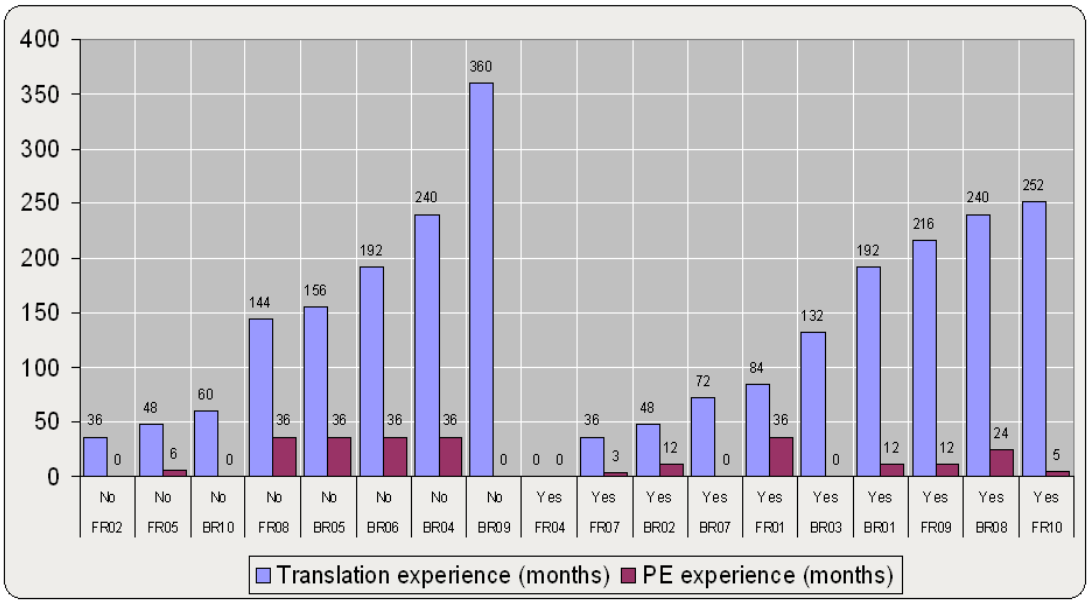
As previously mentioned, the participants were able to spontaneously decide if they wanted to revise their work at the end or not. Additionally, they were informed that they could perform Internet searches in case they wanted to clarify any doubts. This section investigates these two aspects in more detail.

Table 4.29 indicates which participants chose to revise their work at the end. For further clarification, details are also provided about their translation and PE experience, the total PE time and the main categories of the typology.

Table 4.29: Revision - French and Brazilian Portuguese

Participant	Revision at the end	Internet searches	Translation experience	PE experience	Total PE time (minutes)	Essential changes	Preferential changes	Essential changes not impl.	Introduced errors
FR01	Yes	No	84 months (7 years)	36 months (3 years)	49	129	26	18	10
FR02	Yes	Yes	36 months (3 years)	N/A	75	101	54	49	12
FR04	Yes	Yes	0	N/A	105	131	46	22	7
FR05	Yes	Yes	48 months (4 years)	6 months	64	109	95	24	14
FR07	Yes	Yes	36 months (3 years)	3 months	122	135	47	11	4
FR08	No	No	144 months (12 years)	36 months (3 years)	65	123	38	24	6
FR09	No	No	216 months (18 years)	12 months (1 year)	120	130	51	21	8
FR10	No	No	252 months (21 years)	5 months	75	135	52	9	9
BR01	Yes	Yes	192 months (16 years)	12 months (1 year)	93	171	58	21	10
BR02	Yes	Yes	48 months (4 years)	12 months (1 year)	60	174	28	31	10
BR03	Yes	Yes	132 months (11 years)	N/A	95	200	64	14	22
BR04	No	No	240 months (20 years)	36 months (3 years)	61	149	58	53	28
BR05	No	Yes	156 months (13 years)	36 months (3 years)	53	168	50	41	10
BR06	No	No	192 months (16 years)	36 months (3 years)	65	152	9	60	7
BR07	Yes	No	72 months (6 years)	N/A	75	141	37	54	19
BR08	Yes	No	240 months (20 years)	24 months (2 years)	99	178	35	29	8
BR09	No	No	360 months (30 years)	N/A	46	152	27	58	8
BR10	Yes	Yes	60 months (5 years)	N/A	84	172	38	31	15

Figure 4.35: Revision of the post-edited segments



Several observations can be drawn from the data presented in Table 4.29. Firstly, the decision to revise the work at the end does not seem to be linked with the level of translation experience, since a few translators with a lower level of experience (such as FR02, FR07, BR02 and BR10) revised their work, while some of the more experienced translators did not (such as FR08, FR09, FR10, BR04 and BR09). This trend seems to apply to the participants of both target languages. There were also exceptions in both cases. Most of the participants who did not revise their segments at the end had previous experience both with translation and with PE (FR05, FR08, BR05, BR06, and BR04). The other three participants who did not carry out a revision at the end (FR02, BR10, BR09) had experience with translation, but not with PE.

Internet searches to clarify doubts were performed mostly by the participants with a lower level of translation experience (FR02, FR04, FR07, BR02). Participants BR01, BR03, BR05 and BR10, who had a much higher level of translation experience, also performed Internet searches. However, there was a striking difference between these two groups, since the least experienced translators performed a high number of Internet searches to clarify doubts about the meaning of words and also to clear grammar doubts in some cases, while the most experienced translators only performed one Internet search each, to clarify the meaning of words (but not to clarify grammar doubts).

Another interesting aspect observed was that, amongst the participants who performed Internet searches, the most frequently visited website was www.google.com (including its regional links for Ireland, the

UK, France and Brazil). The participants who used Google's search engine to clarify doubts demonstrated some familiarity with Google's search operators, such as enclosing search terms in double quotes (in order to restrict a search). However, none of the participants used Google's advanced search feature nor its advanced search operators (such as "site:" for searching only in specific domains, "define:" to obtain definitions of terms or "translate [...] to" to use Google's MT engine directly from the search field). This was another common trend observed for both languages, and it also suggests that it would be useful to include information on advanced search techniques in PE and/or translation courses.

4.7. Quality and fitness for purpose

As mentioned in the Methodology chapter, the typology devised for the present research was used in a descriptive manner, in order to quantify and classify all the items recorded for each PE session (namely, Essential Changes, Preferential Changes, Essential Changes Not Implemented, and Introduced Errors). The objective of the typology was not to perform an assessment of the final quality of the segments post-edited by the participants (in other words, it was not used as a QA instrument to judge the segments). However, with the values recorded, it is possible to determine which participants achieved the best overall quality based on the following criteria:

- Highest number of Essential Changes
- Lowest number of Preferential Changes

- Lowest number of Essential Changes Not Implemented
- Lowest number of Introduced Errors

The above criteria would also be the desirable outcome for an overall good PE performance, if the definitions of a good post-editor and of fitness for purpose used in this research (as described in Chapter 3) are taken into account, and they should be combined with the lowest possible PE time.

Therefore, by following these criteria and by looking at the values presented in the previous sections, it is possible to identify the participants who could be considered as having achieved the best overall quality and the best PE performance. Tables 4.30 and 4.31 provide more details on this.

Table 4.30: Overall PE performance - French

Participant	Translation experience (months)	PE experience (months)	Extrapolated daily PE productivity	Essential changes	Preferential changes	Essential changes not impl.	Introduced errors
FR01	84	36	9874.24	129	26	18	10
FR02	36	0	6451.20	101	54	49	12
FR05	48	6	7560	109	95	24	14
FR08	144	36	7443.69	123	38	24	6
FR10	252	5	6451.20	135	52	9	9

Table 4.31: Overall PE performance - Brazilian Portuguese

Participant	Translation experience (months)	PE experience (months)	Extrapolated daily PE productivity	Essential changes	Preferential changes	Essential changes not impl.	Introduced errors
BR09	360	0	10518.26	152	27	58	8
BR05	156	36	9129.05	168	50	41	10
BR02	48	12	8064	174	28	31	10
BR04	240	36	7931.76	149	58	53	28
BR06	192	36	7443.69	152	9	60	7

Tables 4.30 and 4.31 include all the participants with a daily PE productivity above the mean values observed for each language (6298.27 words/8 hours for French and 7048.08 words/8 hours for Brazilian Portuguese). The highlighted cells provide the following indications:

- Yellow: values for **Essential Changes** that were **above** the mean value of 124.12 for French and 165.7 for Brazilian Portuguese.
- Blue: values for **Preferential Changes** that were **below** the mean value of 51.12 for French and 40.4 for Brazilian Portuguese.
- Pink: values for **Essential Changes Not Implemented** that were **below** the mean value of 22.25 for French and 39.2 for Brazilian Portuguese.
- Green: values for **Introduced Errors** that were **below** the mean value of 8.75 for French and 13.7 for Brazilian Portuguese.

If all of these factors are taken into account, FR01 may be considered the participant with the best PE performance for French, having the best balance for all the values combined, including extrapolated daily PE productivity. Following the same reasoning, the next best PE performance for French would correspond to FR08.

For Brazilian Portuguese, BR02 may be considered as the participant with the best PE performance, having the best balance for all the values combined, including extrapolated daily PE productivity. The next best PE performance for Brazilian Portuguese would correspond to BR09.

It is also interesting to review their answers to the questionnaire, which are thus summarised here. For French, FR01 considered MT output as having average quality (question 1, answer 3), considered MT useful for translators (question 2, answer 4) and expressed interest in PE (question 3, answer 4). FR08 considered MT output as having good quality (question 1, answer 4), considered MT useful for translators

(question 2, answer 4) and expressed strong interest in PE (question 3, answer 5). For Brazilian Portuguese, BR02 considered MT output as having good quality (question 1, answer 4), considered MT useful for translators (question 2, answer 4) and expressed interest in PE (question 3, answer 4). BR09 considered MT output as having bad quality (question 1, answer 1), considered MT moderately useful for translators (question 2, answer 3) and expressed interest in PE (question 3, answer 4).

These results seem very interesting, in that they signal that three of the four post-editors with the best PE performances expressed positive views regarding MT and PE. On the other hand, BR09, who had the second best PE performance for Brazilian Portuguese, expressed a negative view when answering question 1 and a moderate view when answering question 2. However, it is also relevant to observe that, unlike the other three participants discussed here, BR09 did not have previous PE experience and, at the same time, was the most experienced translator amongst all the participants for both target languages. A more in-depth study with a different setting would be required to investigate this trend, but the results found here seem to suggest a link between PE performance and the post-editors' bias regarding MT and PE. Also, as mentioned in section 4.3.2, the level of translation and PE experience of the participants might also have some influence on their views regarding PE.

Additionally, it is relevant to compare the results of the previous section with the performance of these three participants. FR01 revised the work at the end and did not perform Internet searches. FR08 did not revise

the work at the end and did not perform Internet searches. BR02 revised the work at the end and did a few Internet searches (although not as high a number as FR04, for instance). BR09 did not revise the work at the end and did not perform Internet searches. FR01 had seven years of experience as a translator and no experience with PE, FR08 had 12 years of experience as a translator and three years of experience with PE, BR02 had four years of experience as a translator and one year of experience with PE and BR09 had 30 years of experience as a translator and no experience with PE.

4.8. Comparison between the results of the pilot projects and the results of the main PE project

This section presents the results recorded according to the master and main categories from the typology in the two pilots and in the main project and the values recorded for total PE time. Additional observations are provided after the tables and figures.

There are two objectives when comparing the results of the three projects. First, it will be possible to determine if the same patterns emerge from the results of all three projects and if any similarities can be identified. Second, the comparisons will shed more light on the distribution of measures across the three experiments.

Appendix F provides tables with the breakdown of all of the items from the typology recorded for the main PE project. In this section, for brevity and clarity, the results are provided in a summarised format.

4.8.1. Comparison of the results for the master categories of the typology

This subsection discusses the results recorded for the master categories of the typology, namely, Essential Changes, Preferential Changes, Essential Changes Not Implemented and Introduced Errors. Table 4.32 presents the items from these categories recorded for all three projects. The values indicate the sum of all the items recorded for all the participants from each project. This is followed by the percentage (calculated in relation to the total number of items recorded).

Table 4.32: breakdown of items from the master categories of the typology recorded for the main PE project and the first and the second pilot projects

Project	Essential Changes	Preferential Changes	Essential Changes Not Impl.	Introduced Errors
First pilot project	336 (68.01%)	141 (28.54%)	22 (3.44%)	0 (0%)
Second pilot project	1032 (63.54%)	268 (16.5%)	300 (18.47%)	24 (1.47%)
Main PE project	2650 (62.5%)	813 (19.17%)	570 (13.44%)	207 (4.88%)

4.8.2. General comments

There is no significant discrepancy regarding the percentages of Essential Changes, Preferential Changes and Essential Changes Not Implemented between the second pilot and the main project. The percentage of Essential Changes recorded for the first pilot (68.01%) is slightly higher than the corresponding values for the second pilot (63.54%) and the main project (62.5%), but this cannot be considered as a highly discrepant value.

However, the other percentages of items from the master categories for the first pilot did not follow the same pattern observed in the second pilot and in the main project. The percentages of Preferential Changes (28.54%) and Essential Changes Not Implemented (3.44%) were higher, and no errors were introduced by the post-editors in the first pilot project. It is possible to speculate about the reasons for these differences. The word count used for the first pilot was the lowest one of the three projects carried out (as detailed in Chapter 3). Additionally, all the participants in the first pilot project belonged to translation teams who were working in an actual large-scale localisation project. All of them had already been working on the project for approximately two weeks by the time they took part in the PE experiment; therefore, they were acquainted with the project files (which is relevant, since segments from the project files were used to form the corpus of the first pilot). The familiarity of the participants with the project material and the reduced scope of the first pilot may be reasons for the difference in the percentages of the second

pilot and the main project, particularly the difference in the percentage of introduced errors recorded.

4.9. Similarities and differences in the data from the two target languages of the main PE project

This section examines in more detail the similarities and differences observed in the data collected for French and for Brazilian Portuguese for the main PE project. The objective of this investigation is to determine whether there were common trends between the two target languages, in order to answer the second research question:

Can the same PE strategies and trends be found across languages of the same family? (Test case: French and Brazilian Portuguese.)

Figures are presented to illustrate the results, followed by comments.

Figure 4.36: summary for all participants combined - French

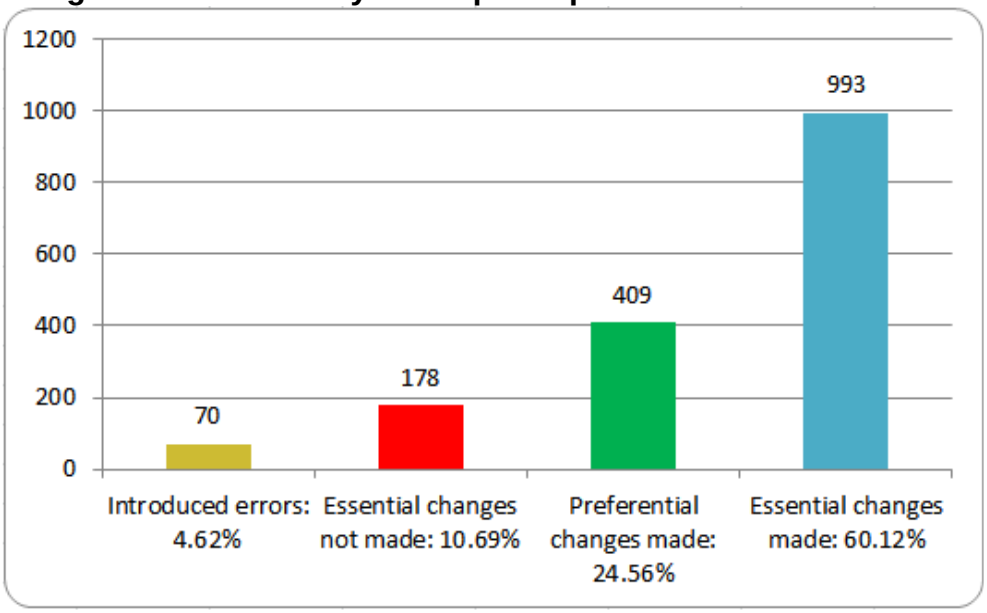


Figure 4.37: Total Essential Changes for French per category

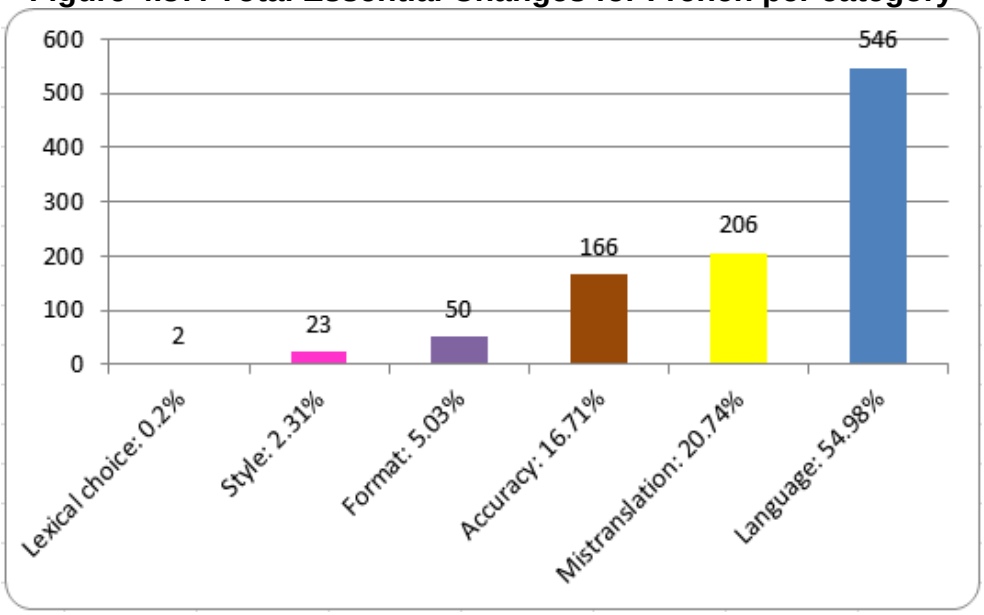


Figure 4.38: Total Preferential Changes for French per category

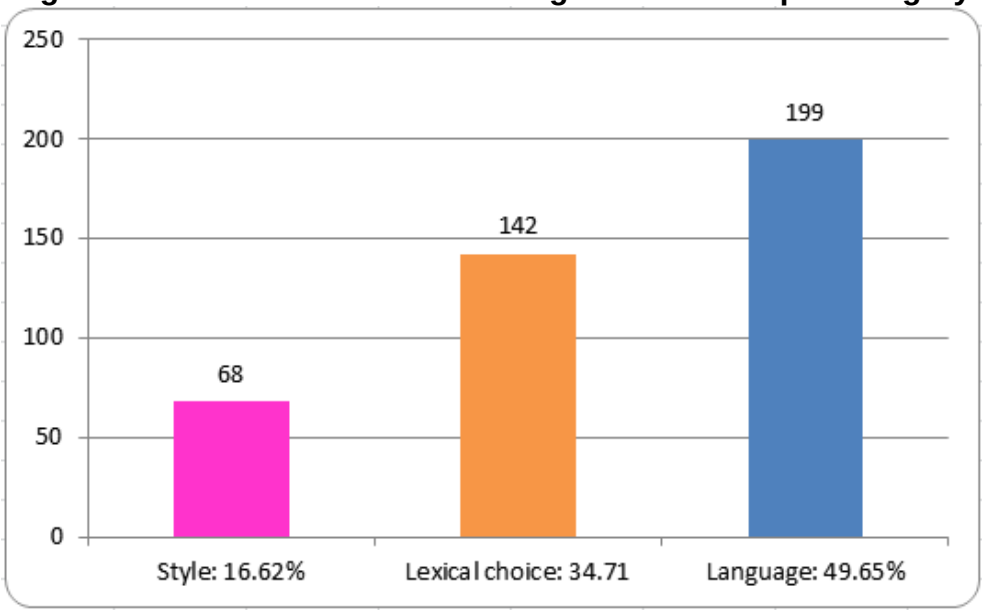


Figure 4.39: Total Essential Changes Not Implemented for French per category

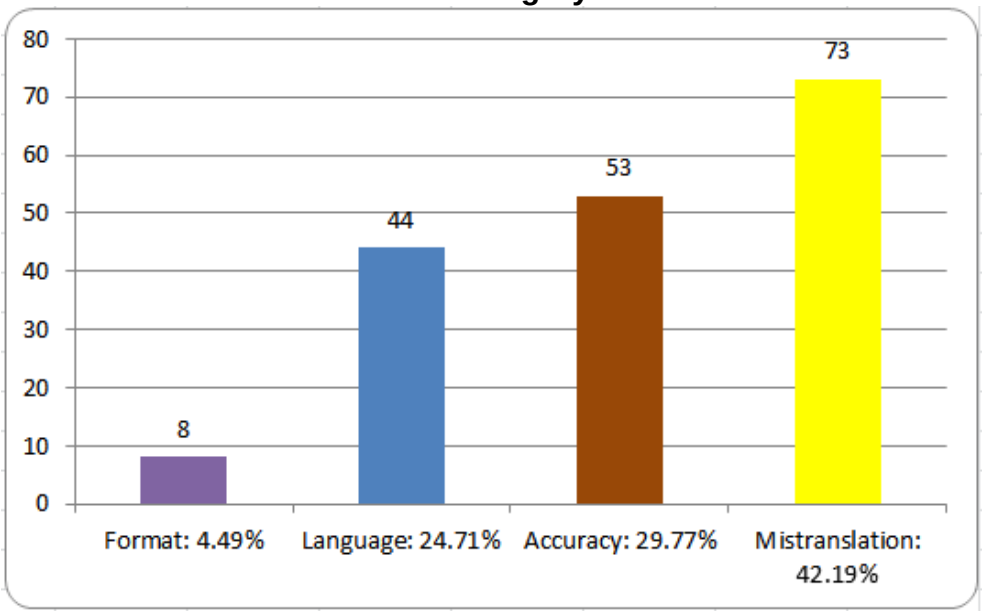
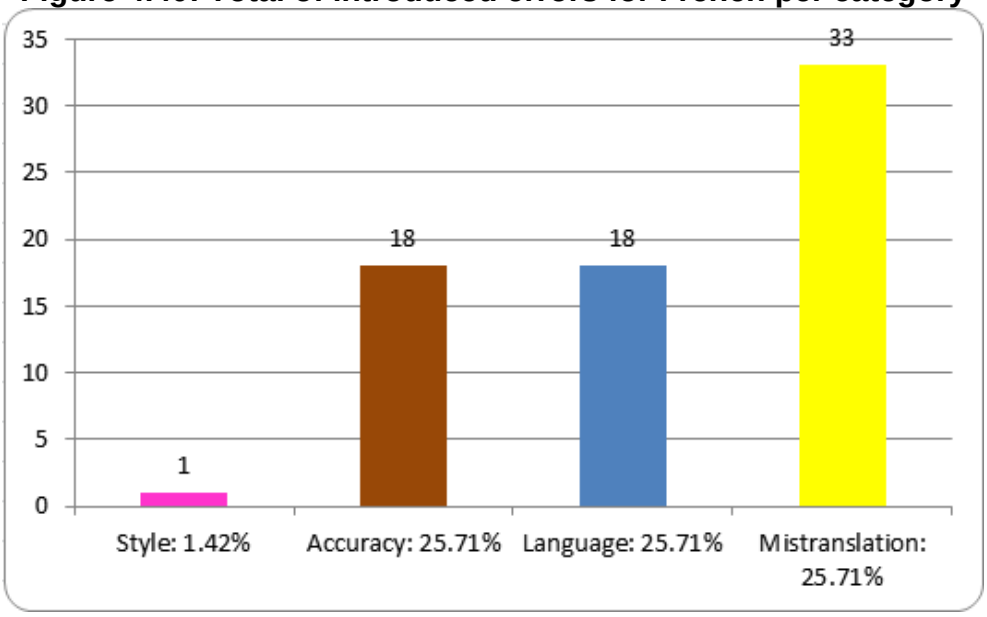


Figure 4.40: Total of introduced errors for French per category



4.9.1. Comments on the data for French

The master category with the highest number of items recorded is Essential changes, corresponding to 60.12% of the total of items recorded. Items in the Language category accounted for the majority of the Essential Changes (54.98%). This corresponds to the findings from the two pilot projects. The combined total of Essential Changes Not Implemented and Introduced Errors is much lower than the total of Essential Changes, accounting for 15.31% of all the items combined. This also corresponds to the findings from the two pilot projects. The total of Preferential Changes is also much lower than the total of Essential Changes, accounting for 24.56% of all the items recorded.

Figure 4.41: summary for all participants combined - Brazilian Portuguese

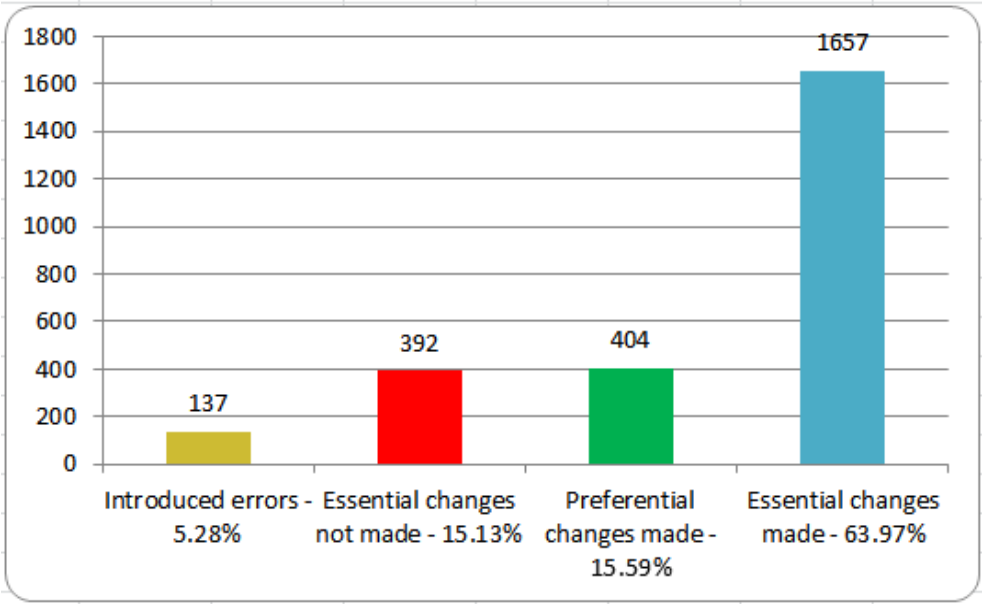


Figure 4.42: Total Essential Changes for Brazilian Portuguese per category

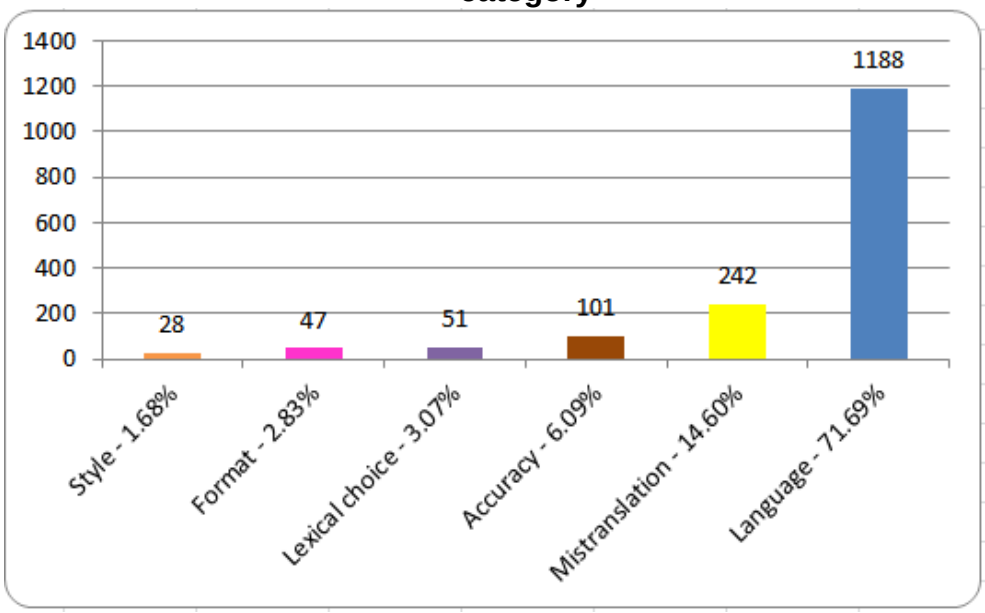


Figure 4.43: Total Preferential Changes for Brazilian Portuguese per category

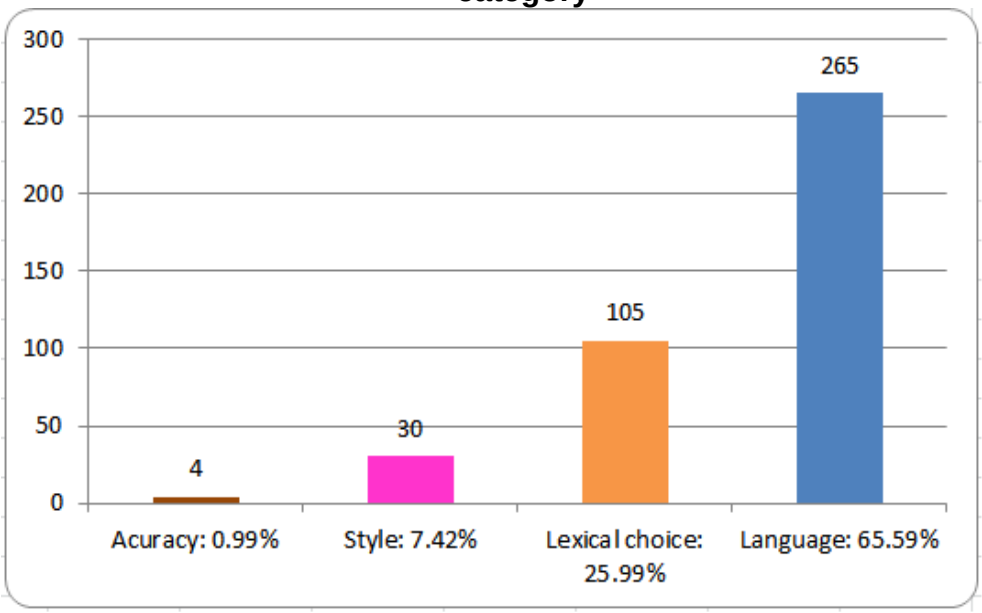


Figure 4.44: Total Essential Changes Not Implemented for Brazilian Portuguese per category

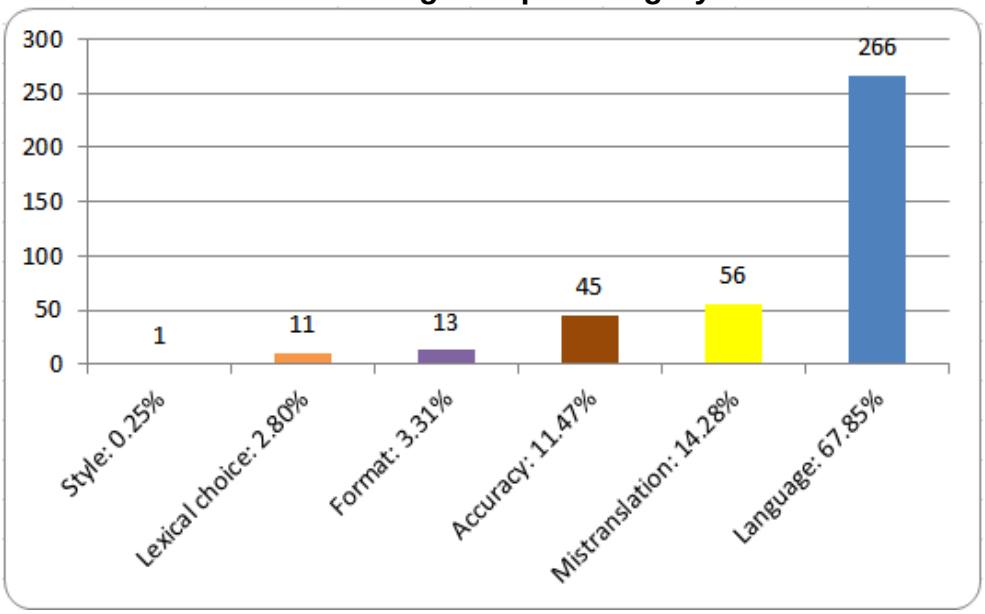
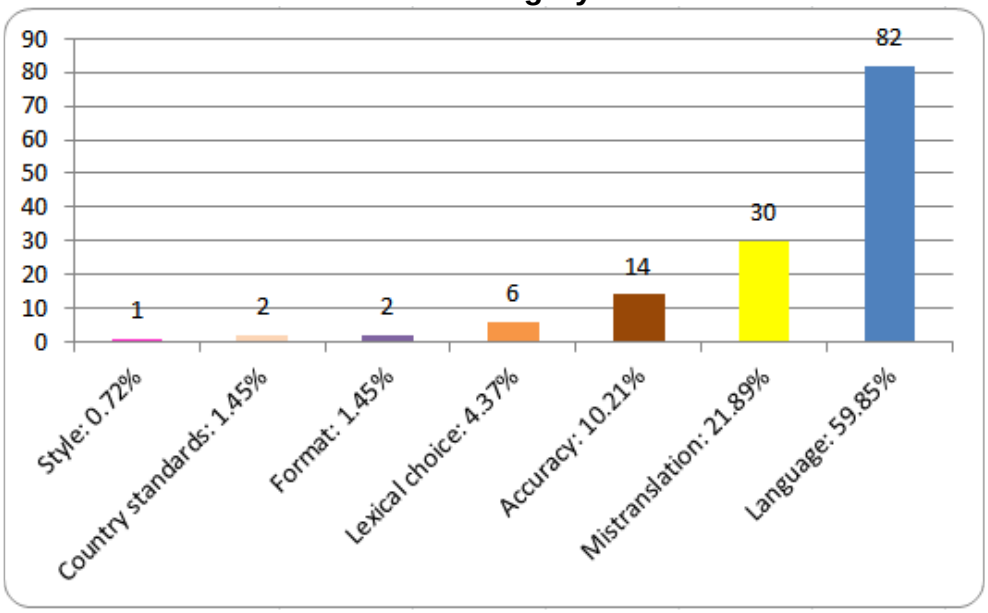


Figure 4.45: Total of introduced errors for Brazilian Portuguese per category



4.9.2. Comments on the data for Brazilian Portuguese

The master category with the highest number of items for Brazilian Portuguese is Essential Changes, corresponding to 63.97% of the total of items recorded. This corresponds to the findings for French in the main PE project (60.12%), and also to the findings from two pilot projects. Items in the Language category accounted for the majority of the Essential Changes (71.69%). Language was also the category with the highest number of Essential Changes for French in the main PE project (54.98%), and for the two pilot projects. The combined total of Essential Changes Not Implemented and Introduced Errors is much lower than the total of Essential Changes, accounting for 20.41% of all the items recorded. This also corresponds to the findings from the two pilot projects and the findings for French in the main PE project (15.31%).

The total of Preferential Changes is also much lower than the total of Essential Changes, accounting for 15.59% of all the items recorded. Again, this also corresponds to the findings for French.

The results therefore point to many similarities between the two target languages in the classification of PE edits according to the typology, which is relevant for our second research question.

4.10. Summary

This chapter provided a discussion of the data collected in the main PE project. Different aspects were analysed: comparisons between the levels of translation and PE experience of the participants and the categories and subcategories of the methodology, the total time taken to complete the task, the use of keyboard and mouse, the option to revise the segments or not at the end of the task, and the participants' answers to the questionnaire. The research questions were referenced during the discussion of these aspects.

Although the Pearson Product Moment Correlation did not indicate correlations between the levels of translation and PE experience and the total PE time, it is important to take into account other elements analysed here. As seen in the section about revision and Internet searches, other aspects played a part in the participants' PE performance. The section on quality and fitness for purpose provides a more complete picture of the results, indicating that BR02 presented the best overall PE performance in terms of all the aspects analysed, despite not having a high level of

translation experience. On the other hand, the two other participants with the best PE performance, BR06 and BR09, were very experienced translators, did not revise their work at the end and did not perform Internet searches (presumably indicating that they had no doubts to clarify). Putting all the pieces together, it is possible to conclude that individual characteristics may play a part in the PE performance, but there does seem to exist a complex relationship between experience and performance, which cannot be measured only by a comparison of the time taken for the task and the level of professional experience (hence the lack of correlations in the Pearson test). This is in tune with Krings' observations that there is more involved in PE effort than temporal measurement (Krings, 2001).

The next chapter, Conclusion, discusses the results in relation to the objectives of the present work, points out the limitations of the project and proposes suggestions for future research.

Chapter 5: Conclusion

5.1. Introduction

This chapter revisits the objectives of the present study, discusses the findings gathered and the limitations of the research, elaborates on the contributions to the literature and provides suggestions for further research.

5.2. Objectives of the study

The present research aimed to shed light on the PE process, establishing whether the previous level of translation and PE experience could have an effect on PE performance (particularly in terms of productivity and the types of changes made or omitted). The second objective was to investigate whether similar trends in the PE process could be observed between two languages of the same family (French and Brazilian Portuguese).

In order to answer the research questions, it was necessary to investigate several aspects. First, the comparison of the participants' productivity with their level of translation and PE experience and with the main categories of the typology. Second, the categories and subcategories of the typology that accounted for the highest number of items recorded in the pilot projects and the main project. Third, whether similar trends could be observed between the two target languages for all the aspects mentioned above. Fourth, the strategies employed by the post-editors when carrying out

a PE task (such as performing a final revision or not), preference for the keyboard or the mouse while editing (or switching frequently between both) and doing online searches to clarify the meaning of words and/or grammar doubts (or not). Lastly, the views of the participants regarding PE and MT (as expressed by their answers to the questionnaire), compared with their level of translation and PE experience and their PE performance.

In order to include all of the several perspectives that the research proposed to cover, the research design included qualitative and quantitative aspects. Quantitative methods included the use of a complex typology to quantify all the PE essential and preferential changes made, the essential changes not made and the errors introduced by the participants, and the calculation of statistical significance for the results and the development. Qualitative methods included screen recording and observation of the individual sessions, a survey to gather data about the participants' previous experience with translation and PE, and a questionnaire prior to the sessions to gauge the participants' views regarding MT and PE. The typology was also used as part of the qualitative methods, since it qualified the PE edits according to several categories and subcategories.

The researcher gathered and classified all the PE items recorded during the sessions according to the typology. The resulting data were analysed and compared, giving consideration to the participants' experience, their answers to the questionnaire, and the changes they implemented in the PE sessions.

5.3. Findings

The classification performed with the use of the typology pointed out that the majority of the items recorded for both target languages in the main PE project pertained to the main category Essential Changes. Next in order were Preferential Changes, Essential Changes Not Made and Introduced Errors. These results were in line with the values found in the two pilot projects.

Within the main category of Essential Changes Made, the subcategory with the highest number of items recorded for both languages was Language. Again, this was consistent with the findings of the two pilot projects. We can, therefore, suggest that, at least for languages from the same family (Romance languages, in the case of the present research), the types of changes made for similar content are alike.

No correlation was detected between the participants' levels of experience and the total time taken to complete the task. This result was found for the participants of both target languages. As discussed in Chapter 4, this is consistent with the findings of Guerberof (2012), who also conducted an experiment with a high number of participants. Furthermore, it signals that PE effort and PE performance involve a high level of complexity that cannot be explained only by analysing temporal values.

The usage of input methods by the participants was analysed, and although no correlation was detected when drawing comparisons with all the participants, it was possible to observe interesting trends amongst clusters of

participants regarding their levels of translation and PE experience and keyboard and mouse usage, as well as the number of switches between the two. This also held true for both target languages.

The PE productivity of the participants was calculated for the task, and their daily productivity was extrapolated from these values, firstly just taking into account the total time taken for each session. Next, the productivity was examined in light of the strategies adopted by the participants, namely, carrying out a revision at the end or not and doing Web searches or not. Interesting findings resulted from these observations: for both target languages: the most experienced participants (in terms of translation and PE) performed little to no Internet searches, while the least experienced participants performed the highest number of searches overall; the searches performed by the least experienced participants included terms that cannot be considered as highly specialised (such as "factory") and grammar doubts. There was no strong correlation between the level of translation and PE experience and the tendency of participants to revise their work at the end or not. This result was found for both target languages.

Taking all of these aspects into account, it was possible to identify the participants who could be considered as having presented the best PE performance, in terms of productivity, adherence to the PE guidelines provided for the task, fitness for purpose and quality. The participants who presented the best PE performance had both previous translation and PE experience and performed few or no searches on the Web (presumably as a

result of their previous professional experience). The participant with the best PE performance overall revised the work at the end, but the two other participants with the best performance did not.

5.4. Limitations of the present research

5.4.1. Language pairs

This study involved one source language (English) and two target languages (French and Brazilian Portuguese) for the main PE project. These language pairs were chosen based on market size (taking into account that many localisation projects that include MT and PE tend to have English as a source language, and localisation into French and Brazilian Portuguese also represents a big market share), the researcher's knowledge and experience with the languages employed and the logistics and availability of corpora and potential participants for the language pairs involved. Comparing the results and findings for two target languages of the same language family presented many advantages, such as determining if the possible similarities identified in the PE process could be used for developing PE guidelines and training for Romance languages, with minimal customisations per language required (such as adding specific examples for each language, but using the same general structure for guidelines aimed at different Romance languages).

However, it would have been beneficial to have one or more target languages from a different language family, or one or more source languages other than English. In this case, it would have been possible to investigate if similar results could have been found across different language families, and if the findings presented here could be generalised for post-editors working with different language combinations. Moreover, it might also have been possible to take cultural differences into account for the analysis (although this was outside the scope of the present research). Tatsumi's study (2010), which has points in common with the present research, included a different language combination (English and Japanese). Although part of the trends identified in her research are common to the present study, some of her findings for the post-edited text in Japanese might not have an exact correspondence with the findings reported for the target languages included in the present research. For instance, the polite form of sentence endings in Japanese may be linked to Japanese-specific PE edits that would not apply to Romance languages.

5.4.2. Participants

While an attempt was made to include participants with different levels of translation and PE experience (ten for Brazilian Portuguese and eight for French), it would have been beneficial to have a higher number of participants to provide more data for the analysis.

Originally, the same number of participants was planned for both languages, and this was indeed achieved, but it was not possible to use the data from two of the French participants (FR03 and FR06), as explained in Chapter 3. For the sake of consistency and balance, it would have been beneficial to have the same number of participants for both languages.

5.4.3. Typology

Some difficulties were encountered when using the typology to classify PE items. Since the typology is complex and includes categories and subcategories, the time required for the classification can be very lengthy. It is not possible to carry out such a classification in just a few days (also taking into account the number of segments and participants involved in the present research). Additionally, since there can be overlapping classifications, as explained in section 3.4.3 of Chapter 3, the classification process requires full attention and concentration at all times, to make sure that nothing is left out. The researcher's familiarity with the LISA QA Model (thanks to years of experience with QAs as part of localisation projects), coupled with considerable experience with PE, including quality assessment of PE projects that involved classifications (albeit with fewer categories) contributed to ensure that the classification was adequately carried out. Nevertheless, it was a lengthy and difficult process.

5.4.4. Metrics

It would have been helpful to include metrics such as GTM and TER, as this could have provided additional insights, such as the assessment of MT quality. However, due to the unavailability of reference human translations for all the segments, these metrics were not used for the main PE project.

5.4.5. Setup

As much as possible, the present research tried to reproduce real work conditions for the PE experiments by allowing the participants to work from their own usual locations (regardless of geographical distance in relation to the researcher), providing guidelines that reflected the instructions normally provided for similar projects, and allowing the participants to make their own choices regarding final revision and Internet searches. However, although less intrusive and artificial than the use of a fully controlled environment with strict limitations (and which possibly would also require the participants to be present at the location chosen for the experiment), this setting was still not exactly identical to that of a real project, due to practical limitations. The participants were aware that their actions onscreen were being recorded and remotely observed by the researcher. This interference may have had an effect on some of the participants, who perhaps felt less at ease than they would when working under normal conditions. If , instead, the participants had worked only on their own local machines, under familiar conditions, knowing that their actions onscreen were not being observed or recorded, some of

them might have been able to focus even more on the task and might have made fewer mistakes, might have been faster or might have achieved a better overall PE performance for other reasons. In order to verify these assumptions, a different test setting would have been required, which was outside the scope of the present research.

5.4.6. PE environment

The PE environment used for the main PE project, Autodesk's Post-editing Workbench, was well-suited for the setup devised for the main experiment. It allowed the participants to carry out the PE task remotely, in a browser window opened in the researcher's screen. The use of the workbench was straightforward and intuitive, and did not require prior training (just a few very simple instructions, sent by the researcher via e-mail before the sessions along with the PE guidelines), which was advantageous. In a different setup, it could have been beneficial to use a more widely adopted tool, such as SDL Trados; however, although this might have provided an extra dimension to the analysis, it was outside the intended scope of the present research.

5.4.7. Questionnaire

Before each PE session, the participants were asked to fill out a short questionnaire to gather insights about their views regarding PE and MT. The

questionnaire was intentionally kept short, because it would be answered in addition to the PE sessions, and some of the participants might have limited time availability and might be discouraged to take part in the experiment if it involved long instructions and/or surveys in addition to the PE task. The difficulty of finding a sufficient number of participants for both target languages, and the large amount of data for the analysis generated by the classification with the typology alone were also taken into account. However, a longer questionnaire and/or retrospective interviews might have also usefully informed the analysis.

5.5. Contribution to the field

The present research offers the following contributions to the field of MT, Post-Editing and Translation research:

Instead of being conducted in a controlled setup, the main PE project was designed to mirror real-life conditions as much as possible, in order to enable the researcher to observe and gather information about participants with different levels of translation and PE experience working on a PE task in real time. The segments used for the PE task were from actual software documentation, the instructions were similar to those provided in PE projects from real life, and the fact that the task was performed remotely also mirrored a method of work that has become increasingly common in recent years for localisation projects, with the adoption of TMs hosted on remote servers by

many companies. In that respect, even Autodesk's workbench can be described as a reflection of real work conditions, since many companies today use proprietary environments for localisation tasks. Together, all of these factors enabled the collection of data that reflect real work conditions and can improve understanding of the PE process as a whole and of some of the difficulties faced by post-editors.

With the use of a complex typology based on an expanded version of the LISA QA Model (also reflecting another method employed in real projects), insights were gathered about the most common changes by post-editors for two target languages. The classification also allowed the identification of common changes not made, preferential changes and errors introduced by the post-editors themselves. This information could contribute towards the design of PE guidelines, which could provide additional examples and instructions highlighting the corrections to be implemented or not (also taking into account the level of PE required for particular projects). The typology itself is also a contribution to the research field.

It was useful to use the LISA QA Model as the basis for the typology, since it demonstrated that this model can be adapted to classify and quantify changes involved in PE projects. The present research used a complex typology to obtain a high level of detail about the PE process, for academic purposes. However, for industrial settings, it can be suggested that it would be possible to use the LISA QA Model with only small adaptations or even in its standard form to classify and quantify PE changes.

Insights were gathered about different strategies employed by the post-editors (such as revising at the end, performing Internet searches, favouring the keyboard or the mouse), as well as the overall PE performance of the participants. These elements were analysed taking into account the profile of a good post-editor proposed in Chapter 3. Loffler-Laurian (1996: 83) suggests that there is a psychological aspect in the reactions of post-editors to the raw MT output, and that professionals who show openness and interest in exploring new activities could be well-suited to perform PE work. These aspects are related to the post-editors' views on PE work, and the results of the present research indicate that there may be indeed a connection between PE performance and the post-editors' opinions regarding MT and PE (Chapter 4, section 4.7). Additionally, the insights gathered in our research suggest that, while there does not seem to be a direct link between translation experience and PE performance, having some level of previous translation experience does seem to be beneficial, as such post-editors may not need to clarify many doubts (if any), may have a more structured approach for dealing with the task (for instance, post-editing the segments sequentially and leaving any necessary corrections to the end, not to break the flow of the work), may already be used to revising their work at the end or may be very proficient, thus minimising the number of corrections left to be implemented at the end. These considerations also suggest that it may be a good idea to tailor PE training according to the level of experience of the participants. PE training may also play an important task for the development

of a skill identified by Offersgaard et al. (2008: 153): the capacity to decide quickly whether a machine-translated segment can be useful (and, therefore, whether it should be post-edited) or whether it should be ignored (and the source segment should be translated from scratch instead). This skill, which is very relevant for PE work, can be developed through practical experience, but if potential post-editors receive training that includes this aspect, they may be better equipped to tackle PE projects.

Although many studies have already been carried out on the topic of PE, they often seem to focus on how to improve metrics, the quality of the raw MT output, or other aspects related to the use of PE in industrial settings. All of these areas are extremely useful and deserve further research; however, the present study aimed to focus instead on the PE process taking into account the perspective of the post-editors. The researcher herself, having extensive experience with software localisation, technical translation and post-editing, has participated in many PE projects that lacked clear guidelines and examples, and has observed from interactions with colleagues over the years that, while the demand for PE has been growing, there is still a lack of training for this activity, and many language professionals still have doubts and even misconceptions about it. Therefore, the present research ultimately aimed to address these issues by first seeking to better understand the PE process itself. The analysis highlighted differences among the participants with varied levels of translation experience and how they approached the PE task. The many insights gathered will be used for the development of a PE

training course, which the researcher has begun to design in parallel with the present research. Some of the areas to be covered, as a direct result from the observation of the participants who took part in the pilot projects and in the main PE project, are practical suggestions and tips to make more efficient use of input methods (keyboard and mouse – with the possibility of including touchscreen and motion-sensing technology in the future); practical tips for optimising the PE task (such as suggestions for the order in which to check the source text and the raw MT output); suggestions on how to optimise the checking of long segments to ensure that nothing is overlooked; suggestions for optimising a PE task as a whole (such as writing down doubts to check them at the end); minimising the time required to do a final revision by taking steps to ensure that as many corrections as possible are covered during the PE phase; other practical suggestions for dealing with PE projects, such as productivity tips; a thorough explanation of the different levels of PE and what each of them entails; an explanation of MT technology, its limitations and its advantages; practical tips for working on projects that involve both MT and TM matches (which are becoming increasingly common in the localisation market); and an introduction to other types of MT-related work that language professionals can perform, such as maintenance of dictionaries and quality assessment of PE projects.

It is hoped that the knowledge gathered will be useful for the industrial partner who sponsored the present research, VistaTEC, for the future improvement of their PE projects, and that it will also be helpful for further

research on PE, for the development and/or improvement of effective PE guidelines and for the improvement of the PE and MT workflow both for contractors and post-editors. In addition, and very importantly, it is hoped that the PE training course to be developed as a direct outcome of this study can help language professionals to better equip themselves for performing PE work, if they so choose, by gaining a better understanding of the PE process as a whole and by learning practical suggestions to make them more productive and efficient post-editors.

5.6. Suggestions for further research

Several future avenues of research can be derived from the present research. The first suggestion would be to conduct a similar PE study including a higher number of participants and different language pairs. This could provide useful insights on whether the same findings could apply to languages of other families, and the knowledge gathered could also help develop and improve PE guidelines and training.

It would be extremely useful to conduct a similar experiment with the inclusion of eye tracking. The project setting would have to be modified, in that the participants would be required to be physically present at the location chosen for the experiment. Still, despite any limitations that this would entail, eye-tracking technology would add a very interesting dimension to the research, and would allow an in-depth investigation of PE strategies pointed

out here, as well as further insights about differences between post-editors with different levels of translation and PE experience, which ultimately could also be helpful for the formulation of guidelines and training.

Although the participants of the main project in the present research expressed mostly positive or neutral views regarding MT and PE, other studies (such as Fulford 2002, Araújo 2004 and Benedetti et al. 2004), as well as the researcher's exchanges with colleagues over the years, suggest that the overall population of translators may still have reservations regarding these technologies or, in some cases, strong negative views. It would be interesting to conduct an experiment similar to the main PE project from the present research that would include two groups of participants – one with extremely negative views on PE and one with very positive opinions about it – and to investigate whether the participants' very strong bias would have a link with their PE performance. This was one of the areas that the present research tried to investigate, but a larger-scale experiment with participants expressing much more radical views could further elucidate this aspect.

For the present research, the participants were instructed to employ an intermediate level of PE for the segments (as explained in section 3.7.4). However, actual PE projects require different levels, ranging from minimal to full PE, according to the target audience and to the purpose of the final text. It would be very interesting to conduct an experiment in which the participants would be asked to employ different levels of PE, in order to investigate the difficulties posed by each level, and the strategies adopted by the participants

for dealing with them. This experiment could include eye-tracking, questionnaires and retrospective interviews, as well as different language pairs.

Finally, since the findings of the present research will be used as the basis for the development of a PE training course, another suggestion for further research would be to compare the post-editors' views on PE and MT and their performance (in terms of productivity, adherence to guidelines, fitness for purpose, types of corrections made or not made and introduced errors, if any) both before and after receiving comprehensive PE training. It would be interesting to do such an experiment with a high number of participants, a larger corpus, different language pairs and the use of eye-tracking, screen recording, keylogging, questionnaires and retrospective interviews. Regarding the number of participants, Bertaux (1981, p. 35) recommends fifteen as the smallest acceptable number, while Morse (1994, p. 225) recommends 30 to 50 participants. Taking these recommendations into account, we suggest that it would be helpful to have at least 30 participants. Regarding the corpus, it would also be helpful to have at least twice as many words as the total word count used in our main experiment. With a higher number of participants and a larger corpus, it would be possible to collect much more data, facilitating a more comprehensive analysis of the PE process.

5.7. Final remarks

The present research has contributed to the existing literature on PE by investigating the PE process taking into consideration the perspective of the post-editors, and by determining areas of difficulty that should be contemplated when formulating PE guidelines and training. Common trends were identified in the two target languages, and similarities and differences between post-editors with different levels of professional experience were analysed. The knowledge gathered can have different applications, not only for improving the PE process in industrial settings, but also for making PE tasks more efficient, less tedious and more gratifying for post-editors.

List of references

Aikawa, T., K. Yamamoto and H. Isahara. 2012. *The impact of crowdsourcing post-editing with the collaborative translation framework*. [Online]. Microsoft Research. Available from: http://research.microsoft.com/pubs/172592/JapTal2012.pdf [Accessed October 2012].

Alfaro, C. 1997. *Descobrindo, compreendendo e analisando a tradução automática*. Monograph. Specialisation course in English-Portuguese Translation, Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, RJ, Brazil.

Alfaro, C. and M. C. Dias. 1998. Sistemas de tradução por máquina: ferramentas de auxílio à tradução. IN: *VII encontro nacional de tradutores e I encontro internacional de tradutores. USP/ABRAPT/CITRAT*, São Paulo, SP, Brazil, September 11.

Allen, J. 2001. Postediting: an integrated part of a translation software program. IN: *Language International – April 2001*. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 26-29.

Allen, J. 2003. Post-editing. IN: Somers, H. L. (ed.), *Computers and translation: a translator's guide*. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 297-316.

Allen, J. 2004. Case study: implementing MT for the translation of pre-sales marketing and post-sales software deployment documentation at Mycom International. IN: *Proceedings of the 6th conference of the Association for Machine Translation in the Americas, AMTA*. Washington, DC, USA, September 28–October 2, pp. 1-6.

Ambati, V., S. Vogel and J. Carbonell. 2012. Collaborative workflow for crowdsourcing translation. IN: *Proceedings of the ACM 2012 conference on computer supported cooperative work*. Seattle, WA, USA, February 11–15, pp. 1191-1194.

Ananthakrishnan, R., P. Bhattacharyya, M. Sasikumar and R. M. Shah. 2007. Some issues in automatic evaluation of English-Hindi MT: more blues for BLEU. IN: *Proceedings of ICON 2007 – 5th International Conference on Natural Language Processing, IIIT*. Hyderabad, India, January 4–6 (no page numbers).

Araújo, J. P. de. 2004. Tradutores experientes falam sobre a tradução automática. [Online]. In: *Comunicar*. Available from: http://www.comunicar.pro.br/artigos/tradut_exp.pdf [Accessed October 2012].

Arnold, D., L. Sadler and R. Lee Humphreys. 1993. Evaluation: an assessment. In: *Machine Translation*. Dordrecht/Boston/London: Kluwer Academic Publishers. 8 (1–2), pp. 1-24.

Ashley, C. W. 1993. *The Ashley book of knots*. London, United Kingdom: Faber and Faber.

Aymerich, J. and H. Camelo. 2009. The MT Maturity Model at PAHO. IN: *Proceedings of the 12th MT summit.* Ottawa, ON, Canada, August 26–31, pp. 403-409.

Benedetti, I. C. and A. Sobral (eds.). 2004. *Conversas com tradutores: balanços e perspectivas da tradução*. São Paulo, SP, Brazil: Parábola Editorial.

Bertaux, D. 1981. From the life-history approach to the transformation of sociological practice. IN: Bertaux, D. (Ed.), *Biography and society: The life history approach in the social sciences*. London, United Kingdom: Sage, pp.29-45.

Brosnan, M. J. 1998. *Technophobia: the psychological impact of information technology*. London, New York: Routledge, p. 10.

Canadian International Trade Tribunal. 2001. [Online]. Available from: http://www.citt.gc.ca/procure/determin/pr2b029_e.asp#P90_10741 [Accessed February 2009].

Carl, M. and A. Way (eds.). 2003. *Recent advances in example-based machine translation*. Dordrecht/Boston/London: Kluwer Academic Publishers.

Carl, M, B. Dragsted, J. Elming, D. Hardt and A. L. Jakobsen. 2011. The process of post-editing: a pilot study. In Sharp, B., Zock, M., Carl, M. and Jakobsen, A. L. (eds), *Proceedings of the 8th International NLPSC workshop. Special theme: human-machine interaction in translation.* Copenhagen Business School, Copenhagen, Denmark, August 20-21, pp. 153-163.

Champollion, Y. 2001. Machine translation (MT) and the future of the translation industry. [Online]. IN: *Translation Journal*, Volume 5, No. 1, January 2001. Available from: http://translationjournal.net/journal/15mt.htm [Accessed October 2012].

Cieri, C., S. Strassel, M. Lammie Glenn and L. Friedman. 2007. Linguistic resources in support of various evaluation metrics (presentation slides). IN: *MT Summit XI, workshop on automatic procedures in MT evaluation*. Copenhagen, Denmark, September 9–14 (no page numbers).

Clarke, R. H. 1996. Public works – engineering in local government. London, United Kingdom: Thomas Telford Publishing.

Craciunescu, O., C. Gerding-Salas and S. Stringer-O'Keeffe. 2004. Machine translation and computer-assisted translation: a new way of translating? [Online]. IN: *Translation Journal, Volume 8, No. 3, July 2004*. Available from: http://translationjournal.net/journal/29computers.htm [Accessed October 2012].

Creswell, J. W. 2003. *Research design: qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA, USA: Sage Publications.

DePalma, D. A., B. B. Sargent and R. M. Powers. 2007. *On the Web, some countries matter more than others. Quantifying the market opportunity for globalizing the Web customer experience*. [Online]. Common Sense Advisory, MA, USA. Available from:

http://www.commonsenseadvisory.com/Portals/_default/Knowledgebase/ArticleImages/070925_R_gdp_analysis_Preview.pdf [Accessed October 2012].

DePalma, D. A. 2010. *SDL Acquires Language Weaver*. [Online]. Common Sense Advisory, MA, USA. Available from: http://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=675&moduleId=391 [Accessed January 2013].

Depraetere, I. 2010. What counts as useful advice in a university post-editing training context? Report on a case study. IN: *Proceedings of the 14th annual EAMT conference*. St. Raphael, France, May 27–28.

Doherty, S. 2012. *Investigating the effects of controlled language on the reading and comprehension of machine translated texts: a mixed-methods approach*. PhD Thesis. School of Applied Language and Intercultural Studies, Dublin City University, Dublin, Ireland.

Doherty, S., D. Kenny and A. Way. 2012. Taking statistical machine translation to the student translator. IN: *Proceedings of the 10th conference of the Association for Machine Translation in the Americas, AMTA*. San Diego, CA, USA, October 28–November 1.

Du, J., Y. He, S. Penkale and A. Way. 2009. MaTrEx: the DCU MT System for WMT 2009. IN: *Proceedings of the Fourth Workshop on Statistical Machine Translation, EACL 2009*. Athens, Greece, March 30-31.

Dunne, K. 2009. Assessing software localization - For a valid approach. IN: Angelelli, C. V. and Jacobson, H. E. (eds.), *Testing and Assessment in Translation and Interpreting Studies: A Call For Dialogue Between Research and Practice.* Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 200.

Editing skills for computer translation. 2009. [Online]. IN: *The Chinese University of Hong Kong, Department of Translation*. Available from: http://traserver.tra.cuhk.edu.hk/pgcourses/outline/TRAN6602.pdf [Accessed October 2012].

Elliott, D., A. Hartley and E. Atwell. 2004. A fluency error categorization scheme to guide automated machine translation evaluation. IN: *Proceedings of the 6th conference of the Association for Machine Translation in the Americas, AMTA*. Washington, DC, USA, September 28–October 2, p. 66.

Flanagan, M. A. 1994. Error classification for MT evaluation. IN: *Proceedings of the 1st conference of the Association for Machine Translation in the Americas, AMTA.* Columbia, MD, USA, October 5–8, pp. 65-72.

Flickinger, D., I. Sag, J. Nerbonne and T. Wasow. 1987. Toward evaluation of NLP systems. Unpublished. IN: *Forum for the Association of Computational Linguistics*, Stanford University, Palo Alto, CA, USA, July 6.

Font Llitjós, A. 2007. *Automatic improvement of machine translation systems*. PhD Thesis. Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA.

Frey, L. R., C. H. Botan and G. L. Kreps. 1991. *Investigating communication – an introduction to research methods*. Englewood Cliffs, NJ, USA: Prentice Hall.

Fulford, H. 2002. Freelance translators and machine translation: an investigation of perceptions, uptake, experience and training needs. IN: *6th EAMT workshop teaching machine translation*. Manchester, United Kingdom, November 14–15, 2002, pp. 117-122.

García, I. 2009. Beyond translation memory: computers and the professional translator. IN: *Journal of Specialised Translation*, Issue 12. Available from: http://www.jostrans.org/issue12/art_garcia.pdf [Accessed October 2012].

García, I. 2010. Is machine translation ready yet? IN: *Target*, 22*, 1*, pp. 7-21.

Gow, F. 2003. *Metrics for evaluating translation memory software*. MA Thesis. School of Translation and Interpretation, University of Ottawa, ON, Canada.

Guerberof Arenas, A. 2008. *Productivity and quality in the post-editing of outputs from translation memories and machine translation*. Diploma of Advanced Studies thesis. Universitat Rovira i Virgili, Catalonia, Spain.

Guerberof Arenas, A. 2012. *Productivity and quality in the post-editing of outputs from translation memories and machine translation*. PhD Thesis. Universitat Rovira i Virgili, Catalonia, Spain.

Guerberof Arenas, A. 2012. *Post-editing MT output/Posedición de traducción automática*. [Online]. Colegio de Traductores Públicos de la Ciudad de Buenos Aires, Argentina. Available from: http://www.traductores.org.ar/nuevo_org/home/cursos_presenciales/?id_ruta=6&nivel2=526&nivel3=130&nivel4=131&id_curso=1020&mes=05%ano=2010 [Accessed October 2012].

Guerra Martínez, L. 2003. *Human translation versus machine translation and full post-editing of raw machine translation output*. MA Thesis. School of Applied Language and Intercultural Studies, Dublin City University, Dublin, Ireland.

Guzmán, R. 2007. Automating MT post-editing using regular expressions. IN: *Multilingual, #90, vol. 18, issue 6*, pp. 49-52.

Guzmán, R. 2008. Advanced automatic MT post-editing. IN: *Multilingual, #95, vol. 19, issue 3*, pp. 52-57.

Halliday, M. A. K. and J. Webster. 2003. *On language and linguistics*. London, New York: Continuum Publishing Group, p. 183.

Han, J., M. Kamber and J. Pei. 2012. *Data mining – concepts and techniques*. Waltham, MA, USA: Elsevier Inc.

Hatch, J. A. 2002. *Doing qualitative research in education settings*. New York: State University of New York, NY, USA, p. 50.

Horning, A. S. and A. Becker (eds.). 2006. *Revision: history, theory, and practice*. West Lafayette, IN, USA: Parlor Press LLC.

Hutchins, W. J. and H. L. Somers. 1992. *An introduction to machine translation*. London, United Kingdom: Academic Press.

Hutchins, W. J. 2000. *Early years in machine translation: memoirs and biographies of pioneers*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Hutchins, J. 2003. Commercial systems: the state of the art. IN: Somers, H. (ed.), *Computers and translation: a translator's guide*. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 161-174.

IRCSET – Irish Research Council for Science, Engineering & Technology. [Online]. Available from: http://www.research.ie/ [Accessed October 2012].

Isahara, H., M. Utiyama, E. Yamamoto, A. Terada and Y. Abe. 2008. Application of resource-based machine translation to real business scenes. IN: *Proceedings of the language resources and evaluation conference 2008*. Marrakech, Morocco, May 26–June 1, p. 3071.

Itagaki, M., T. Aikawa and X. He. 2007. Automatic validation of terminology translation consistency with statistical method. IN: *Proceedings of MT summit XI*. Copenhagen, Denmark, September 9–14 (no page numbers).

Jackson, S. L. 2011. *Research methods and statistics: a critical thinking approach*. Belmont, CA, USA: Cengage Learning.

Kelly, N. and D. A. DePalma. 2012. *The top 100 language service providers.* Common Sense Advisory, MA, USA. Available from: http://www.commonsenseadvisory.com/Portals/0/downloads/120531_QT_Top_100_LSPs.pdf [Accessed October 2012].

Kittur, A. 2010. Crowdsourcing, collaboration and creativity. IN: *sroads, The ACM Magazine for Students – Comp-YOU-Ter, Volume 17 Issue 2, Winter 2010*. ACM New York, NY, USA, pp. 22-26.

Koehn, P. 2010. Enabling monolingual translators: post-editing vs. options. IN: *HLT '10 Human language technologies: the 2010 annual conference of the North American chapter of the Association for Computational Linguistics*. Los Angeles, CA, USA, June 1–6, pp. 537-545.

Koehn, P. 2010. *Statistical machine translation*. Cambridge, United Kingdom: Cambridge University Press.

Koehn, P. 2011. What is a better translation? Reflections on six years of running evaluation campaigns. [Online]. IN: *Tralogy Conference*. CNRS, Paris, France, March 3–4. Available from: http://homepages.inf.ed.ac.uk/pkoehn/publications/tralogy11.pdf [Accessed October 2012] (no page numbers).

Krings, H. P. 2001. *Repairing texts: empirical investigations of machine translation post-editing processes*. Kent, OH, USA: The Kent State University Press, edited/translated by G. S. Koby.

Lagoudaki, E. 2006. *Translation memories survey. translation memory systems: enlightening users' perspective*. [Online]. Imperial College London, United Kingdom, November 2006. Available from: http://www3.imperial.ac.uk/portal/pls/portallive/docs/1/7307707.PDF [Accessed February 2009].

Loffler-Laurian, A.-M. 1984. Machine translation : what type of post-editing on what type of documents for what type of users. IN: *Proceedings of the 10th international conference on computational linguistics and 22nd annual meeting of the Association for Computational Linguistics*. Stanford, CA, USA, July 2–6, pp. 236-238.

Loffler-Laurian, A.-M. 1996. *La traduction automatique*. Paris, France: Presses Universitaires du Septentrion.

Machine translation post-editing guidelines. 2011. [Online].TAUS (Translation Automaton User Society). Available from: http://www.translationautomation.com/guidelines/machine-translation-post-editing-guidelines [Accessed October 2012].

Marrafa, P. and A. Ribeiro. 2001. Quantitative evaluation of machine translation systems: sentence level. IN: *Proceedings of the 8th MT Summit*, Santiago de Compostela, Spain, September 18–22 (no page numbers).

Màster en Tradumática 2012–2013. [Online]. Universitat Rovira i Virgili, Catalonia, Spain. Available from: http://pagines.uab.cat/mastertradumatica/sites/pagines.uab.cat.mastertradumatica/files/calendari.pdf [Accessed October 2012].

Microsoft® Language Portal. [Online]. Available from: http://www.microsoft.com/language/en/us/default.mspx [Accessed October 2012].

Morse, J. M. 1994. Designing funded qualitative research. IN: Denzin, N. K. and Lincoln, Y. S. (eds.), *Handbook of qualitative research, 2nd ed*. Thousand Oaks, CA, USA: Sage, pp. 220-35.

Mukherjee, P. N. 2006. *Total quality management*. New Delhi, India: Prentice Hall of India Private Limited.

Newton, J. (ed.). 1992. *Computers in translation: a practical appraisal*, London; New York: Routledge, p. 8.

O'Brien, S. 2002. Teaching post-editing: a proposal for course content. IN: *6th EAMT workshop teaching machine translation*. Manchester, United Kingdom, November 14–15, pp. 99-106.

O'Brien, S. 2006. *Machine-translatability and post-editing effort: an empirical study using translog and choice network analysis*. PhD Thesis. School of Applied Language and Intercultural Studies, Dublin City University, Dublin, Ireland.

O'Brien, S. and R. Fiederer. 2009. Quality and machine translation: a realistic objective? IN: *The Journal of Specialised Translation, Issue 11 – January 2009*. Jostrans, pp. 52-74.

O'Brien, S., J. Roturier and G. de Almeida. 2009. Postediting Machine Translation. [Online]. Tutorial presented at: *The 12th conference of the Association for Machine Translation in the Americas, AMTA*. Ottawa, Ontario, Canada, Aug. 26-30. Available from: http://www.mt-archive.info/MTS-2009-OBrien-ppt.pdf [Accessed December 2012].

O'Brien, S. 2011. Towards predicting post-editing productivity. IN: *Machine Translation, Vol. 25(3)*. Netherlands: Springer, pp. 197-215.

O'Brien, S. 2011. *TAUS Workshop: Best practices for postediting*. Presentation given at the Park Plaza Victoria Hotel, Amsterdam, Netherlands, Feb. 8.

Offersgaard, L., C. Povlsen, L. Almsten and B. Maegaard. 2008. Domain-specific MT in use. IN: *Proceedings of the 12th EAMT conference*. Hamburg, Germany, September 22–23, pp. 153-154.

O'Hagan, M. 2009. Evolution of user-generated translation: fansubs, translation hacking and crowdsourcing. IN: *The Journal of Internalisation and Localisation Volume I, 2009*. Lessius University College/K.U.Leuven Department of Applied Language Studies/Localization Industry Standards Association, pp. 94-121.

Papineni, K., S. Roukos, T. Ward and W.-J. Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. IN: *Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL)*. Philadelphia, PA, USA, July 6–12, pp. 311-318.

Parton, K., N. Habash, K. McKeown, G. Iglezias and A. de Gispert. 2012. Can automatic post-editing make MT more meaningful? IN: *Proceedings of the 16th International workshop of the European Association for Machine Translation (EAMT 12)*. Trento, Italy, May 28–30.

Plitt, M. and F. Masselot. 2010. A productivity test of statistical machine translation post-editing in a typical localisation context. IN: *The Prague Bulletin of Mathematical Linguistics*. Prague, Czech Republic: Universita Karlova, pp. 7-16.

Pym, A. 1992. Translation error analysis and the interface with language teaching. IN: Dollerup, C. and A. Loddegaard (eds.), *Teaching translation and interpreting*. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 279-288.

Ribas López, C. 2008. *Translation memories as vehicles for error propagation*. PhD Thesis. Universitat Rovira i Virgili, Catalonia, Spain.

Ribeiro, G. Castelo Branco. 2006. *Avaliação de tradução automática no mercado de localização de software: um estudo de caso*. MSc. Thesis. Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, RJ, Brazil.

Rieche, A. Ceschin. 2004. *Memória de tradução: auxílio ou empecilho?* M.A. Thesis. Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, RJ, Brazil.

Rubin, A. and E. Babbie. 2009. *Research methods for social work (second edition)*. Belmont, CA, USA: Cengage Learning.

Ryan, J. P. 1993. Machine translation: matching reality to expectations. IN: Nirenburg, S. (ed.), *Progress in machine translation*. Amsterdam, Netherlands: IOS Press, pp. 225-235.

Schäfer, F. 2003. MT post-editing: How to shed light on the "unknown task". Experiences made at SAP. IN: *8th International workshop of the European Association for Machine Translation (EAMT 03)*. Dublin City University, Dublin, Ireland, May 15–17.

Seeburg, D. 2008. Trends in language technology: machine translation. IN: *Language Tech News*. American Translators Association, Vol. 2, No. 2, June 2008, pp. 11-13.

Silverman, D. 2011. *Interpreting qualitative data (4th edition)*. Thousand Oaks, CA, USA: Sage Publications.

Simard, M., N. Ueffing, P. Isabelle and R. Kuhn. 2007. Rule-based translation with statistical phrase-based post-editing. IN: *ACL 2007, Second workshop on statistical machine translation*. Prague, Czech Republic, June 23, pp. 203-206.

Sinkovics, R. R., B. Stöttinger, B. B. Schlegelmilch and S. Ram. 2002. Reluctance to use technology-related products: development of a technophobia scale. IN: *Thunderbird International Business Review*. Thunderbird School of Global Business Management, AZ, USA, Volume 44, July/August 2002, pp. 477-494.

Slocum, J. 1984. Machine translation: its history, current status, and future prospects. IN: *Proceedings of the 10th international conference on computational linguistics*. Stanford, CA, USA, July 02–06, p. 552.

Somers, H. L. 2001a. Machine translation, applications. IN: Baker, M. (ed.), *Routledge Encyclopedia of Translation Studies*. London, New York: Routledge, pp. 136-141.

Somers, H. L. 2001b. Machine translation, history. IN: Baker, M. (ed.), *Routledge Encyclopedia of Translation Studies*. London, New York: Routledge, pp. 140-142.

Somers, H. L. (ed.). 2003. *Computers and translation: a translator's guide*. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 4-5.

Sousa, S. C. M., W. Aziz and L. Specia. 2011. [Online] Assessing the post-editing effort for automatic and semi-automatic translations of DVD subtitles. IN: *Recent Advances in Natural Language Processing (RANLP-2011)*, Hissar, Bulgaria. Available from: clg.wlv.ac.uk/papers/ranlp-2011_sousa.pdf [Accessed January 2013].

Specia, L., N. Cancedda, M. Dymetman, M. Turchi and N. Cristianini. 2009. Estimating the sentence-level quality of machine translation systems. IN: *Proceedings of the 13th annual conference of the EAMT*. Barcelona, Spain, May 14–15, pp. 28-35.

Specia, L. 2011. Exploiting objective annotations for measuring translation post-editing effort. IN: Forcada, M. L., H. Depraetere and V. Vandeghinste (eds.), *Proceedings of the 15th annual EAMT conference*. Leuven, Belgium, May 30–31, pp. 73-80.

Spector, P.E. 1992. *Summated rating scale construction – an introduction*. Thousand Oaks, CA, USA: Sage Publications.

Systran. [Online]. Available from: http://www.systran.co.uk/ [Accessed October 2012].

Systran and MultiCorpora integrate technologies for increased translation quality and volume. [Online]. IN: *MultiCorpora.* Available from: http://www.multicorpora.com/about/news/about_news_1971_en/ [Accessed February 2009].

Tatsumi, M. 2009. Correlation between automatic evaluation metric scores, post-editing speed, and some other factors. IN: *Proceedings of the 12th MT summit*. Ottawa, ON, Canada, August 26–31, pp. 332-339.

Tatsumi, M. 2010. *Post-editing machine translated text in a commercial setting: observation and statistical analysis*. PhD Thesis. Dublin City University. Dublin, Ireland.

Tatsumi, M. and J. Roturier. 2010. Source text characteristics and technical and temporal post-editing effort: what is their relationship? IN: Zhechev, V. (ed.), *Proceedings of the 2nd Joint EM+CNGL workshop: Bringing MT to the user: research on integrating MT in the translation industry*. Denver, CO, USA, November 4, pp. 43-51.

Tatsumi, M., T. Aikawa, K. Yamamoto and H. Isahara. 2012. How good is crowd post-editing? Its potential and limitations. IN: IN: *Proceedings of the 10th conference of the Association for Machine Translation in the Americas, AMTA*. San Diego, CA, USA, October 28–November 1.

TAUS Research - Postediting in Practice. 2010. [Online]. IN: Translation Automation User Society (TAUS). Available from: http://www.translationautomation.com/images/stories/pdf/benchmarch-data-for-postediting-practices-globally.pdf [Accessed January 2013].

Translation Memory Survey 2002. [Online] IN: The Localization Industry Standards Association. Available from: http://www.lisa.org/Translation-Memory-S.523.0.html [Accessed 10 February 2009].

Van der Meer, J. 2003. The business case for MT – the breakthrough is for real. IN: *LISA Newsletter Vol. XII (2.6).*

Vasconcellos, M. 1986. Post-editing on-screen: machine translation from Spanish into English. IN: *Proceedings of Translating and the Computer 8*. London, United Kingdom, November 13–14, pp. 133-146.

Vieira. L. and L. Specia. 2011. A review of machine translation tools from a post-editing perspective. IN: *3rd Joint EM+/CNGL workshop bringing MT to the user: research meets translators (JEC 2011)*. Luxembourg, October 14.

VistaTEC. [Online]. Available from: http://www.vistatec.com/en/home.htm [Accessed October 2012].

Wendt, C. 2008. Large-scale deployment of statistical machine translation: example Microsoft. Presentation given at: *The 8th conference of the Association for Machine Translation in the Americas, AMTA*. Waikiki, HI, USA, Oct. 26.

White, J. S. 2003. How to evaluate machine translation. IN: Somers, H. L. (ed.), *Computers and translation: a translator's guide*. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 211-244.

Winiwarter, W. 2007. Machine translation using corpus-based acquisition of transfer rules. IN: *Proceedings of the second international conference on digital information management (IEEE Engineering Management Society)*. University of Lyon, France, October 27–29, pp. 345-350.

Wolochwianski, R. 2008a. What is really at stake with machine translation? IN: *Language Tech News*. American Translators Association, Vol. 2, No. 3, November 2008, pp. 9-14.

Wolochwianski, R. 2008b. MT Post-editing: a threat or an opportunity? (Presentation slides). IN: *Localization Latin America*. Buenos Aires, Argentina, November 17–20 (no page numbers).

Yunker, J. 2008. The end of translation as we know it. IN: *MultiLingual Magazine*. December 2008, #100, Volume 19, Issue 8, pp. 30-31.

Zaidan, O.F. and C. Callison-Burch. 2011. Crowdsourcing translation: professional quality from non-professionals. IN: *Proceedings of the 49th annual meeting of the Association for Computational Linguistics*. Portland, OR, USA, June 19–24, pp. 1220-1229.

Zhuang, X. 2002. The emerging role of translation experts in the coming MT era. [Online]. IN: *Translation Journal*. Volume 6, No. 4, October 2002. Available from: http://translationjournal.net/journal/22mt.htm [Accessed October 2012].

Appendices

Appendix A: Questionnaire

1. Taking into account your previous experience with machine translation (using online MT engines, using MT as part of translation/localisation projects or in any other context, as a translator or as a user), please highlight one of the options below to indicate your opinion regarding the general quality of translated texts produced by machine translation only (without post-editing):
1 - Very bad quality
2 - Bad quality
3 - Average quality
4 - Good quality
5 - Very good quality

2. Please highlight one of the options below to indicate your opinion about the following statement:

Machine translation can be helpful for translators (as a productivity tool, for instance).
1 - Strongly disagree
2 - Disagree
3 - Neither agree nor disagree
4 - Agree
5 - Strongly agree

3. Please highlight one of the options below to indicate your opinion about the following statement:

Post-editing texts produced by machine translation is an activity that interests me as a translator, as it can provide me with new sources of work and new professional skills.
1 - Strongly disagree
2 - Disagree
3 - Neither agree nor disagree
4 - Agree
5 - Strongly agree

Appendix B: Segments used for the main PE experiment

1	{1}May 13, 2010{2}
2	Do not install this Client Update unless you have a clear need to access an Autodesk Vault 2011 SP1 Server release from your Autodesk Inventor 2010 Service Pack 3 installation.
3	Verify you have administrator privileges on your local machine to install this Client Update.
4	On Windows Vista installations, disable the User Account Control feature.
5	A minimum of 39 MB of free disk space is required on the drive where your temporary files are directed.
6	If required, reboot the system after the installation is finished.
7	We thank all our customers for your continued business and for feedback regarding the content of this release.
8	How project files work with factory layout files
9	If the Default project is active the files are all stored in the My Documents folder.
10	{14}Change the Active Project{15}
11	Check boxes enable layer visibility to be toggled on or off as required.
12	The template settings enable you to specify the template file to be used when creating a new factory layout.
13	The Factory Options settings are global so that all new documents use the settings.
14	The Factory Options dialog box contains three tabs.
15	The option to turn on or off the visibility of selected drawing layers is also provided.
16	This check box is disabled by default.
17	The black and white buttons at the upper right of the dialog box let you toggle the preview background color between black and white.
18	As you zoom in, the actual distance decreases; the distance increases as you zoom out.
19	The components selected in step 2 move to align with the reference component.
20	New components are placed in space, and you have to constrain them manually.
21	Layer visibility can be toggled on or off as required.
22	The preview updates as layer visibility is toggled on or off.
23	Click the Browse button to navigate to an alternative folder location if you wish to specify a different template file.
24	The preference settings only affect the current document.
25	For existing documents, you have to change the settings in each document if you want the new behavior.
26	By default, the floor is specified as visible in a new factory layout.

27	It is set to resize automatically once components are moved outside the default boundary.
28	You can control the floor color and whether a border displays.
29	Controls the floor visibility, size, and style.
30	The floor is resized such that components always remain positioned on the floor.
31	{134}User Defined{135}
32	Select this option to set the length and width of the floor manually.
33	{160}Restore Defaults{161}
34	You can navigate through the data by selecting folders or you can do a property search.
35	A list of valid search terms is dynamically displayed.
36	The component remains listed in the folder until the next search is conducted when it is replaced with the new search result.
37	From this location, you can double-click folders to move lower or use the Folder up button to move higher in the tree structure.
38	If multiple insertion points have been defined on the model, press the {193}TAB{194} key to cycle through each of the insertion points until you identify the point you want.
39	The aligning components automatically rotate so their axes align with the axes on the reference component.
40	In the rare event that a connection cannot be made, the connector points change from yellow to red spheres.
41	You can modify the location and orientation of components after they are placed by using the commands on the Layout Tools panel on the Factory tab.
42	How to modify parameter values in the properties browser
43	Select any of the parameter values in the right-hand column and enter a new value.
44	To finalize the change and update the model, press the {207}Enter{208} key on the keyboard or click the Update button {209}.
45	Using the Update button enables you to update multiple parameter values at one time.
46	A Factory Layout can be populated with component data from multiple sources.
47	For example, infrequently used components are not usually published into the library.
48	For these situations, you can add a model directly to the layout.
49	Drag the selected component to the required location in the layout and press the left mouse button to insert the component.
50	How to move components to align with another component
51	How to distribute components so they are equally spaced
52	Components may also be distributed equally at horizontal or vertical intervals.
53	The selected components rotate so their orientation matches the reference component.

54	Aligns the selected component(s) to the X axis orientation of the reference component.
55	Aligns the distribution of the selected component(s) horizontally by their left edges.
56	The Align command lets you align multiple components in various directions and orientations, based on a selected reference component.
57	The alignments occurs between the selected components, and does not use a reference component.
58	The publishing tool saves the file to the library.
59	To provide a more descriptive name for the new folder, right-click over the New Folder name and select Rename Folder from the pop-up context menu.
60	The pop-up context menu also provides the option to create an additional subfolder under the new folder you have created.
61	The parameters are automatically selected if they are marked for export in the model parameters.
62	Enter model information in the property fields.
63	The connector point displays as a green sphere.
64	Continue to create additional connector points, as necessary.
65	Although the primary orientation is defined, you might have to rotate it for certain applications.
66	Only the component you select has the new orientation.
67	Any changes made to the selected component only reflect in that specific instance of the component.
68	The drawing can also be created in another CAD program capable of outputting a file in DWG format.
69	The drawing may be repositioned, scaled, rotated, or deleted.
70	The drawing preview updates to reflect any changes to layer visibility.
71	A drop-down menu displays.
72	Enter a value in the Y: text box to move the drawing in the Y direction.
73	Selection: Assigns a reference component and components to align to it.
74	Align Position: Moves selected components to a location defined by another component without changing their orientation.

Appendix C: Post-editing instructions - French

1. Objective

The main objective of post-editing is to make the machine-translated output understandable and grammatically correct, and to fully convey the meaning from the source text.

You should make the MT output have the correct meaning, using understandable language, in **as few edits as possible**.

2. Level of post-editing required

Specifically for this exercise, intermediate post-editing is required. Intermediate post-editing consists of making all necessary corrections in the machine-translated text, but avoiding preferential corrections.

Please note that **purely stylistic changes should be avoided**.

Please also note that post-editing **is not the same as revising**. As such, you should try to **minimise** the number of edits whenever possible, without compromising the meaning.

3. Instructions and examples

In most cases, there will be different ways to edit the machine-translated output. When choosing between two correct options, try and select the one that involves the **lowest number of edits**.

Do not spend too much time over a problem. If you cannot think quickly of a way to improve the output, leave it unchanged (there is no point in trying a few alternatives and reverting eventually to the initial suggestion).

The types of issues that you should correct are:

- Grammatical errors, such as incorrect agreement, incorrect capitalisation, accents missing etc.;
- Misspellings;
- Mistranslations;
- Incorrect tags (tags missing, tags incorrectly positioned etc.);
- Incorrect spaces (spaces missing, extra spaces etc.).

Some of the types of changes that **should not be made** are:

- Replacing a correct term with a synonym: if a term is grammatically correct and it conveys the correct meaning, it should not be replaced for stylistic purposes;
- Changing the order of the terms in a sentence for stylistic preferences: if the original order is grammatically correct and the text is understandable, it should not be changed;
- Replacing the passive voice with the active voice or vice-versa: if the original translation is grammatically correct and it conveys the same meaning as the source text, this type of change should not be made.

Examples are provided in the next section.

Examples of issues that should be corrected:

- **Missing information:**

In case the machine-translated output is missing information contained in the source text, you should add the missing information.

Example:

Source text	MT output	Post-edited text	Comments
Booting from a **USB** hard drive or floppy disk	Le démarrage depuis un disque dur ou une disquette	Démarrage depuis un disque dur **USB** ou une disquette	"USB" was missing from the machine-translated output.

- **Extra information:**

If the machine-translated output includes information not contained in the source text, you should remove it.

Example:

Source text	MT output	Post-edited text	Comments
If this switch is turned on, messages are sent.	Si ce paramètre est activé, les **fichiers de** messages sont envoyés.	Si ce paramètre est activé, les messages sont envoyés.	"fichiers de" was not part of the source text, so it should be removed.

Mistranslations:

Incorrect transfers of meaning should be corrected by the post-editors.

Example:

Source text	MT output	Post-edited text	Comments
During a silent upgrade, the user will not be prompted to download and start the upgrade, but they will still be prompted to confirm that they want to actually perform the upgrade.	Lors d' une mise à niveau silencieuse, l'utilisateur ne sera pas invité à télécharger et à démarrer la mise à niveau, mais ils **ne peut être invité** à confirmer qu'il souhaite réellement effectuer la mise à niveau.	Lors d'une mise à niveau silencieuse, l'utilisateur ne sera pas invité à télécharger ni à démarrer la mise à niveau, mais il **sera invité** à confirmer qu'il souhaite réellement effectuer la mise à niveau.	In the raw MT output, "they will still be prompted" was incorrectly translated as "ils ne peut être invité", which does not convey the same meaning as the source text.

- **Punctuation:**

Incorrect punctuation should also be corrected in the post-edited text.

Example:

Source text	MT output	Post-edited text	Comments
Click this to decompress compressed files as they are backed up.	Cliquez sur cette option pour decompress**,** les fichiers compressés ils sont sauvegardés.	Cliquez sur cette option pour décompresser les fichiers compressés au fur et à mesure qu'ils sont sauvegardés.	There was an incorrect comma in the raw MT output.

Verb tenses:

The post-edited text must have the same verb tense as the source text **when this makes a difference in the meaning**.

Source text	MT output	Post-edited text	Comments
The files are available to the installation system.	Les fichiers **seront** accessibles au système d'installation	Les fichiers **sont** accessibles au système d'installation.	The verb is in the present tense in the source text, but it was in the future in the MT output, so it was corrected in the post-edited

			text.

- **Determiners:**

In some cases, determiners may be required to convey the correct meaning and to render the text grammatically correct.

Example:

Source text	MT output	Post-edited text	Comments
Proper configuration is essential for a smoothly running system.	Configuration appropriée est indispensable pour assurer un bon fonctionnement système.	**Une** bonne configuration est indispensable pour assurer un bon fonctionnement du système.	"Une" was missing from the raw machine translation output, so this was corrected in the post-edited text.

Tags:

Extra, missing or incorrectly positioned tags should be corrected.

Example:

Source text	MT output	Post-edited text	Comments
The respective media must be created with the utility {7}mkbootdisk{8}, which can be found together with its documentation on the first CD or DVD in the directory {9}/boot/<architecture>{10}.	**{7}**Le respectifs doivent être créées avec l'utilitaire mkbootdisk, qui se trouvent en collaboration avec sa documentation sur le premier CD ou DVD dans le répertoire /boot/<architecture> **{8}{9}{10}**	Les supports correspondants doivent être créés avec l'utilitaire **{7}**mkbootdisk**{8}**, qui se trouve avec sa documentation sur le premier CD ou DVD dans le répertoire **{9}**/boot/<architecture>**{10}**	All the tags were misplaced in the raw MT output. If this type of issue is not corrected, it can cause formatting errors and other types of problems.

Examples of items that do not need to be changed:

- **Synonyms:**

Correct terms should not be replaced with synonyms.

Example:

Source text	MT output	Post-edited text	Comments
The exact protocol depends on your hardware.	Le moment **exact** protocole dépend de votre matériel.	Le protocole **précis** dépend de votre matériel.	It was not necessary to replace "exact" with "précis". The post-edited text could have been left as: Le protocole **exact** dépend de votre matériel.

A similar approach should be applied to phrasal ordering: the order of phrases should only be changed if this is necessary to correct the meaning of the text. Otherwise, it should be left unchanged.

Appendix D: Post-editing instructions - Brazilian Portuguese

1. Objective

The main objective of post-editing is to make the machine-translated output understandable and grammatically correct, and to fully convey the meaning from the source text.

You should make the MT output have the correct meaning, using understandable language, in **as few edits as possible**.

2. Level of post-editing required

Specifically for this exercise, intermediate post-editing is required. Intermediate post-editing consists of making all necessary corrections in the machine-translated text, but avoiding preferential corrections.

Please note that **purely stylistic changes should be avoided**.

Please also note that post-editing **is not the same as revising**. As such, you should try to **minimise** the number of edits whenever possible, without compromising the meaning.

3. Instructions and examples

In most cases, there will be different ways to edit the machine-translated output. When choosing between two correct options, try and select the one that involves the **lowest number of edits**.

Do not spend too much time over a problem. If you cannot think quickly of a way to improve the output, leave it unchanged (there is no point in trying a few alternatives and reverting eventually to the initial suggestion).

The types of issues that you should correct are:

- Grammatical errors, such as incorrect agreement, incorrect capitalization, accents missing etc.;
- Misspellings;
- Mistranslations;
- Incorrect tags (tags missing, tags incorrectly positioned etc.);
- Incorrect spaces (spaces missing, extra spaces etc.).

Some of the types of changes that **should not be made** are:

- Replacing a correct term with a synonym: if a term is grammatically correct and it conveys the correct meaning, it should not be replaced for stylistic purposes;
- Changing the order of the terms in a sentence for stylistic preferences: if the original order is grammatically correct and the text is understandable, it should not be changed;
- Replacing the passive voice with the active voice or vice-versa: if the original translation is grammatically correct and it conveys the same meaning as the source text, this type of change should not be made.

Examples are provided in the next section.

Examples of issues that should be corrected:

- **Missing information:**

In case the machine-translated output is missing information contained in the source text, you should add the missing information.

Example:

Source text	MT output	Post-edited text	Comments
Booting from a hard drive **or floppy disk**	Inicialização por meio de um hard drive	Inicialização por meio de um disco rígido **ou disquete**	"ou disquete" was missing from the machine-translated output.

- **Extra information:**

If the machine-translated output includes information not contained in the source text, you should remove it.

Example:

Source text	MT output	Post-edited text	Comments
If this switch is turned on, **messages** are sent.	Se essa opção for ativada, **arquivos de mensagens** serão enviados.	Se essa opção for ativada, **mensagens** serão enviadas.	"arquivos de" was not part of the source text, so it should be removed.

Mistranslations:

Incorrect transfers of meaning should be corrected by the post-editors.

Example:

Source text	MT output	Post-edited text	Comments
During a silent upgrade, the user will not be prompted to download and start the upgrade, but **they will still be** prompted to confirm that they want to actually perform the upgrade.	Durante uma atualização silenciosa, o usuário não será solicitado a fazer o download e começar a atualização, mas **não será** solicitado a confirmar que quer realmente fazer a atualização.	Durante uma atualização silenciosa, o usuário não será solicitado a fazer o download e começar a atualização, mas **será** solicitado a confirmar que quer realmente fazer a atualização.	In the raw MT output, "will still be prompted" was incorrectly translated as "não será solicitado", which does not convey the same meaning as the source text.

Punctuation:

Incorrect punctuation should also be corrected in the post-edited text.

Example:

Source text	MT output	Post-edited text	Comments
Click this to decompress compressed files as they are backed up.	Clique aqui para descompactar**,** arquivos compactados enquanto é feito o backup.	Clique aqui para descompactar arquivos compactados enquanto é feito o backup.	There was an incorrect comma in the raw MT output.

- **Verb tenses:**

The post-edited text must have the same verb tense as the source text **when this makes a difference in the meaning**.

Source text	MT output	Post-edited text	Comments
The files **are** available to the installation system.	Os arquivos **estarão** disponíveis para a instalação do sistema.	Os arquivos **estão** disponíveis para o sistema de instalação.	The verb is in the present tense in the source text, but it was in the future in the MT output, so it was corrected in the post-edited text.

- **Determiners:**

In some cases, determiners may be required to convey the correct meaning and to render the text grammatically correct.

Example:

Source text	MT output	Post-edited text	Comments
Proper configuration is essential for a smoothly running system.	Configuração adequada é essencial para a execução correta do sistema.	**A** configuração adequada é essencial para a execução correta do sistema.	"A" was missing from the raw machine translation output, so this was corrected in the post-edited text.

- **Tags:**

Extra, missing or incorrectly positioned tags should be corrected.

Example:

Source text	MT output	Post-edited text	Comments
The respective media must be created with the utility **{7}**mkbootdisk**{8}**, which can be found together with its documentation on the first CD or DVD in the directory **{9}**/boot/<architecture>**{10}**.	A mídia respectivos devem ser criados com o utilitário mkbootdisk **(7) (8)**, que podem ser encontradas juntamente com a documentação sobre o primeiro CD ou DVD no diretório **(9)** / boot / <architecture> **(10)**.	A respectiva mídia deve ser criada com o utilitário **{7}**mkbootdisk**{8}**, que pode ser encontrado juntamente com a documentação no primeiro CD ou DVD no diretório **{9}**/boot/<architecture>**{10}**.	All the tags were misplaced in the raw MT output (as well as having parentheses instead of curly brackets). If this type of issue is not corrected, it can cause formatting errors and other types of problems.

Examples of items that do not need to be changed:

- **Synonyms:**

Correct terms should not be replaced with synonyms.

Example:

Source text	MT output	Post-edited text	Comments
The **exact** protocol depends on your hardware.	O protocolo **exato** depende do seu hardware.	O protocolo **específico** depende do seu hardware.	It was not necessary to replace "exato" with "específico". The post-edited text could have been left as: O protocolo **exato** depende do seu hardware.

A similar approach should be applied to phrasal ordering: the order of phrases should only be changed if this is necessary to correct the meaning of the text. Otherwise, it should be left unchanged.

Appendix E: DCU Ethics forms

DUBLIN CITY UNIVERSITY

I. Research Study Title

Investigation of Post-Editing Strategies

Giselle de Almeida, IRCSET/Vistatec, Dublin City University

Sharon O'Brien, School of Applied Language and Intercultural Studies, Dublin City University

II. Clarification of the purpose of the research

(1) *Our objectives are to verify what parts of text cause difficulties for post-editing in the languages studied, to verify if there is a correlation between translation/post-editing experience and speed/efficiency in the post-editing process, and to analyse if similar strategies, difficulties and results can be found for languages of the same language family.*

III. Confirmation of particular requirements as highlighted in the Plain Language Statement

You will be asked to participate in one session, which should last approximately two hours. In this session you will be asked to post-edit a text comprising 1008 words on-screen. The text will have been machine-translated in advance from English into your native language (French or Brazilian Portuguese). Your actions on-screen to post-edit the text will be automatically recorded by a keylogging software and by a screen-capture program, as well as being monitored by the researchers. You will also be asked in advance to indicate your translation/post-editing experience in years.

Participant – please complete the following (Circle Yes or No for each question)	
Have you read or had read to you the Plain Language Statement	*Yes/No*
Do you understand the information provided?	*Yes/No*
Have you had an opportunity to ask questions and discuss this study?	*Yes/No*
Have you received satisfactory answers to all your questions?	*Yes/No*

IV. Confirmation that involvement in the Research Study is voluntary

Involvement in this study is voluntary. You may withdraw from the Research Study at any point. If you would rather not be part of the research study, please let us know as soon as possible.

V. Advice as to arrangements to be made to protect confidentiality of data, including that confidentiality of information provided is subject to legal limitations

Your anonymity will be protected at all times. You will be given an identifier such as "FR01" or "BR01" and no link will ever be made to your real identity. The data collated will be used only by the researchers named above and will not be given to anybody else.

VII. Signature:

I have read and understood the information in this form. My questions and concerns have been answered by the researchers, and I have a copy of this consent form. Therefore, I consent to take part in this research project.

Participants' Signature: _____

Name in Block Capitals: _____

Witness: _____

Date: _____

DUBLIN CITY UNIVERSITY

I. Introduction to the Research Study

This is a study of post-editing strategies for machine translation output. The study is being carried out by Sharon O'Brien and Giselle de Almeida from the School of Applied Language and Intercultural Studies and IRCSET, DCU. The main point of contact for this study at VistaTEC is Phil Ritchie (philr@vistatec.ie). The contact details for Sharon O'Brien and Giselle de Almeida are:

t. 700 5832 e. sharon.obrien@dcu.ie

t. 700 5385 e. giselle.dealmeida2@mail.dcu.ie

II. Details of what involvement in the Research Study will require

Your involvement in this study will involve the following:

- On-screen post-editing of machine-translated segments.

- Indicating your translation/post-editing experience in years.

III. Potential risks to participants from involvement in the Research Study (if greater than that encountered in everyday life)

There are no risks involved in participating in this study.

IV. Benefits (direct or indirect) to participants from involvement in the Research Study

The indirect benefits of your participation in this study are:

- *You are helping to identify some of the main strategies and difficulties found in the post-editing process, and to identify whether or not a correlation exists between languages of the same language family. You are thereby contributing to the field, since little work has been done on this topic to date.*

V. Advice as to arrangements to be made to protect confidentiality of data, including that confidentiality of information provided is subject to legal limitations

Your anonymity will be protected at all times. You will be given an identifier such as "FR 01" or "BR 01" and no link will ever be made to your real identity. The data collated will be used only by the researchers mentioned above and will not be given to anybody else.

VI. Advice as to whether or not data is to be destroyed after a minimum period

The data will be stored in a secure location only at DCU. The data will be destroyed within five years of its acquisition.

VII. Statement that involvement in the Research Study is voluntary

Involvement in this study is voluntary. You may withdraw from the Research Study at any point. If you would rather not be part of the research study, please let us know as soon as possible.

If participants have concerns about this study and wish to contact an independent person, please contact: The Secretary, Dublin City University Research Ethics Committee, c/o Office of the Vice-President for Research, Dublin City University, Dublin 9. Tel 01-7008000

Appendix F: Presentation of the data

F.1. Overview of this appendix

This appendix presents the data collected during the main PE experiment. It is subdivided in sections containing detailed tables and charts with the data for the participants for each language, indicating the numbers for each master category from the typology (Essential Changes, Preferential Changes, Essential Changes Not Implemented and Introduced Errors), as well as the numbers for the subcategories and the totals, first for French, then for Brazilian Portuguese. In addition, there is a breakdown of the data per language, also including tables and charts. Examples are provided for French and for Brazilian Portuguese after the presentation of the breakdown of the data.

The analysis of the data is presented in a separate chapter (Chapter 4, Analysis). The analysis and the full data are presented separately in order to provide better organisation and clarity, and also to avoid having an extremely long chapter, which might be cumbersome to browse and to consult.

As previously mentioned in Chapter 3, the data from French participants 3 and 6 is not included here due to unavailability (one of these participants did not finish the PE section, and the data from the other participant was corrupted).

F.2. Results of the linguistic analysis per individual participant

This section presents the results participant by participant, first for French, then for Brazilian Portuguese. The typology of PE changes is presented in tabular format, commencing with Essential Changes, followed by Preferential Changes, Essential Changes Not Implemented and Introduced Errors. Each of these categories is followed by a bar chart.

F.3. French - Participant 1

Table F.1: French - Participant 1: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	23	Accuracy - extra information: 1 Accuracy - information missing: 22
Format	4	
Language	75	Language - adverbs: 3 Language - capitalisation: 10 Language - conjunctions: 2 Language - determiners: 4 Language - gender: 14 Language - number: 14 Language - phrasal ordering: 3 Language - prepositions: 7 Language - pronouns: 5 Language - punctuation: 5 Language - verbs: 8
Mistranslation	25	
Style	2	
Total	**129**	

Figure F.1: French - Participant 1: Essential Changes

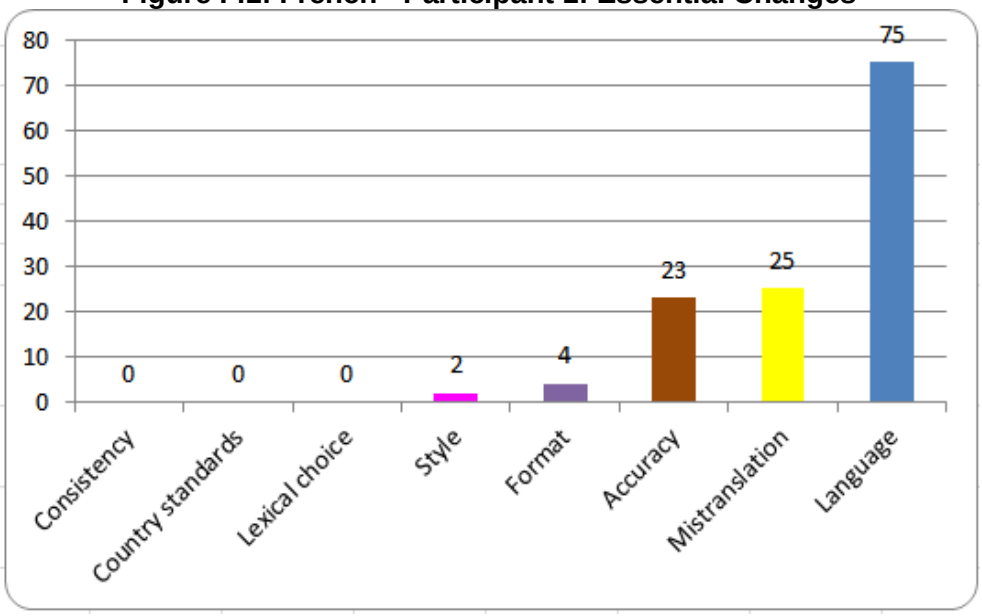


Table F.2: French - Participant 1: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	15	Language - adverbs: 1 Language - determiners: 5 Language - number: 4 Language - phrasal ordering: 1 Language - prepositions: 3 Language - punctuation: 1
Lexical choice	9	
Style	2	
Total	**26**	

Figure F.2: French - Participant 1: Preferential Changes

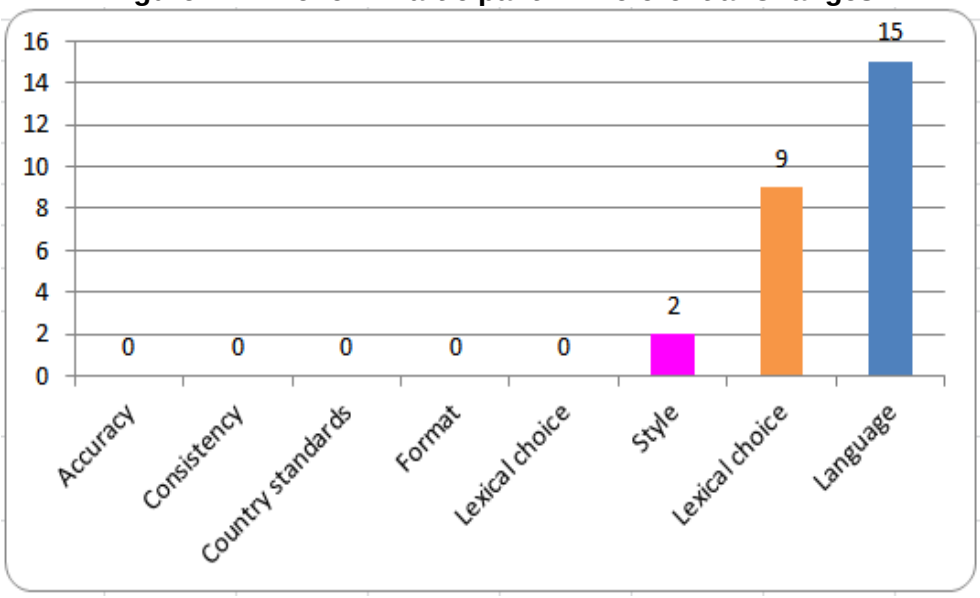


Table F.3: French - Participant 1: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	5	Accuracy - information missing: 5
Format	1	
Language	3	Language - determiners: 1 Language - prepositions: 1 Language - pronouns: 1
Mistranslation	9	
Total	**18**	

Figure F.3: French - Participant 1: Essential Changes Not Implemented

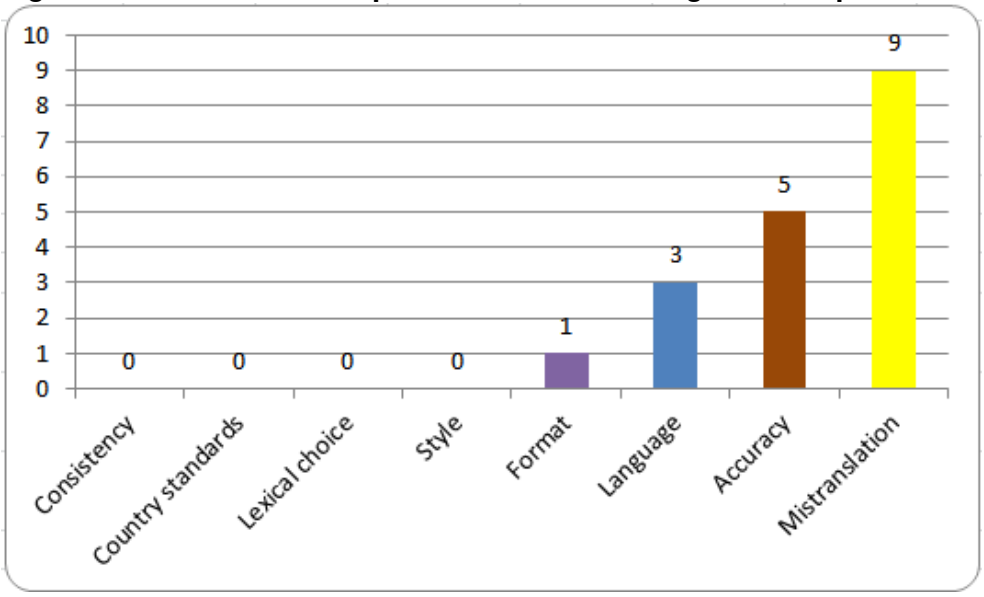


Table F.4: French - Participant 1: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	3	Accuracy - information missing: 1
Language	3	Language - number: 2 Language - prepositions: 1
Mistranslation	4	
Total	**10**	

Figure F.4: French - Participant 1: Introduced Errors

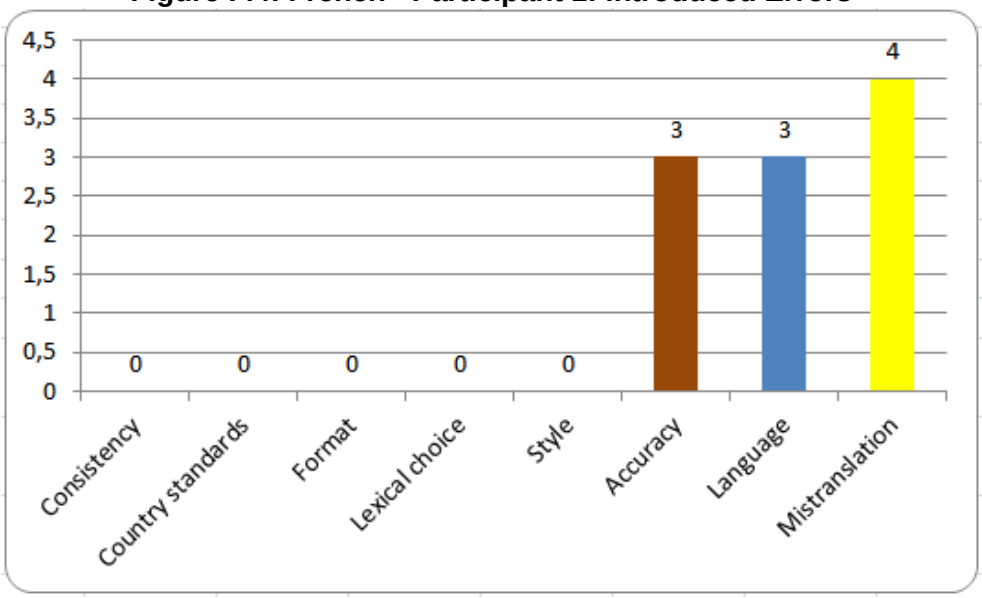
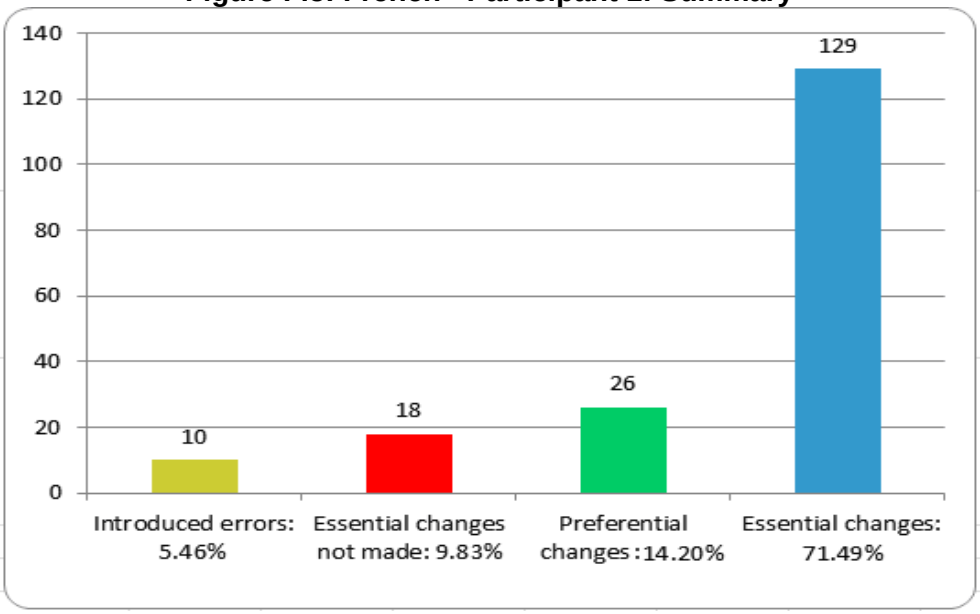


Table F.5: French - Participant 1: Summary

ESSENTIAL CHANGES	129
Preferential Changes	26
Essential Changes Not Implemented	18
Introduced Errors	10
TOTAL	183

Figure F.5: French - Participant 1: Summary



F.4. French - Participant 2

Table F.6: French - Participant 2: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	15	Accuracy - information missing: 15
Format	1	
Language	66	Language - adverbs: 5 Language - capitalisation: 7 Language - conjunctions: 2 Language - determiners: 4 Language - gender: 12 Language - number: 9 Language - phrasal ordering: 2 Language - prepositions: 7 Language - pronouns: 5 Language - punctuation: 4 Language - verbs: 9
Lexical choice	1	
Mistranslation	17	
Style	1	
Total	**101**	

Figure F.6: French - Participant 2: Essential Changes

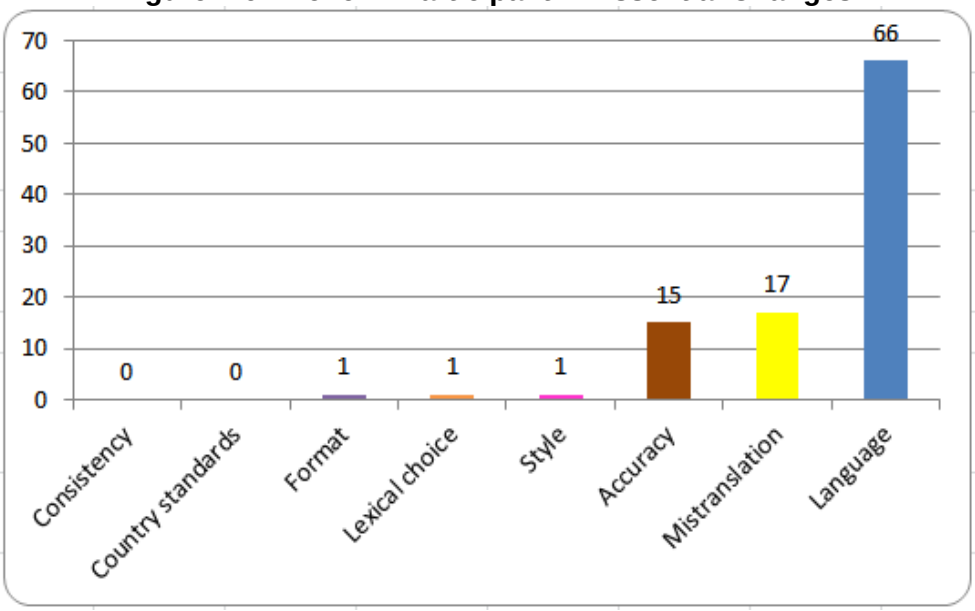


Table F.7: French - Participant 2: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	31	Language - capitalisation: 1 Language - determiners: 6 Language - number: 7 Language - phrasal ordering: 3 Language - prepositions: 9 Language - pronouns: 1 Language - punctuation: 1 Language - verbs: 3
Lexical choice	15	
Style	8	
Total	**54**	

Figure F.7: French - Participant 2: Preferential Changes

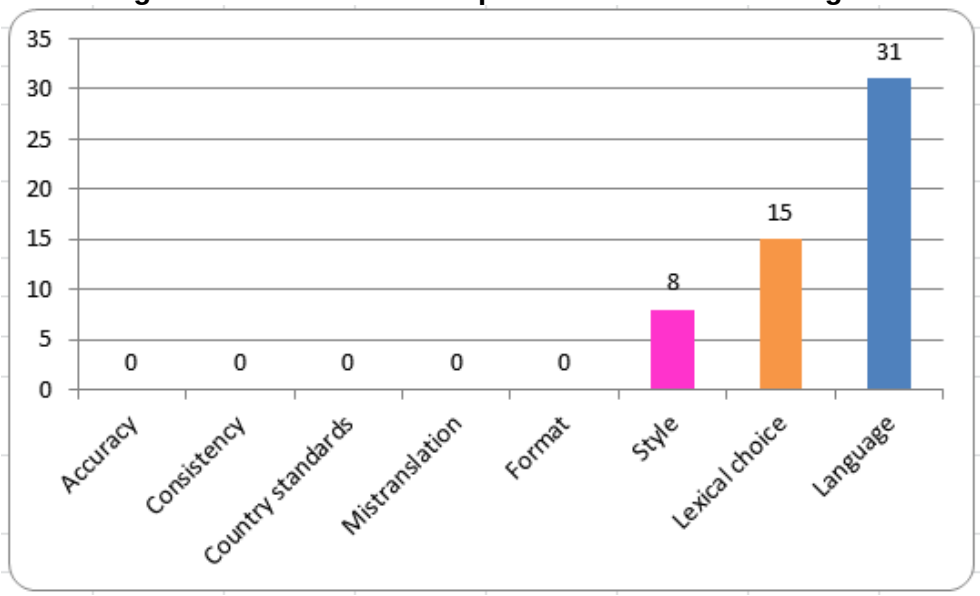


Table F.8: French - Participant 2: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	13	Accuracy - extra information: 1 Accuracy - information missing: 12
Format	7	
Language	12	Language - capitalisation: 3 Language - determiners: 2 Language - gender: 1 Language - number: 2 Language - phrasal ordering: 1 Language - pronouns: 1 Language - punctuation: 2
Mistranslation	17	
Total	**49**	

Figure F.8: French - Participant 2: Essential Changes Not Implemented

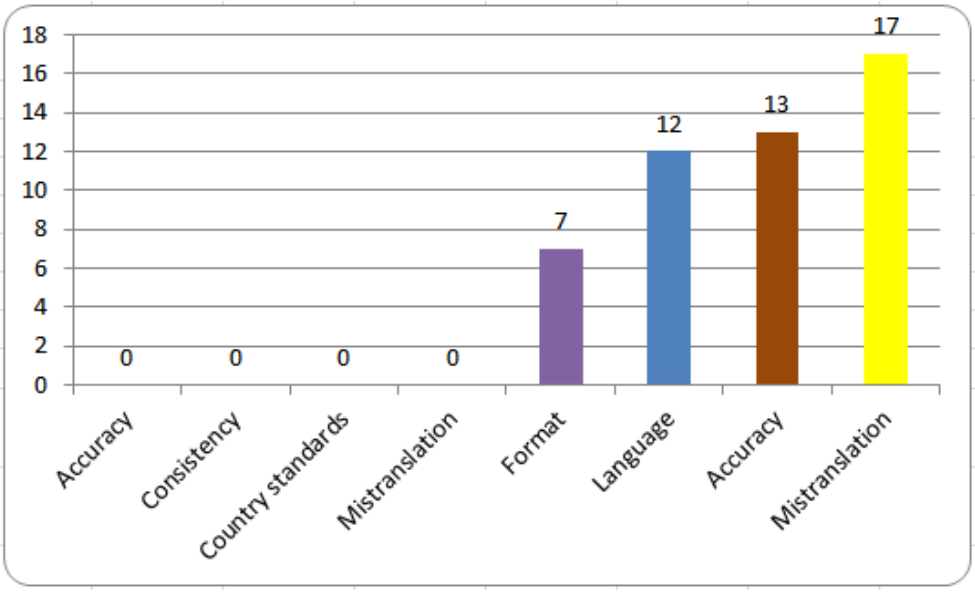


Table F.9: French - Participant 2: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	2	Language - information missing: 1 Language - untranslated text: 8
Language	1	Language - prepositions: 1
Mistranslation	9	
Total	**12**	

Figure F.9: French - Participant 2: Introduced Errors

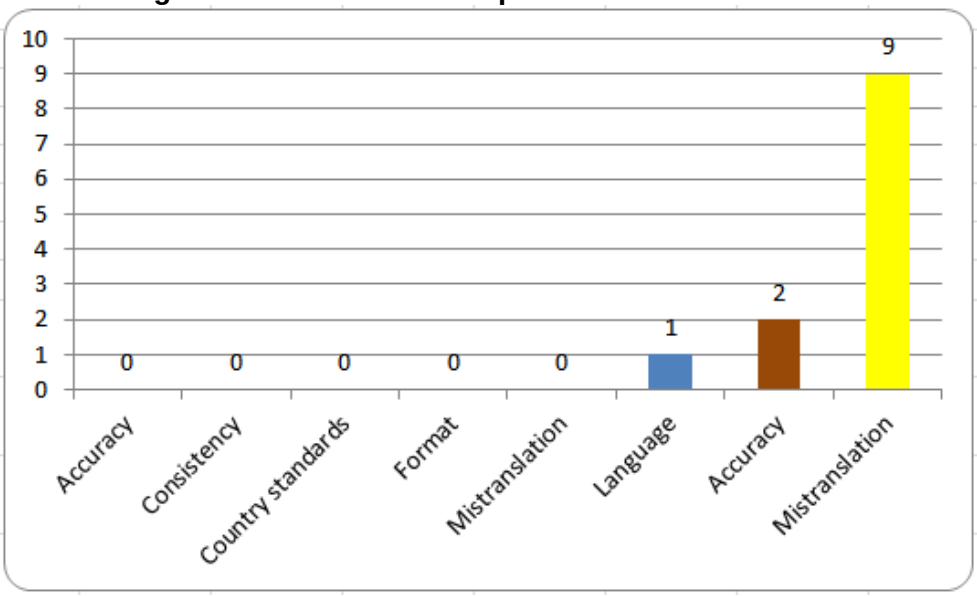
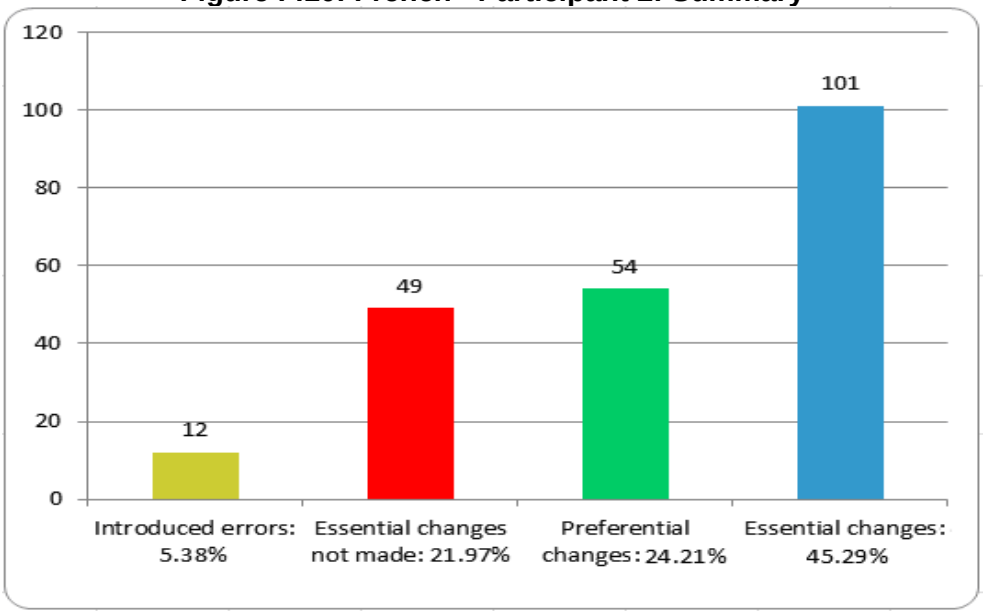


Table F.10: French - Participant 2: Summary

ESSENTIAL CHANGES	101
Preferential Changes	54
Essential Changes Not Implemented	49
Introduced Errors	12
TOTAL	223

Figure F.10: French - Participant 2: Summary



F.5. French - Participant 4

Table F.11: French - Participant 4: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	22	Accuracy - extra information: 1 Accuracy - information missing: 21
Format	8	
Language	74	Language - adjectives: 1 Language - adverbs: 3 Language - capitalisation: 10 Language - conjunctions: 2 Language - determiners: 4 Language - gender: 15 Language - number: 11 Language - phrasal ordering: 3 Language - prepositions: 7 Language - pronouns: 5 Language - punctuation: 5 Language - verbs: 8
Mistranslation	25	
Style	2	
Total	**131**	

Figure F.11: French - Participant 4: Essential Changes

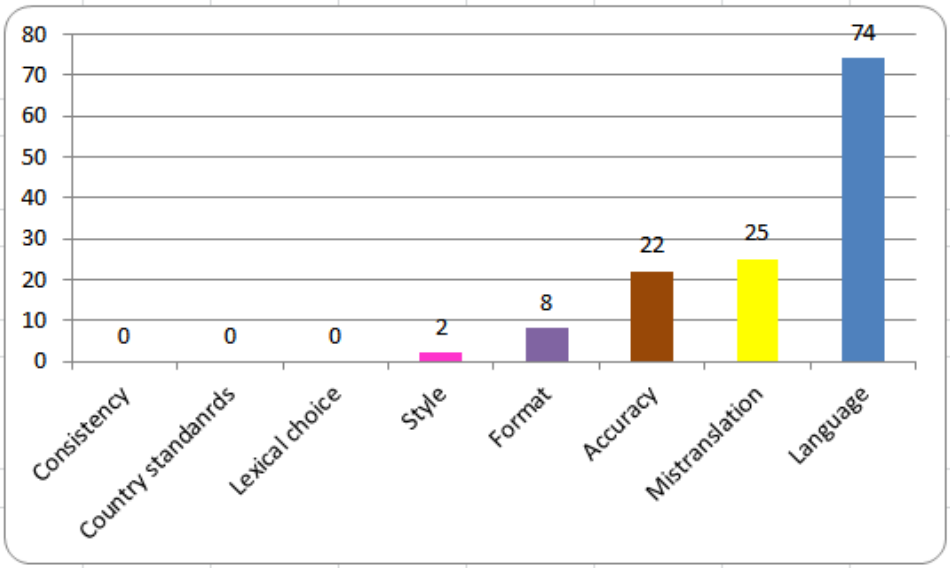


Table F.12: French - Participant 4: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	21	Language - adverbs: 1 Language - conjunctions: 1 Language - determiners: 1 Language - number: 2 Language - phrasal ordering: 1 Language - prepositions: 10 Language - pronouns: 1 Language - punctuation: 1 Language - verbs: 3
Lexical choice	21	
Style	4	
Total	**46**	

Figure F.12: French - Participant 4: Preferential Changes

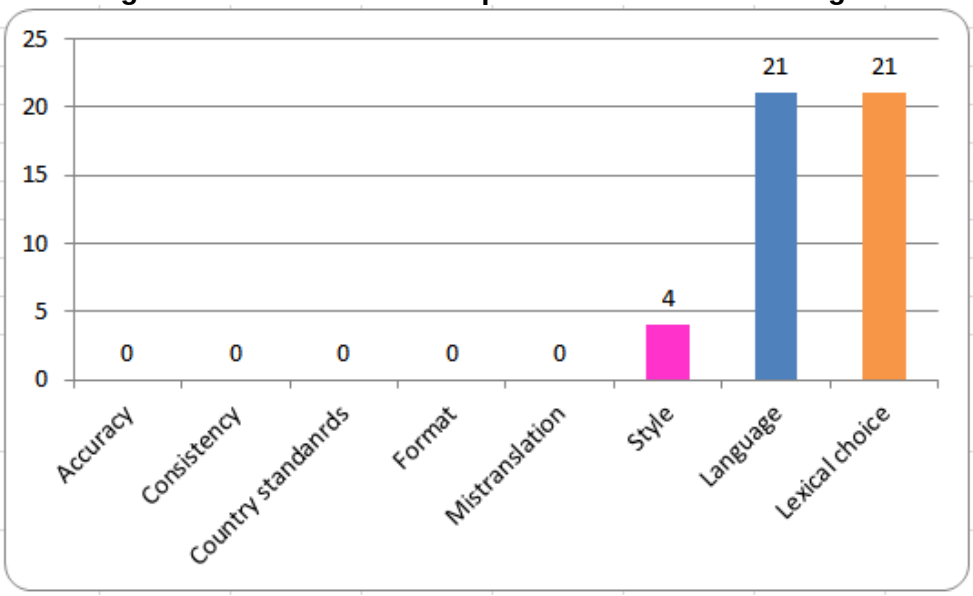


Table F.13: French - Participant 4: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	7	Accuracy - information missing: 7
Language	4	Language - determiners: 1 Language - phrasal ordering: 1 Language - pronouns: 1 Language - punctuation: 1
Mistranslation	11	
Total	**22**	

Figure F.13: French - Participant 4: Essential Changes Not Implemented

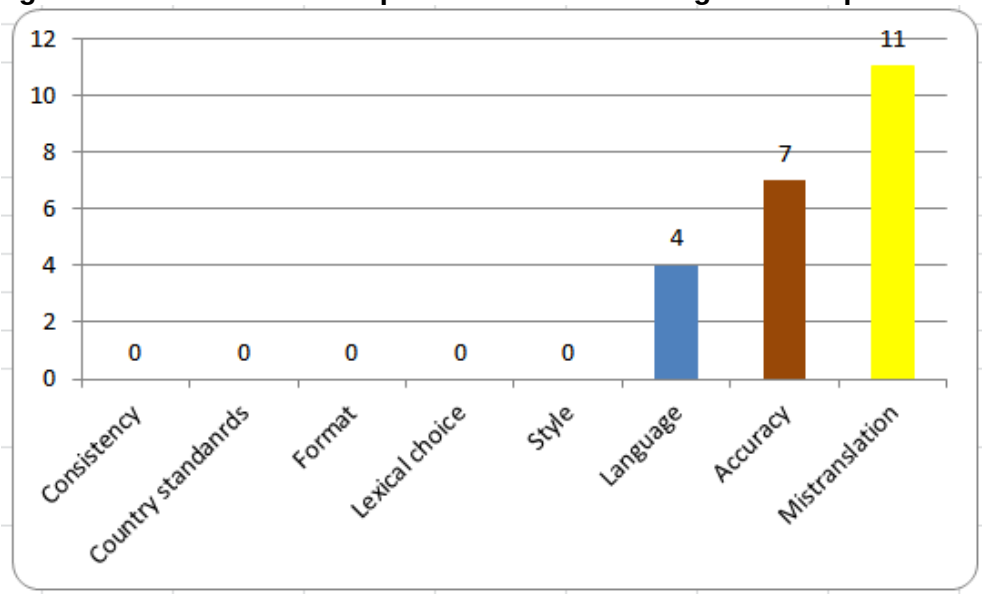


Table F.14: French - Participant 4: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	2	Accuracy - information missing: 2
Language	1	Language - determiners: 1
Mistranslation	4	
Total	**7**	

Figure F.14: French - Participant 4: Introduced Errors

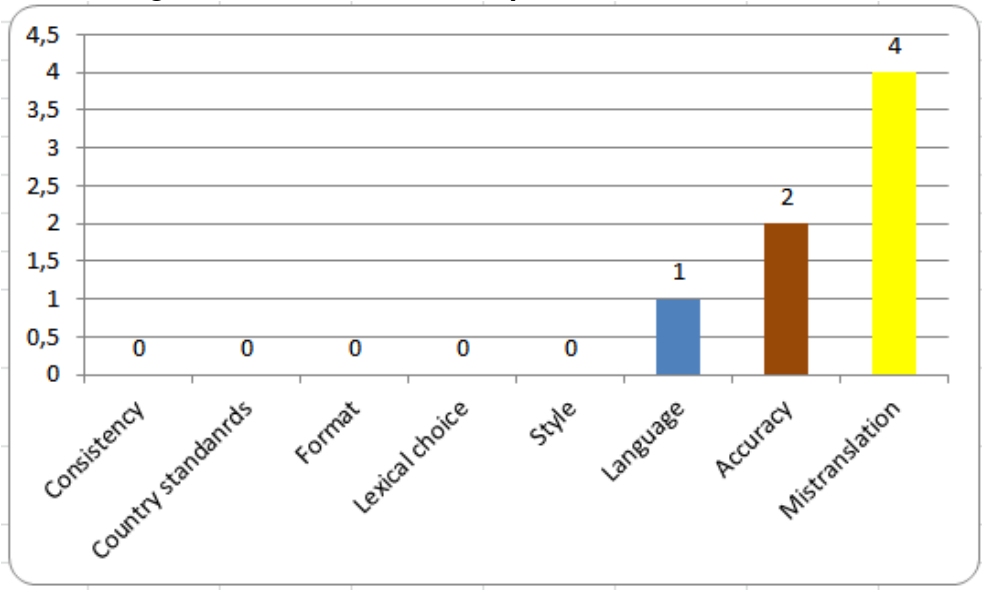
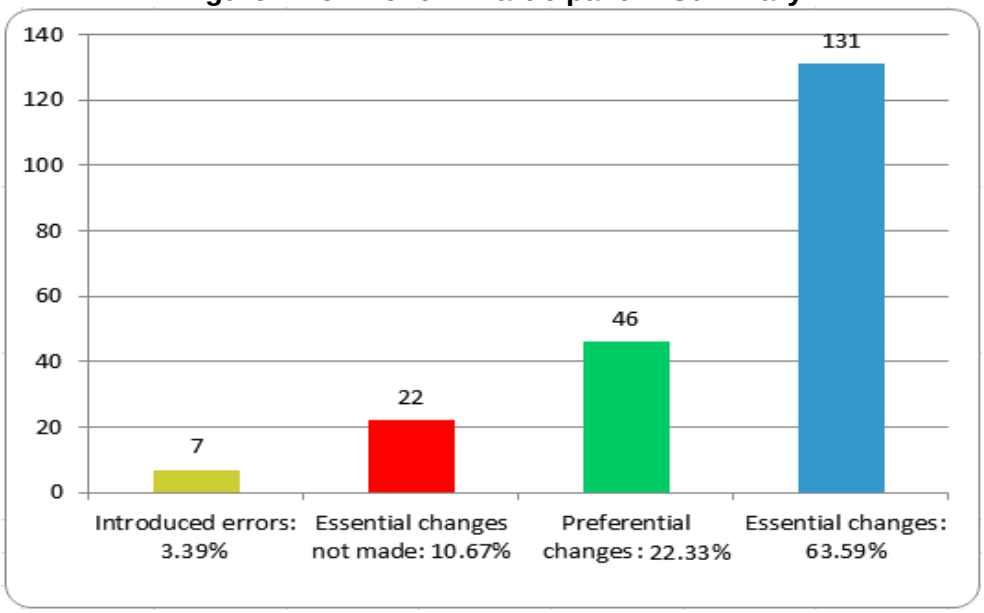


Table F.15: French - Participant 4: Summary

ESSENTIAL CHANGES	**131**
Preferential Changes	**46**
Essential Changes Not Implemented	**22**
Introduced Errors	**7**
TOTAL	**206**

Figure F.15: French - Participant 4: Summary



F.6. French - Participant 5

Table F.16: French - Participant 5: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	10	Accuracy - extra information: 1 Accuracy - information missing: 20
Format	8	
Language	59	Language - adjectives: 1 Language - adverbs: 3 Language - capitalisation: 9 Language - determiners: 4 Language - gender: 10 Language - number: 11 Language - phrasal ordering: 3 Language - prepositions: 6 Language - pronouns: 3 Language - punctuation: 4 Language - verbs: 5
Lexical choice	1	
Mistranslation	27	
Style	4	
Total	**109**	

Figure F.16: French - Participant 5: Essential Changes

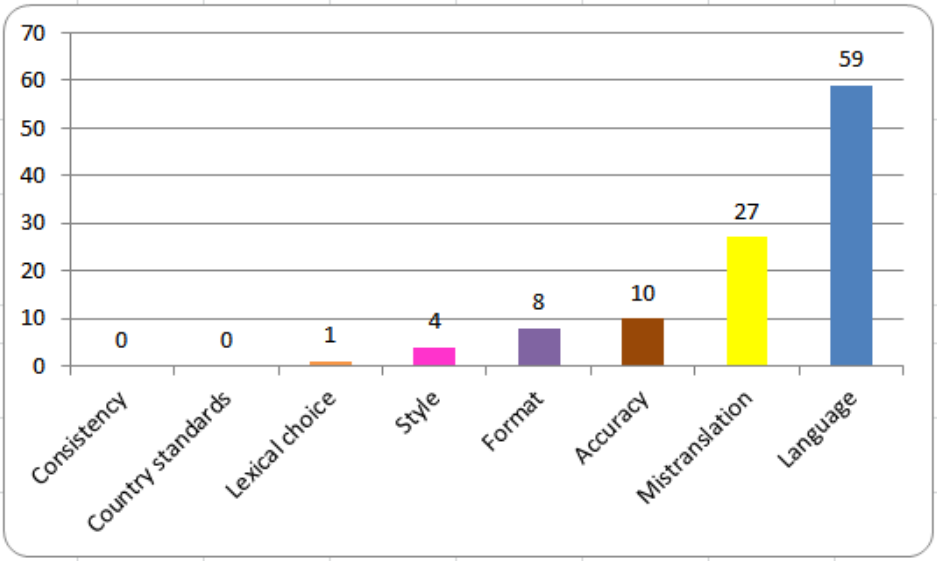


Table F.17: French - Participant 5: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	39	Language - adverbs: 1 Language - capitalisation: 1 Language - determiners: 12 Language - gender: 2 Language - number: 3 Language - phrasal ordering: 4 Language - prepositions: 12 Language - pronouns: 4
Lexical choice	36	
Style	20	
Total	**95**	

Figure F.17: French - Participant 5: Preferential Changes

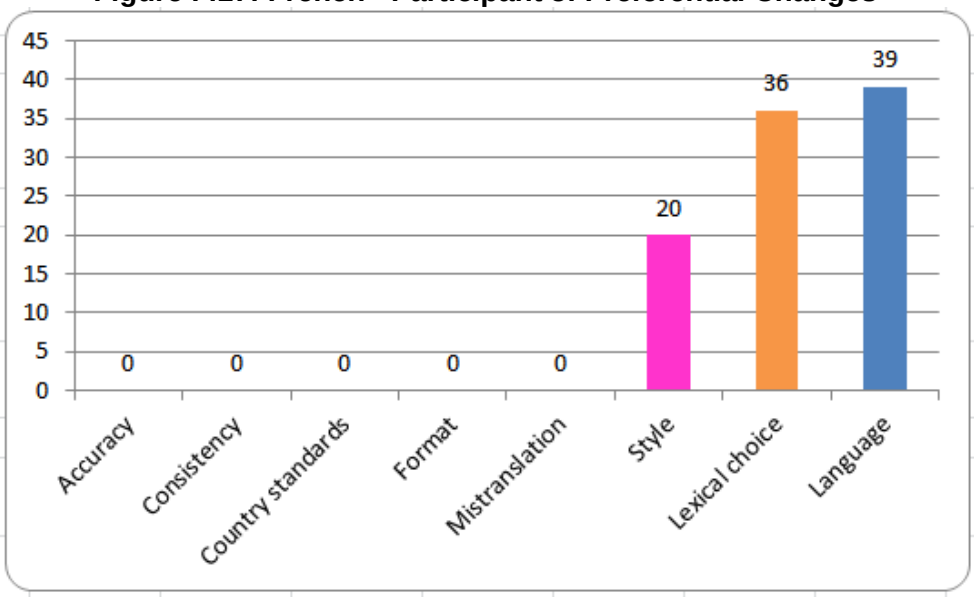


Table F.18: French - Participant 5: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	7	Accuracy - information missing: 7
Language	7	Language - capitalisation: 1 Language - determiners: 2 Language - gender: 1 Language - prepositions: 1 Language - pronouns: 2
Mistranslation	10	
Total	**24**	

Figure F.18: French - Participant 5: Essential Changes Not Implemented

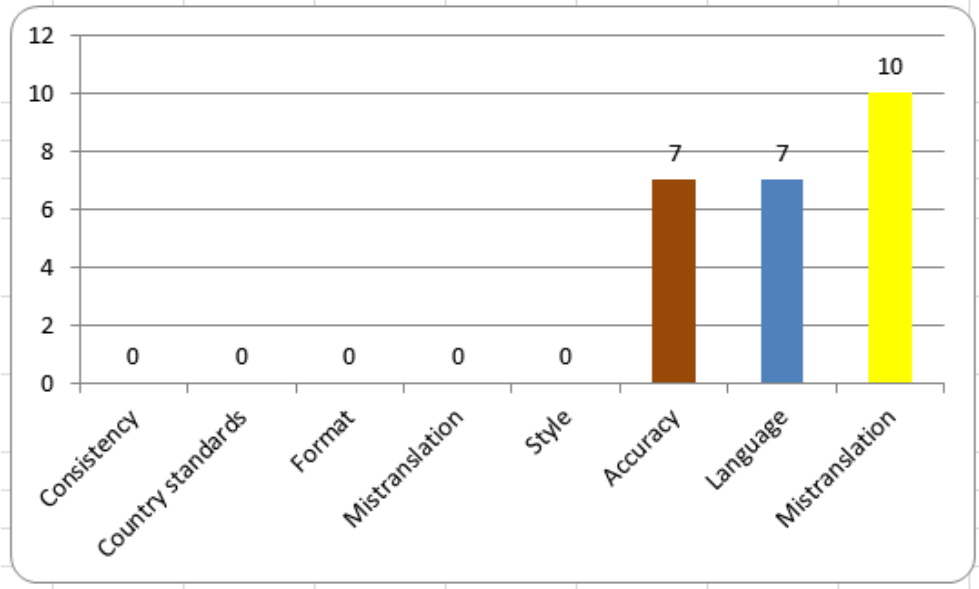


Table F.19: French - Participant 5: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	6	Accuracy - extra information: 1 Accuracy - information missing: 5
Language	1	Language - spelling: 1
Mistranslation	6	
Style	1	
Total	**14**	

Figure F.19: French - Participant 5: Introduced Errors

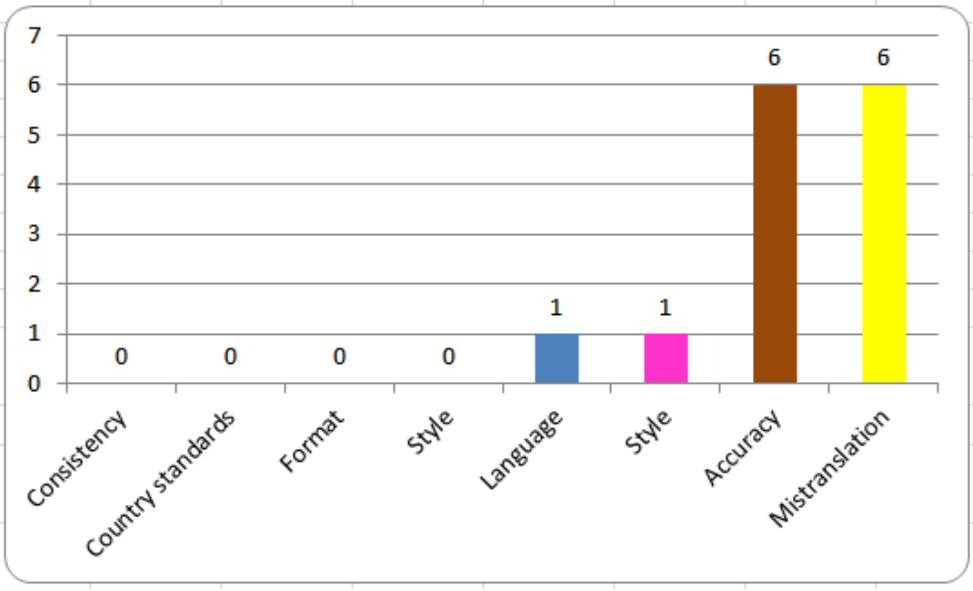
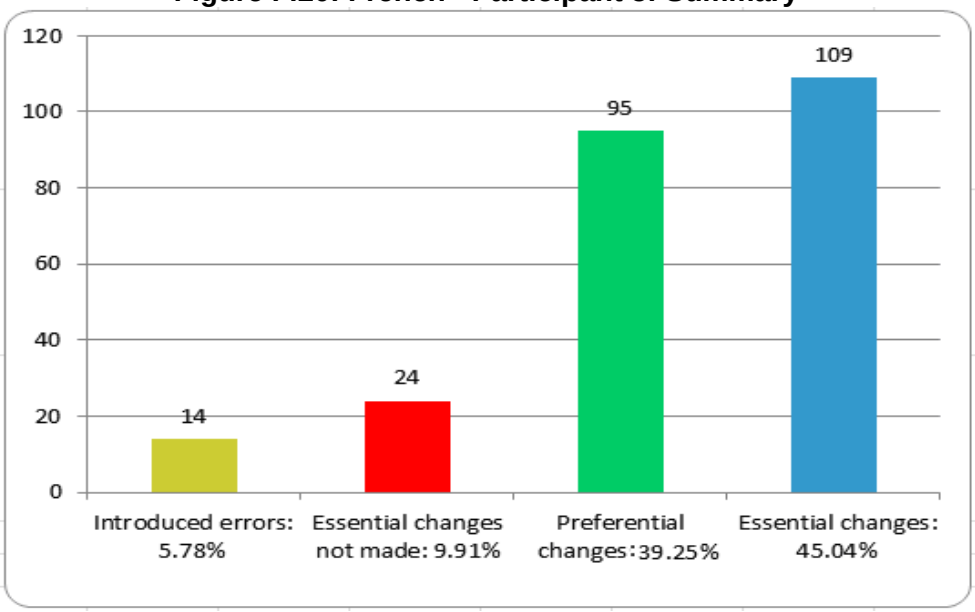


Table F.20: French - Participant 5: Summary

ESSENTIAL CHANGES	109
Preferential Changes	95
Essential Changes Not Implemented	24
Introduced Errors	14
TOTAL	242

Figure F.20: French - Participant 5: Summary



F.7. French - Participant 7

Table F.21: French - Participant 7: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	26	Accuracy - extra information: 1 Accuracy - information missing: 25
Format	8	
Language	69	Language - adverbs: 5 Language - capitalisation: 9 Language - determiners: 4 Language - gender: 14 Language - number: 11 Language - phrasal ordering: 2 Language - prepositions: 6 Language - pronouns: 6 Language - punctuation: 5 Language - verbs: 7
Mistranslation	30	
Style	2	
Total	**135**	

Figure F.21: French - Participant 7: Essential Changes

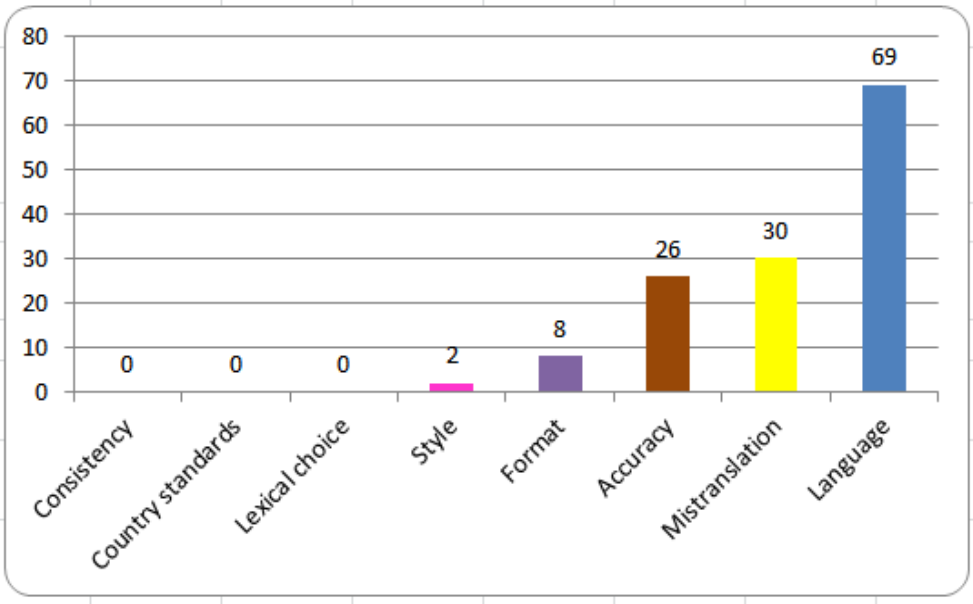


Table F.22: French - Participant 7: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	23	Language - adjectives: 1 Language - determiners: 4 Language - number: 9 Language - phrasal ordering: 1 Language - prepositions: 6 Language - punctuation: 1 Language - verbs: 1
Lexical choice	19	
Style	5	
Total	**47**	

Figure F.22: French - Participant 7: Preferential Changes

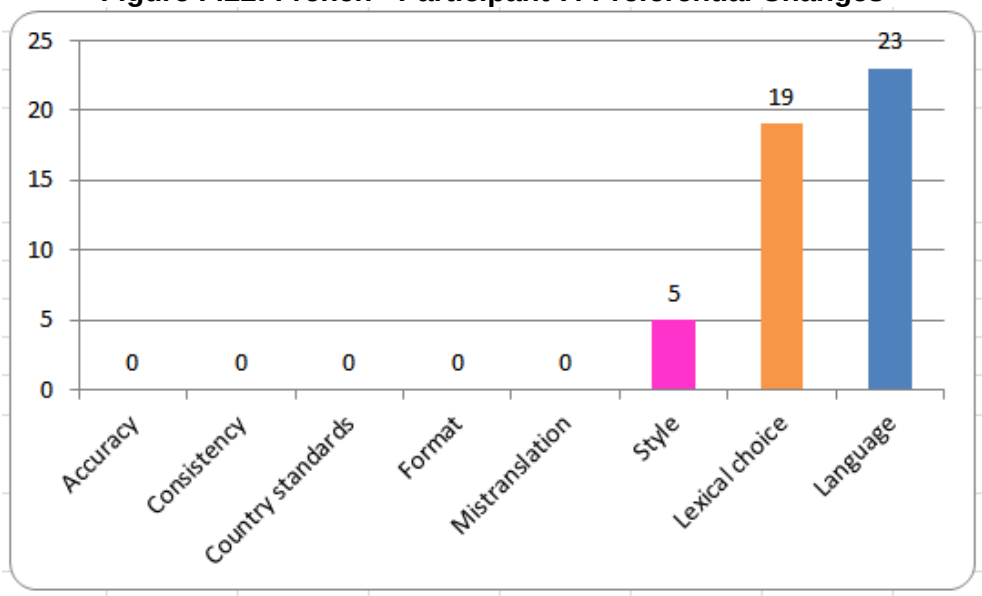


Table F.23: French - Participant 7: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	3	Accuracy - information missing: 3
Language	4	Language - capitalisation: 1 Language - determiners: 1 Language - phrasal ordering: 1 Language - punctuation: 1
Mistranslation	4	
Total	**11**	

Figure F.23: French - Participant 7: Essential Changes Not Implemented

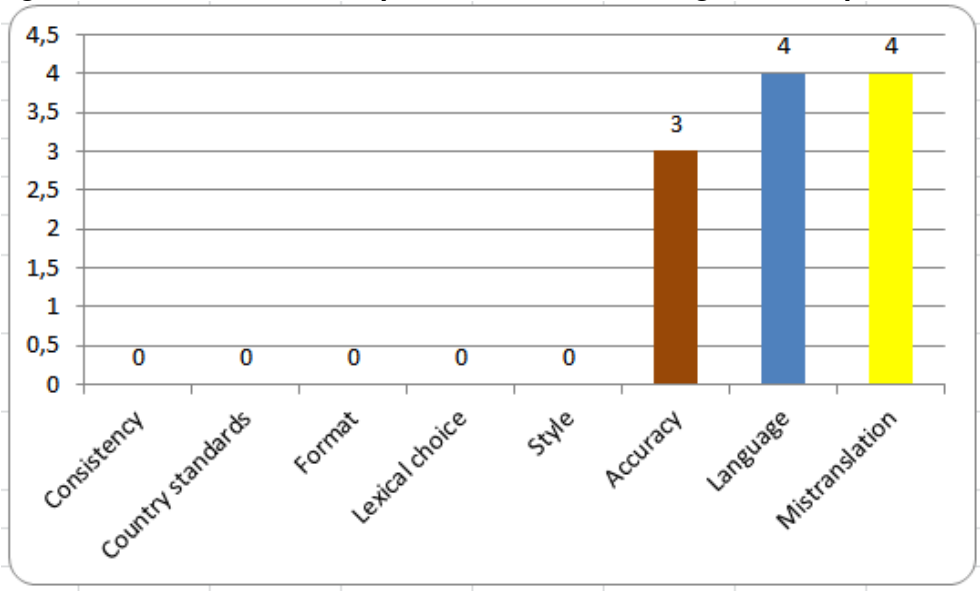


Table F.24: French - Participant 7: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	3	Accuracy - information missing: 2 Accuracy - untranslated text: 1
Language	1	Language - capitalisation: 1
Total	**4**	

Figure F.24: French - Participant 7: Introduced Errors

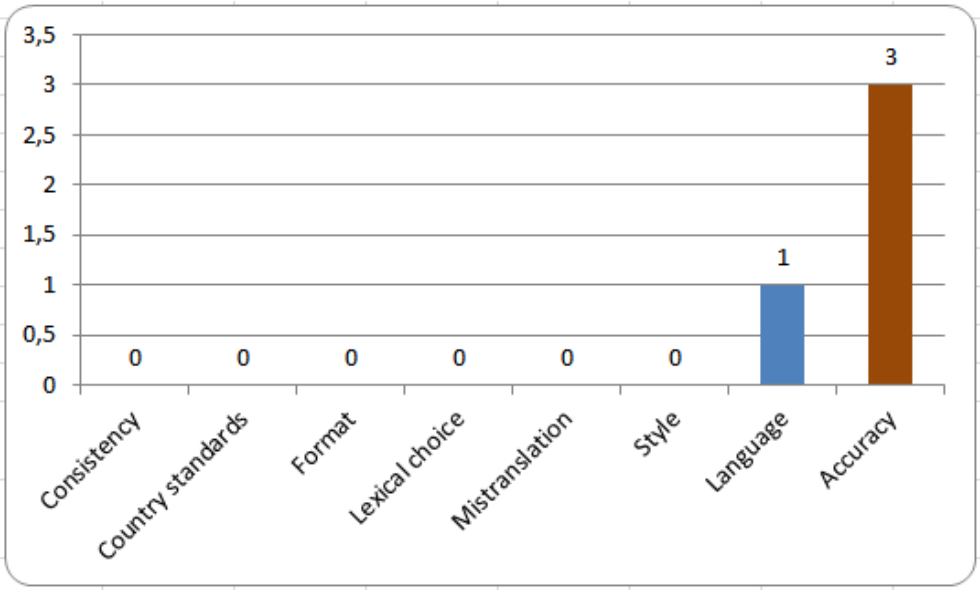
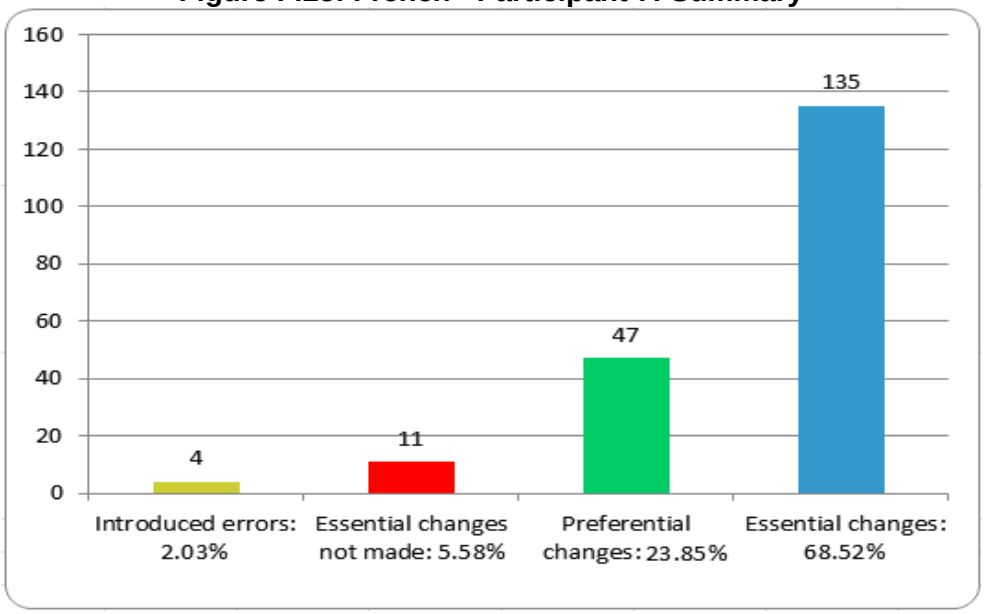


Table F.25: French - Participant 7: Summary

ESSENTIAL CHANGES	135
Preferential Changes	47
Essential Changes Not Implemented	11
Introduced Errors	4
TOTAL	197

Figure F.25: French - Participant 7: Summary



F.8. French - Participant 8

Table F.26: French - Participant 8: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	22	Accuracy - extra information: 1 Accuracy - information missing: 21
Format	5	
Language	69	Language - adverbs: 5 Language - capitalisation: 7 Language - determiners: 4 Language - gender: 14 Language - number: 13 Language - phrasal ordering: 2 Language - prepositions: 8 Language - pronouns: 5 Language - punctuation: 5 Language - verbs: 6
Mistranslation	24	
Style	3	
Total	**123**	

Figure F.26: French - Participant 8: Essential Changes

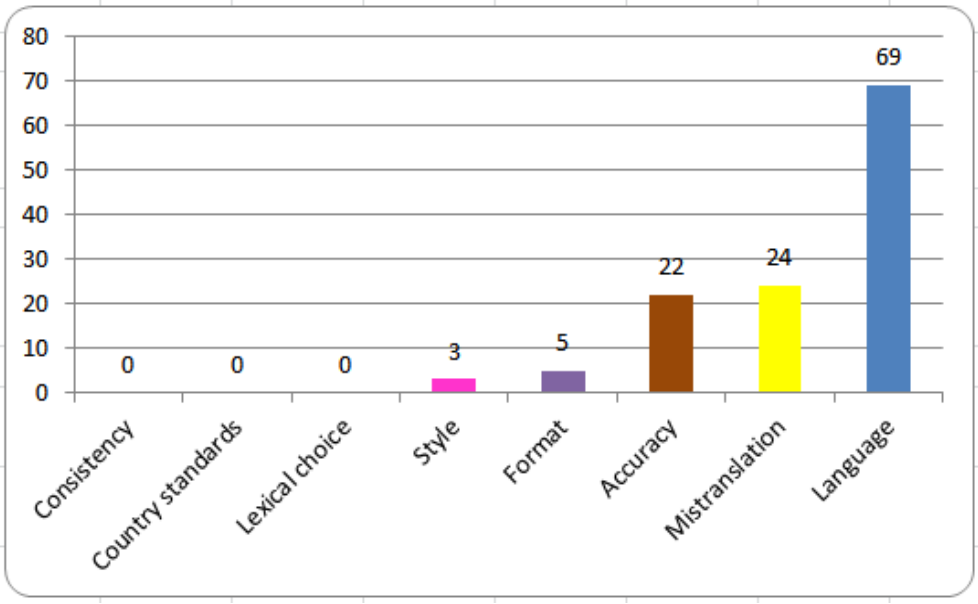


Table F.27: French - Participant 8: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	19	Language - adverbs: 1 Language - determiners: 4 Language - number: 1 Language - phrasal ordering: 1 Language - prepositions: 7 Language - pronouns: 1 Language - punctuation: 4
Lexical choice	10	
Style	9	
Total	**38**	

Figure F.27: French - Participant 8: Preferential Changes

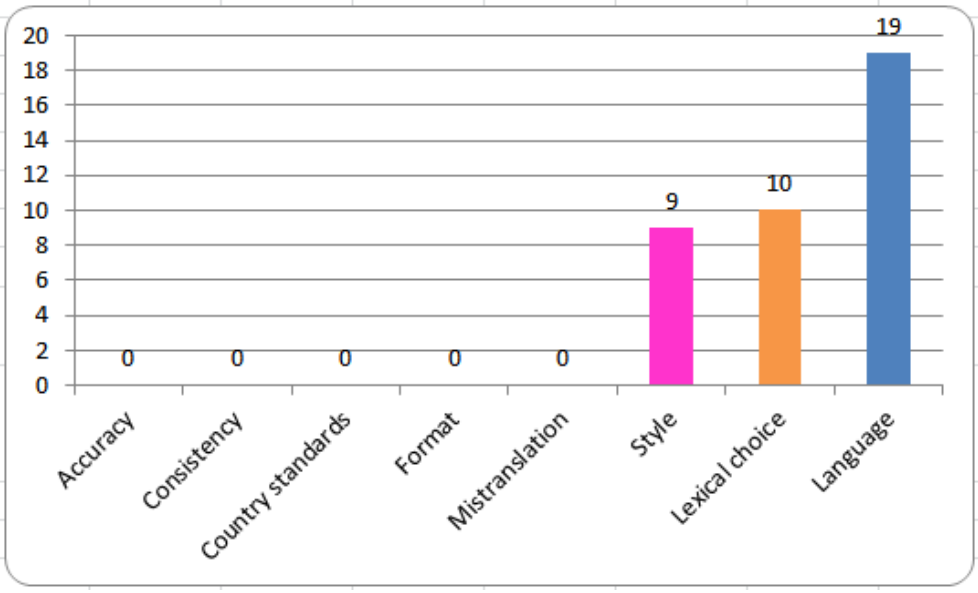


Table F.28: French - Participant 8: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	7	Accuracy- missing information: 7
Language	7	Language - capitalisation: 3 Language - determiners: 1 Language - phrasal ordering: 1 Language - pronouns: 1 Language - punctuation: 1
Mistranslation	10	
Total	**24**	

Figure F.28: French - Participant 8: Essential Changes Not Implemented

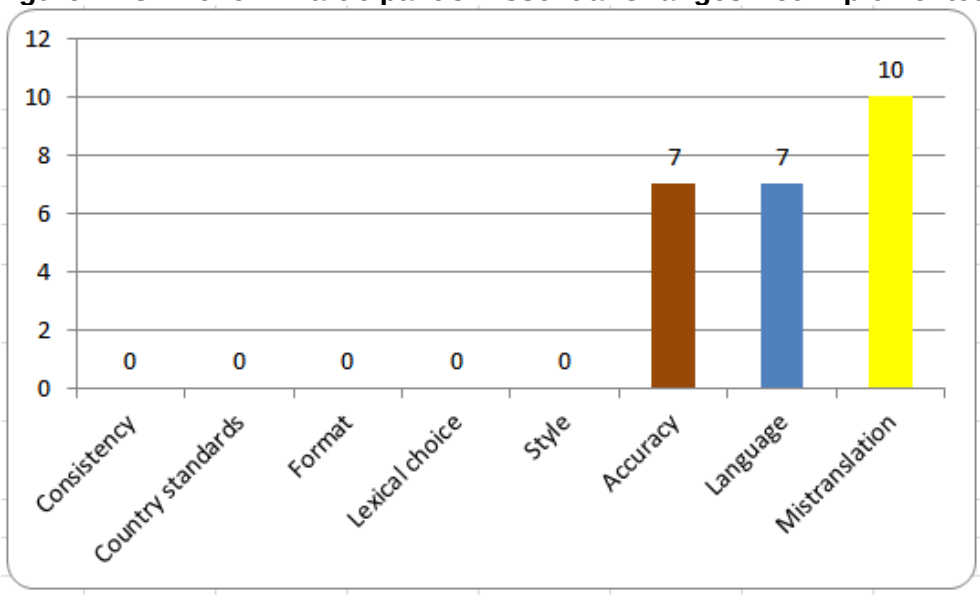


Table F.29: French - Participant 8: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	1	Accuracy - missing information: 1
Language	2	Language - number: 2
Mistranslation	3	
Total	**6**	

Figure F.29: French - Participant 8: Introduced Errors

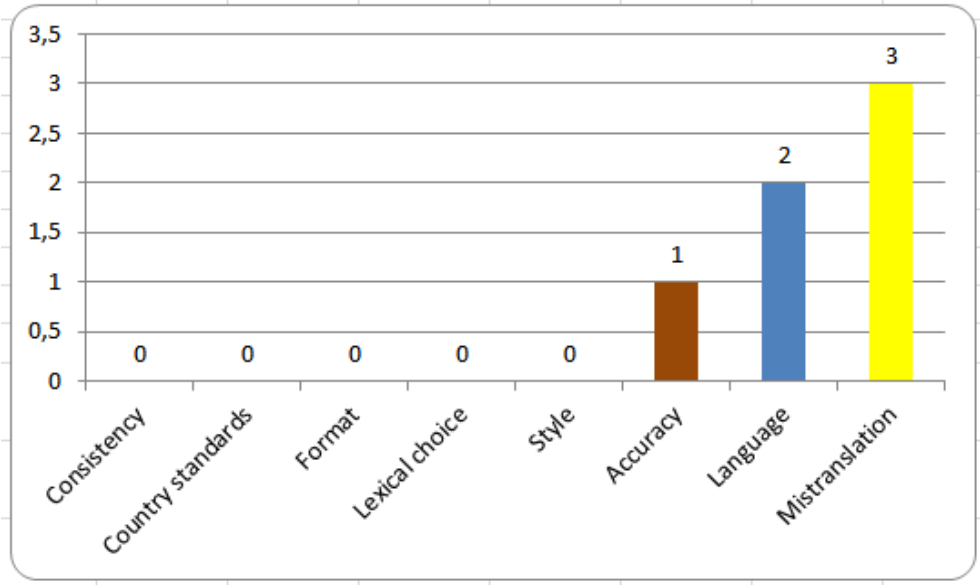
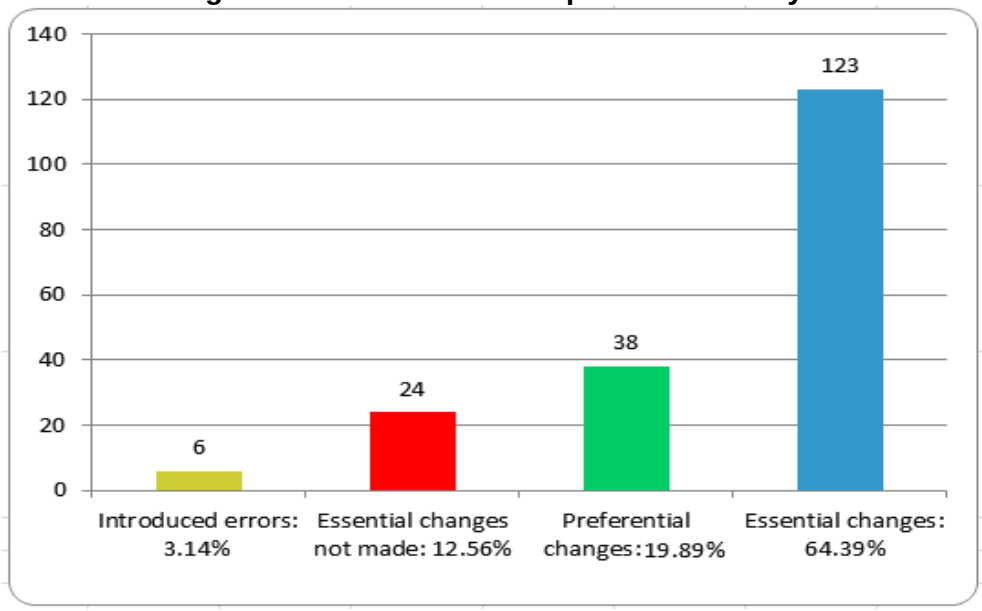


Table F.30: French - Participant 8: Summary

ESSENTIAL CHANGES	123
Preferential Changes	38
Essential Changes Not Implemented	24
Introduced Errors	6
TOTAL	191

Figure F.30: French - Participant 8: Summary



F.9. French - Participant 9

Table F.31: French - Participant 9: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	24	Accuracy - extra information: 1 Accuracy - information missing: 23
Format	8	
Language	66	Language - adverbs: 4 Language - capitalisation: 9 Language - conjunctions: 1 Language - determiners: 5 Language - gender: 14 Language - number: 9 Language - phrasal ordering: 3 Language - prepositions: 5 Language - pronouns: 5 Language - punctuation: 6 Language - verbs: 5
Mistranslation	26	
Style	6	
Total	**130**	

Figure F.31: French - Participant 9: Essential Changes

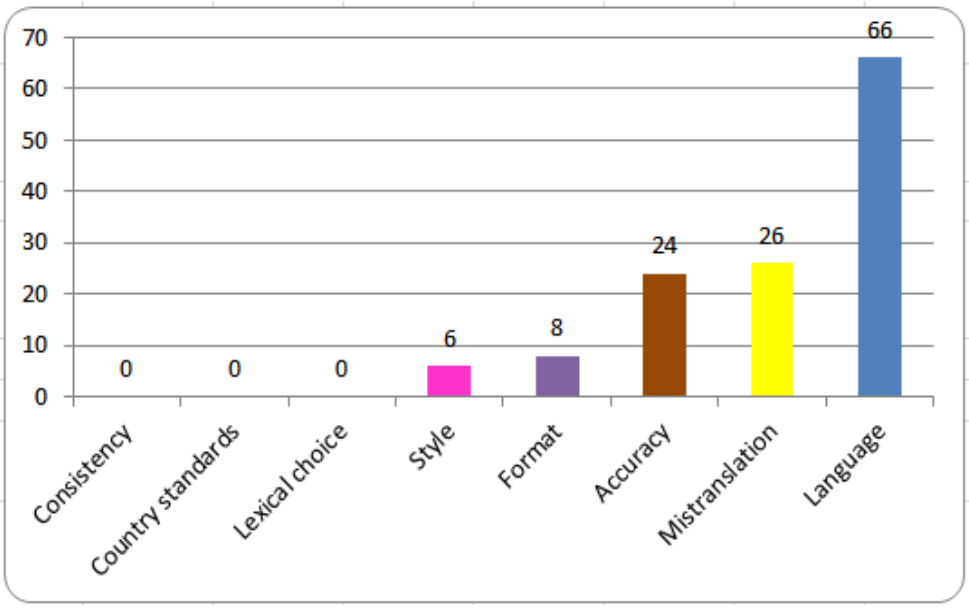


Table F.32: French - Participant 9: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	23	Language - adverbs: 2 Language - conjunctions: 1 Language - determiners: 5 Language - phrasal ordering: 4 Language - prepositions: 7 Language - pronouns: 2 Language - punctuation: 1 Language - verbs: 1
Lexical choice	18	
Style	10	
Total	**51**	

Figure F.32: French - Participant 9: Preferential Changes

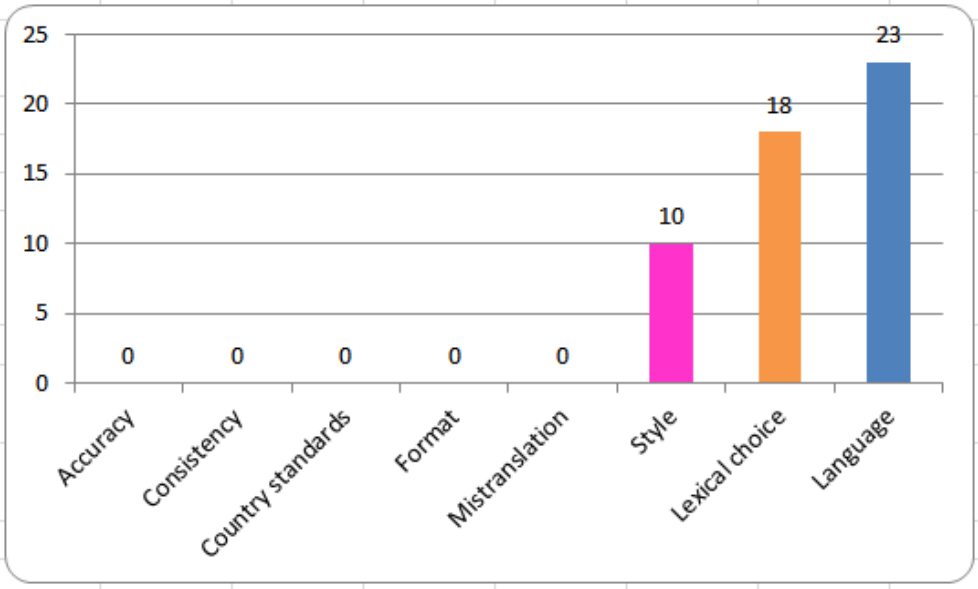


Table F.33: French - Participant 9: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	6	Accuracy - information missing: 6
Language	5	Language - capitalisation: 2 Language - number: 1 Language - prepositions: 1 Language - pronouns: 1
Mistranslation	10	
Total	**21**	

Figure F.33: French - Participant 9: Essential Changes Not Implemented

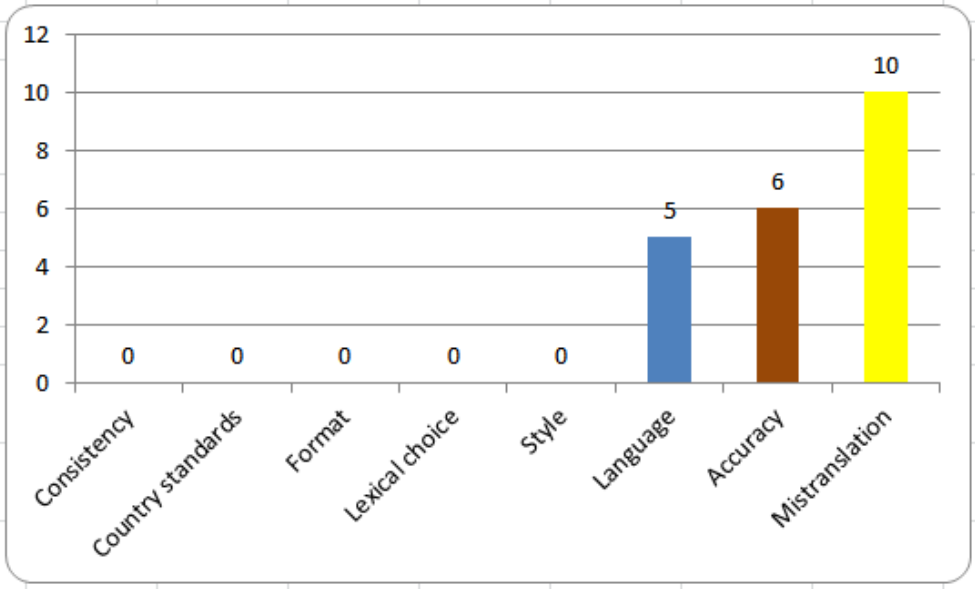


Table F.34: French - Participant 9: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	1	Accuracy - information missing: 1
Language	4	Language - number: 4
Mistranslation	3	
Total	**8**	

Figure F.34: French - Participant 9: Introduced Errors

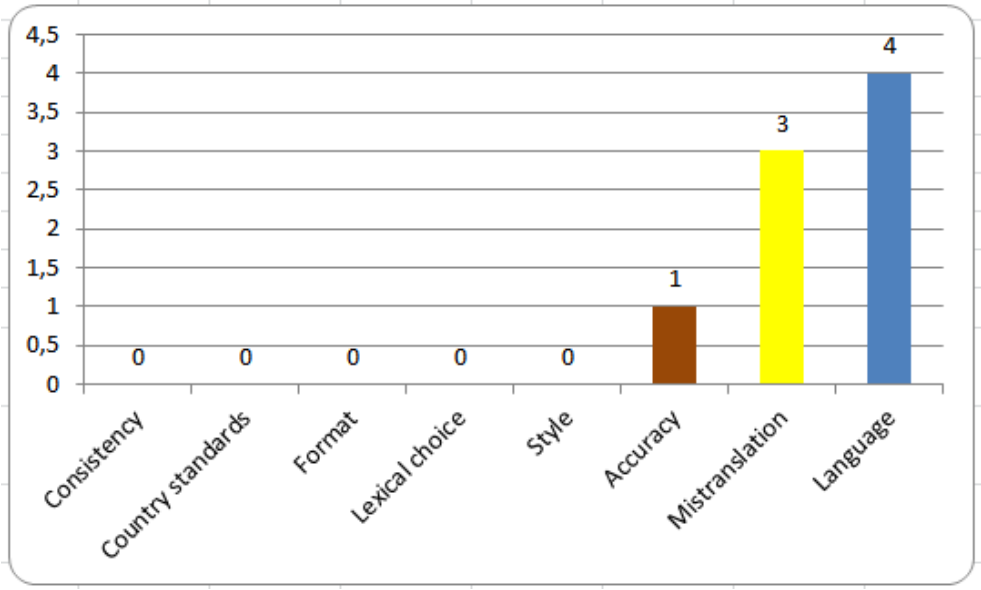
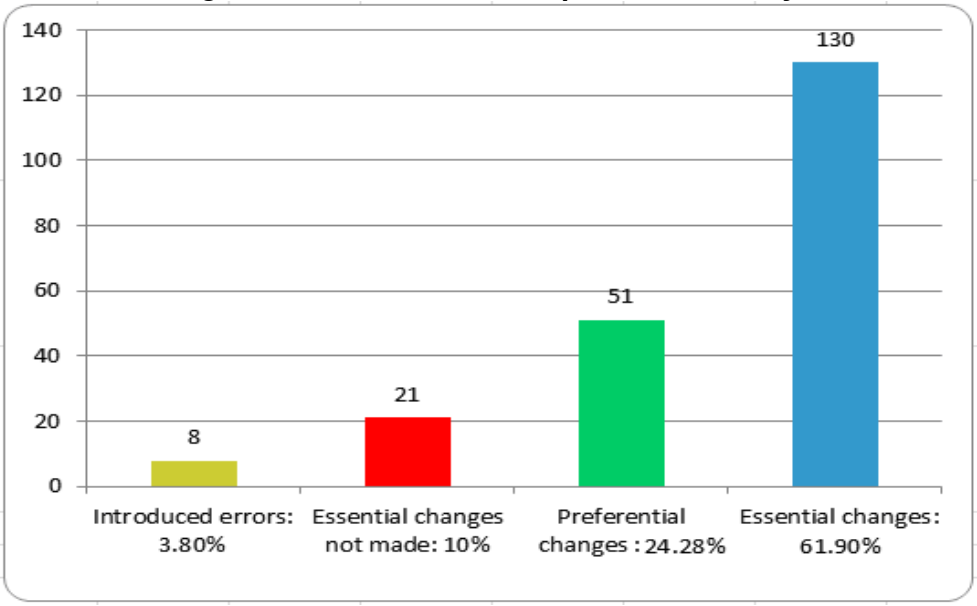


Table F.35: French - Participant 9: Summary

ESSENTIAL CHANGES	130
Preferential Changes	51
Essential Changes Not Implemented	21
Introduced Errors	8
TOTAL	210

Figure F.35: French - Participant 9: Summary



F.10. French - Participant 10

Table F.36: French - Participant 10: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	24	Accuracy - extra information: 1 Accuracy - information missing: 23
Format	8	
Language	68	Language - adverbs: 4 Language - capitalisation: 9 Language - conjunctions: 1 Language - determiners: 4 Language - gender: 12 Language - number: 10 Language - phrasal ordering: 3 Language - prepositions: 7 Language - pronouns: 5 Language - punctuation: 6 Language - verbs: 7
Mistranslation	32	
Style	3	
Total	**135**	

Figure F.36: French - Participant 10: Essential Changes

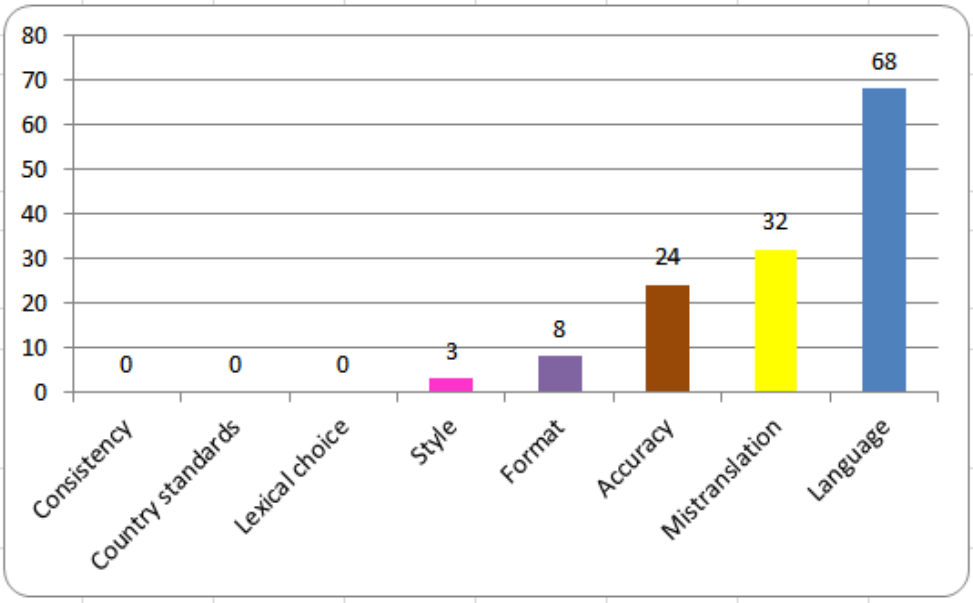


Table F.37: French - Participant 10: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	28	Language - adverbs: 1 Language - conjunctions: 1 Language - determiners: 6 Language - number: 1 Language - phrasal ordering: 3 Language - prepositions: 9 Language - pronouns: 2 Language - punctuation: 4 Language - verbs: 1
Lexical choice	14	
Style	10	
Total	**52**	

Figure F.37: French - Participant 10: Preferential Changes

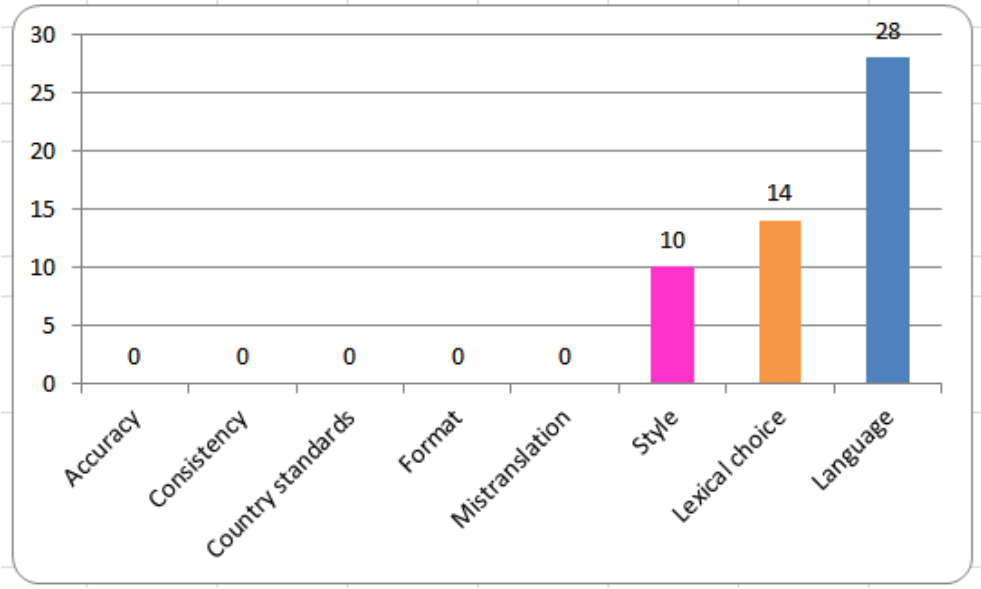


Table F.38: French - Participant 10: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	5	Accuracy - information missing: 5
Language	2	Language - capitalisation: 1 Language - determiners: 1
Mistranslation	2	
Total	**9**	

Figure F.38: French - Participant 10: Essential Changes Not Implemented

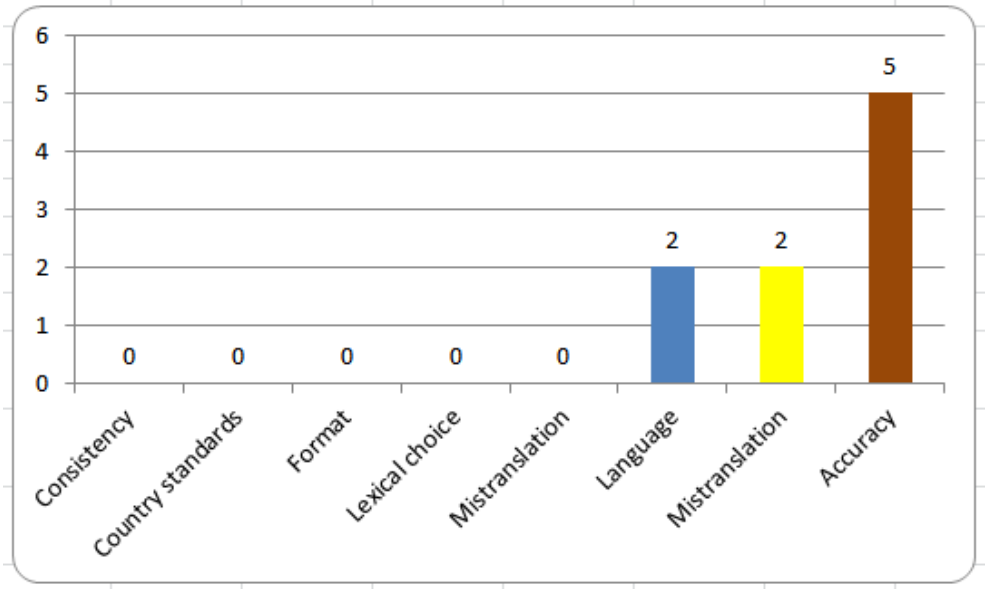


Table F.39: French - Participant 10: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Language	5	Language - number: 3 Language - prepositions: 2
Mistranslation	4	
Total	**9**	

Figure F.39: French - Participant 10: Introduced Errors

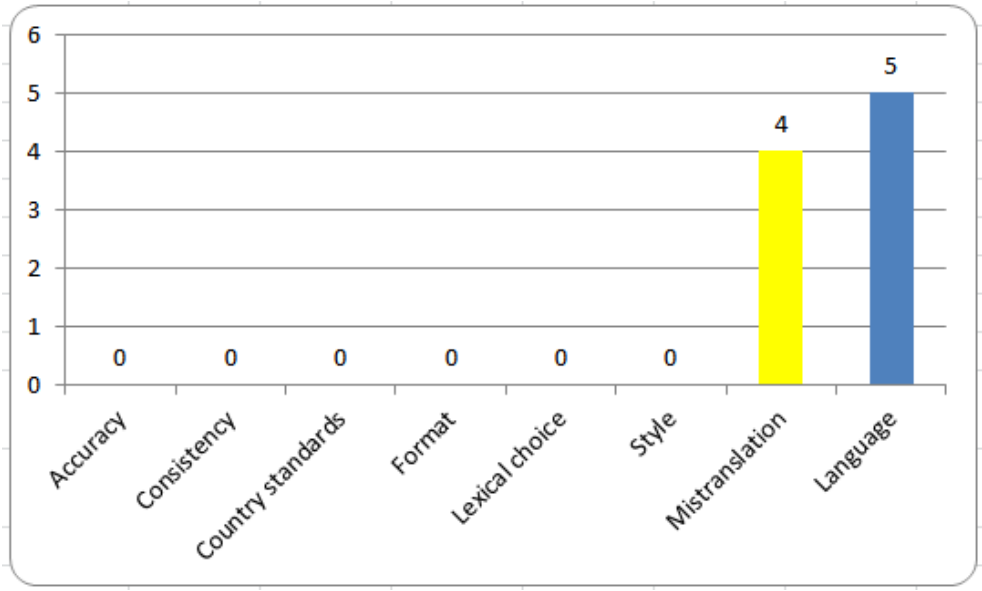
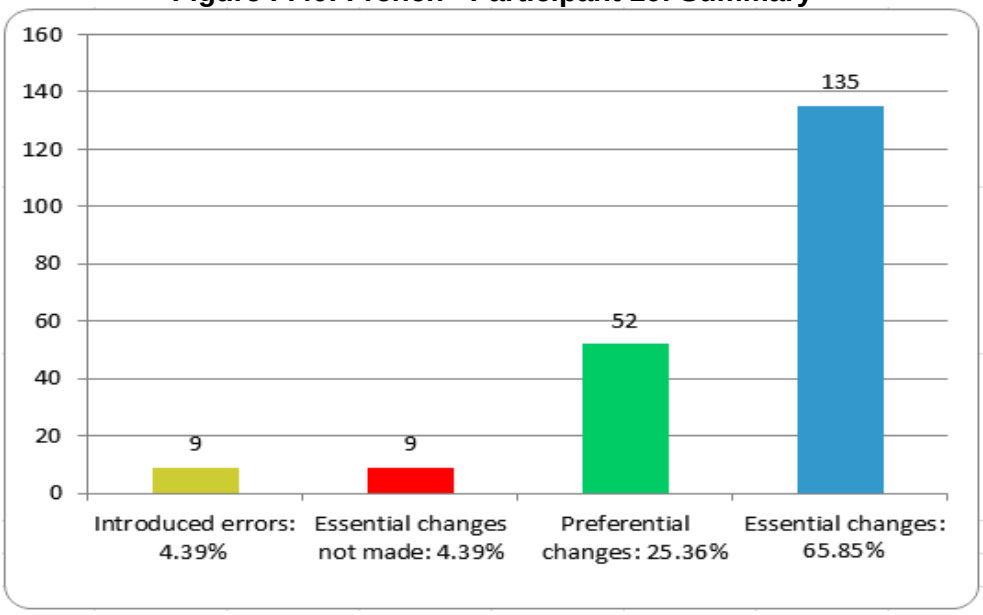


Table F.40: French - Participant 10: Summary

ESSENTIAL CHANGES	**135**
Preferential Changes	**52**
Essential Changes Not Implemented	**9**
Introduced Errors	**9**
TOTAL	**205**

Figure F.40: French - Participant 10: Summary



F.11. Summary for all participants combined - French

Figure F.41: Summary for all participants combined - French

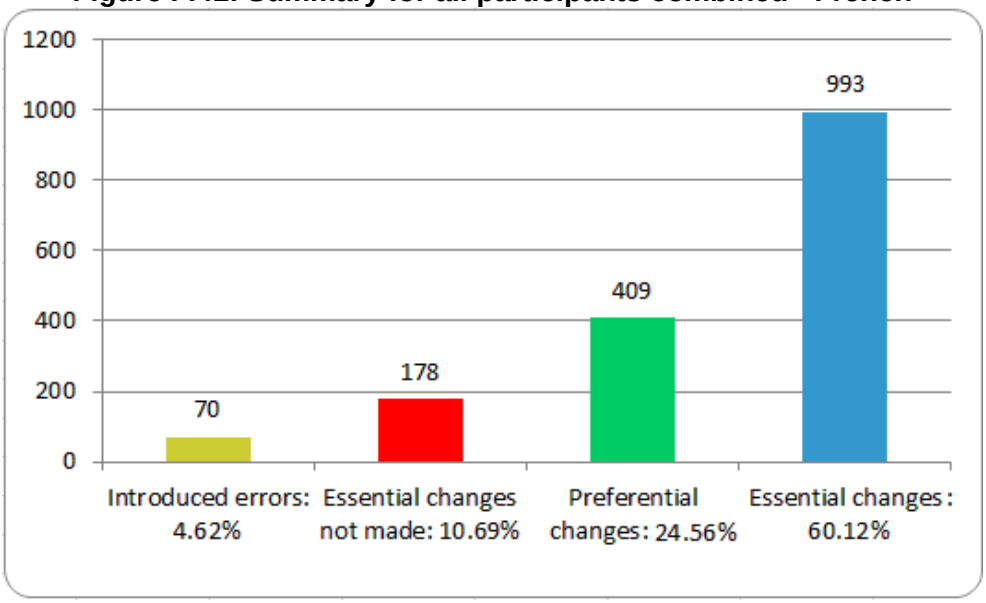
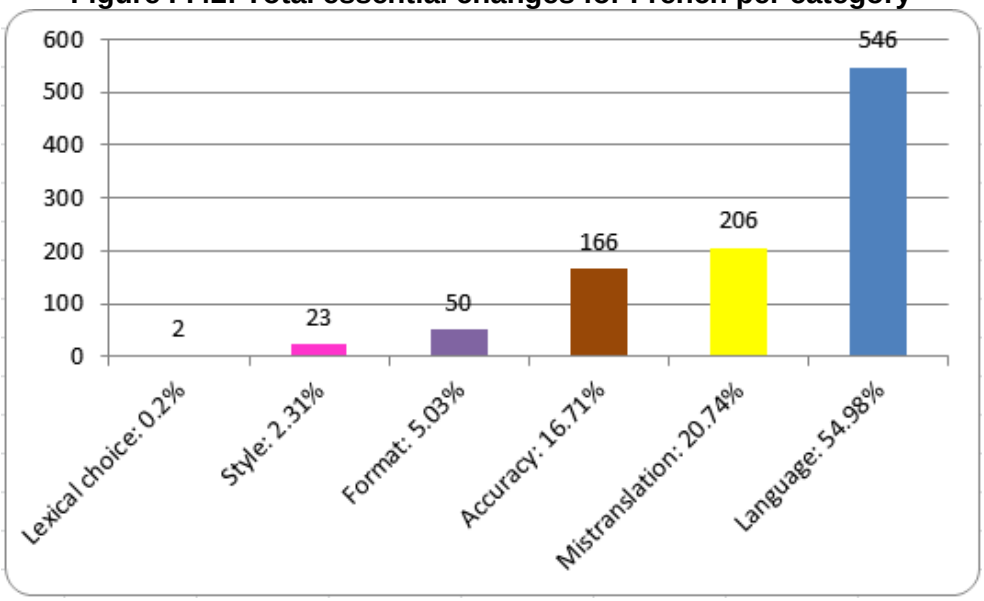


Figure F.42: Total essential changes for French per category



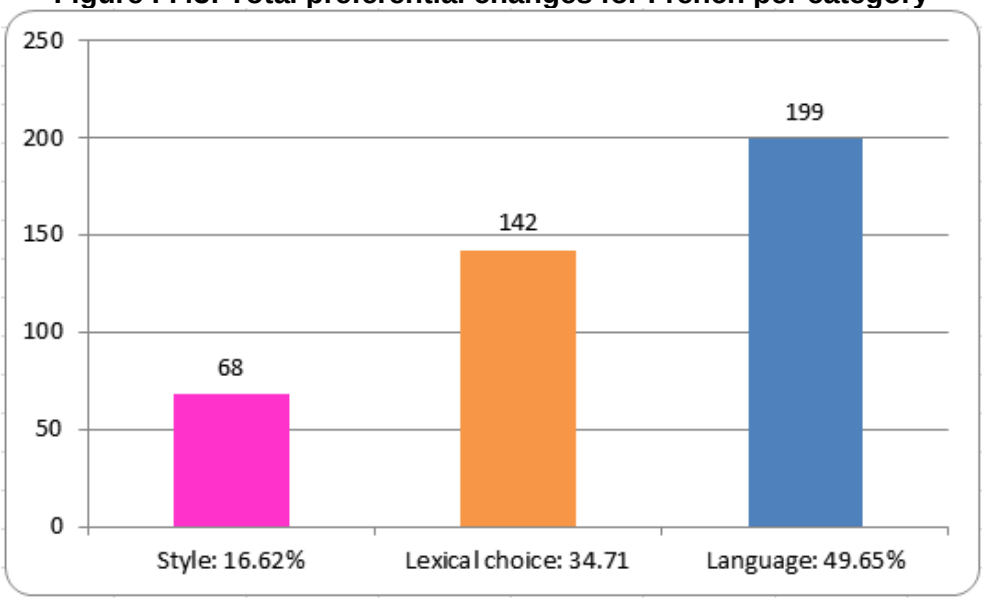
Total of essential changes recorded: 993

Table F.41: Examples of Essential Changes for French

Source text	Raw MT output	Post-edited text	Comments
How to move components	**comment** déplacer des composants	**Comment** déplacer des composants	Language - Capitalisation: The word "Comment" was not capitalised in the raw MT output, so this was corrected by the post-editor.
Components may also be distributed	Les composants peuvent également être **distribuée**	Les composants peuvent également être **distribués**	Language - Gender and Language - Number: The adjective "distribuée" (feminine, singular) was corrected to "distribués" (masculine, plural) to agree in gender and number with the noun to which it refers, "composants".
The selected components rotate	Les composants sélectionnés **Rotation**	Les composants sélectionnés pivotent	Mistranslation: The verb "rotate" was mistranslated as the noun "Rotate" in the raw MT output, so it was corrected to the verb "pivotent" by the post-editor.
The Align command lets you align multiple components	L'**option** Aligner vous permet d'aligner plusieurs composants	La **commande** Aligner vous permet d'aligner plusieurs composants	Mistranslation: The term "command" was mistranslated as "option" in the raw MT output, so it was corrected to "commande" by the post-editor.

Source text	Raw MT output	Post-edited text	Comments
Although the primary orientation is defined, you might have to rotate it for certain applications.	Bien que l'orientation **est** définie, vous pouvez être amené à **le** faire pivoter certaines applications.	Bien que l'orientation **principale soit** définie, vous pouvez être amené à **la** faire pivoter **pour** certaines applications.	Accuracy - Information Missing: The adjective "principale", which was missing from the raw MT output, was added by the post-editor. Language - Verb: The verb "est" was corrected from the present of the indicative to the present of the subjunctive. Language - Determiner: The article "le" was corrected to "la" to agree in gender with "orientation". Accuracy - information missing: The preposition "pour", which was missing from the raw MT output, was added by the post-editor.
{160}Restore Defaults{161}	Restaurer valeurs par défaut**{161} {160}**	**{160}**Restaurer valeurs par défaut**{161}**	Format: The order of the tags was corrected.

Figure F.43: Total preferential changes for French per category

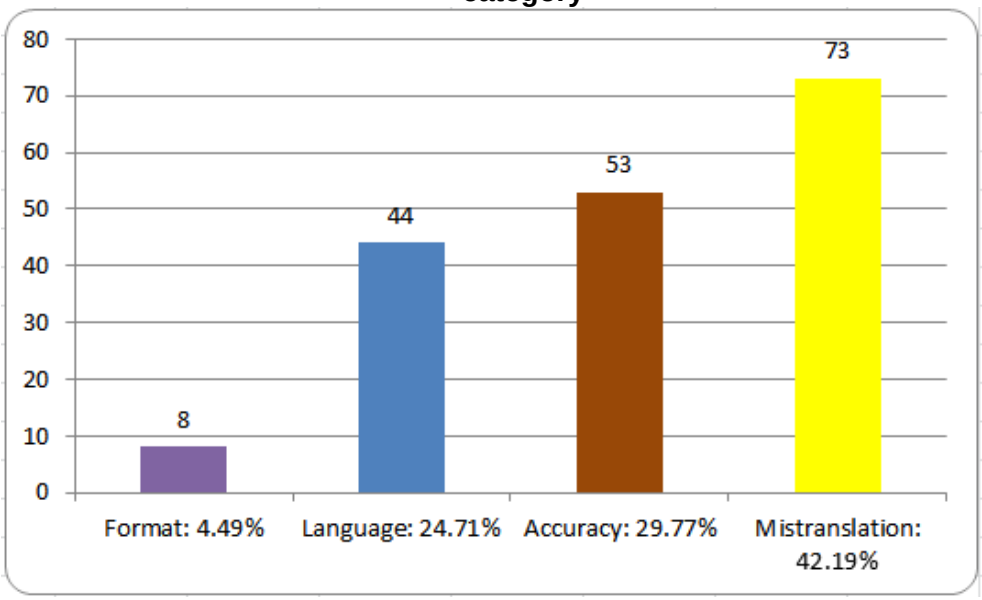


Total of preferential changes recorded: 409

Table F.42: examples of preferential changes for French

Source text	Raw MT output	Post-edited text	Comments
Select any of the parameter values in the right-hand column and enter a new value.	Sélectionnez l'une des valeurs de paramètre dans la colonne de droite et **entrez** une nouvelle valeur.	Sélectionnez l'une des valeurs de paramètre dans la colonne de droite et **saisissez** la nouvelle valeur.	Lexical Choice: The verb "entrez" was unnecessarily replaced with the verb "saisissez**".**
How to modify parameter values	**Comment modifier les** valeurs des paramètres	**Modification des** valeurs des paramètres	Style: The style of the sentence in the raw MT output was unnecessarily changed in the post-edited text.
You can navigate through the data by selecting folders or you can do a property search.	Vous pouvez parcourir les données en sélectionnant des dossiers ou **vous pouvez** effectuer une recherche de propriété.	Vous pouvez parcourir les données en sélectionnant des dossiers ou effectuer une recherche de propriété.	Style: "Vous pouvez" was unnecessarily removed from the post-edited text.
Select this option to set the length and width of the floor manually.	Sélectionnez cette option pour définir la longueur et la largeur du sol **manuellement**.	Sélectionnez cette option pour définir **manuellement** la longueur et la largeur du sol.	Language - Phrasal Ordering: The placement of the adverb "manuellement" in the sentence was unnecessarily changed.
If the Default project is active the files are all stored in the My Documents folder.	Si le projet par défaut est active, les fichiers sont **tous stockés** dans le dossier Mes documents.	Si le projet par défaut est actif, **tous** les fichiers sont **sauvegardés** dans le dossier Mes documents.	Language - Phrasal Ordering: The placement of "tous" was unnecessarily changed. Lexical Choice: "stockés" was unnecessarily replaced with "sauvegardés".

Figure F.44: Total essential changes not implemented for French per category

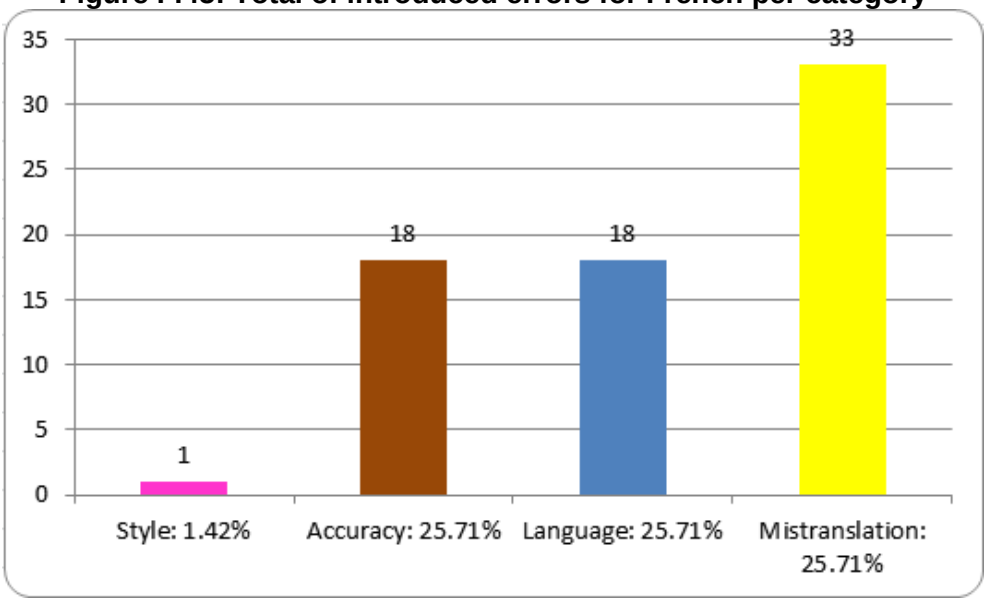


Total of essential changes not implemented recorded: 178

Table F.43: examples of essential changes not implemented for French

Source text	Raw MT output	Post-edited text	Comments
The option to turn on or off the visibility of **selected** drawing layers is also provided.	L'option permettant d'activer ou de désactiver la visibilité des calques du dessin sont également fournies.	L'option permettant d'activer ou de désactiver la visibilité des calques du dessin est également fournie.	Accuracy - Information Missing: The term "selected" is missing from the raw MT output, so "sélectionnés" should have been added after "dessin".
The floor is resized such that components always remain positioned on the floor.	Le sol est redimensionnée, tels que les composants restent toujours **placé** sur le sol.	Le sol est redimensionné de sorte que les composants restent toujours **placé** sur le sol.	Language - Number: The term "placé" should have been corrected to "placés" to agree in number with "composants".
{134}User Defined{135}	{134}{135}défini par l'utilisateur	{134}{135}défini par l'utilisateur	Format: The second formatting tag should have been placed after the text. Language - Capitalisation: The term "défini" should have been capitalised.
A list of **valid** search terms is dynamically displayed.	Une liste de termes de recherche s'**affichent** dynamiquement.	Une liste de termes de recherche s'affiche dynamiquement.	Accuracy - Information Missing: The term "valid" is missing from the raw MT output, so "valides" should have been added after "recherche".

Figure F.45: Total of introduced errors for French per category



Total of introduced errors recorded: 70

Table F.44: examples of introduced errors for French

Source text	Raw MT output	Post-edited text	Comments
By default, the floor is specified as visible in a new factory layout.	Par défaut, le sol est spécifiée comme visibles dans une nouvelle **famille**.	Par défaut, le sol est spécifié comme visible dans les nouveaux **layout Factory**.	Mistranslation: "A new factory layout" was mistranslated as "une nouvelle famille" in the raw MT output. Instead of correcting it, the post-editor replaced it with another mistranslation, "les nouveaux layout Factory".
The black and white buttons	Les boutons **en** noir et blanc	Les boutons **em** noir et blanc	Language - Spelling: The post-editor misspelled "en" as "em".
The option to turn on or off the visibility of selected **drawing layers** is also provided.	L'option permettant d'activer ou de désactiver la visibilité **des calques du dessin** sont également fournies.	L'option permettant d'activer ou de désactiver la visibilité de certaines **barres d'outils Dessin** est également fournie.	Mistranslation: The post-editor incorrectly replaced "calques du dessin" with "barres d'outils Dessin", which does not correspond to the source text.
right-click over the New Folder name	avec le bouton droit de la souris sur le nouveau nom du dossier	cliquer avec le bouton droit de la souris Nouveau nom de dossier	Accuracy - Information Missing: The post-editor incorrectly removed the preposition "sur" after "souris".

F.12. Brazilian Portuguese - Participant 1

Table F.45: Brazilian Portuguese - Participant 1: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	10	Accuracy - information missing: 8 Accuracy - untranslated text: 2
Format	5	
Language	117	Language - adverbs: 1 Language - capitalisation: 16 Language - conjunctions: 3 Language - determiners: 10 Language - gender: 23 Language - number: 23 Language - prepositions: 18 Language - pronouns: 5 Language - punctuation: 5 Language - verbs: 13
Lexical choice	6	
Mistranslation	28	
Style	5	
Total	**171**	

Figure F.46: Brazilian Portuguese - Participant 1: Essential Changes

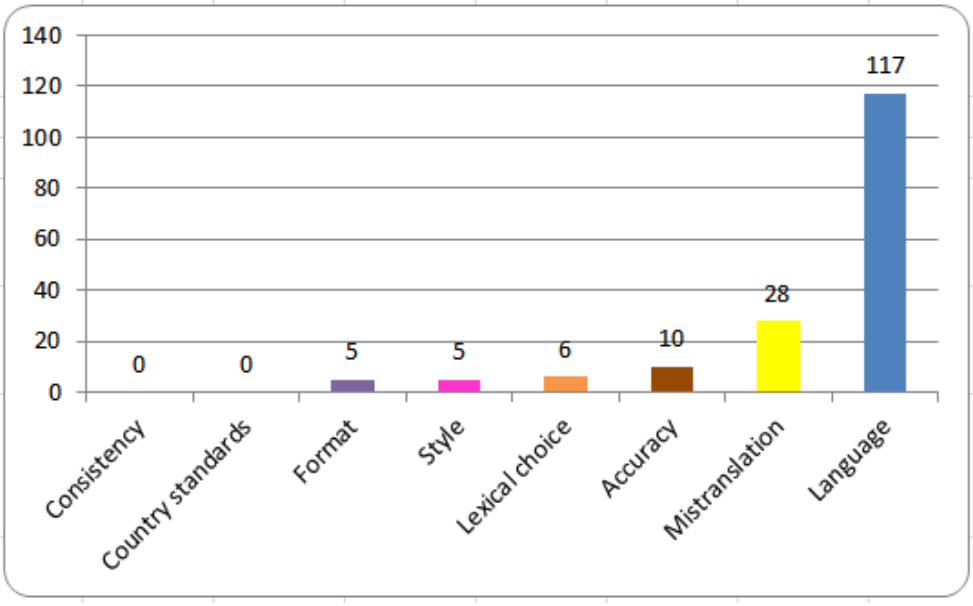


Table F.46: Brazilian Portuguese - Participant 1: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	38	Language - determiners: 6 Language - gender: 5 Language - number: 2 Language - phrasal ordering: 3 Language - prepositions: 9 Language - pronouns: 5 Language - verbs: 6
Lexical choice	18	
Style	2	
Total	**58**	

Figure F.47: Brazilian Portuguese - Participant 1: Preferential Changes

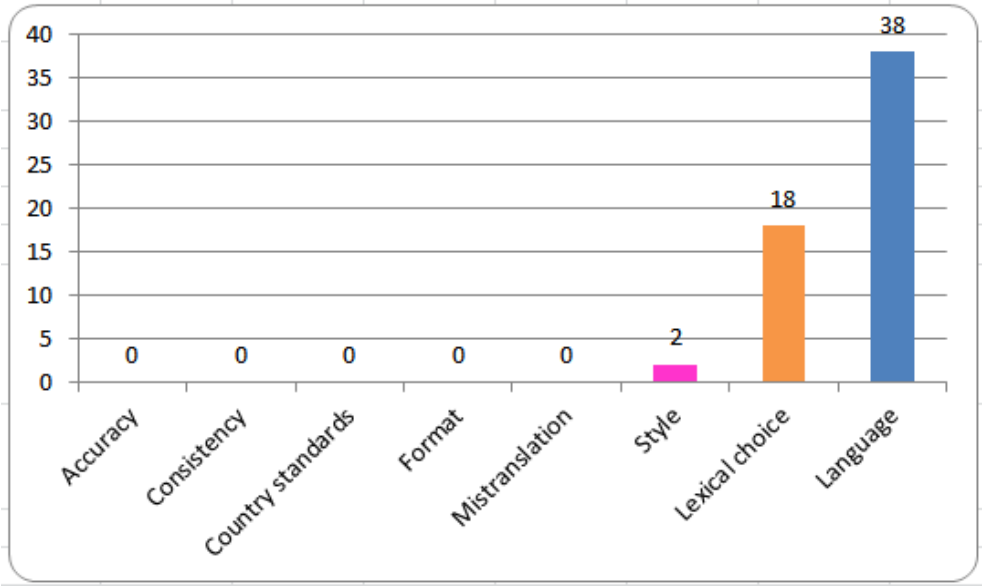


Table F.47: Brazilian Portuguese - Participant 1: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	4	Accuracy - information missing -4
Format	1	
Language	14	Language - adverbs: 1 Language - capitalisation: 6 Language - gender: 1 Language - number: 4 Language - phrasal ordering: 2
Mistranslation	1	
Style	1	
Total	**21**	

Figure F.48: Brazilian Portuguese - Participant 1: Essential Changes Not Implemented

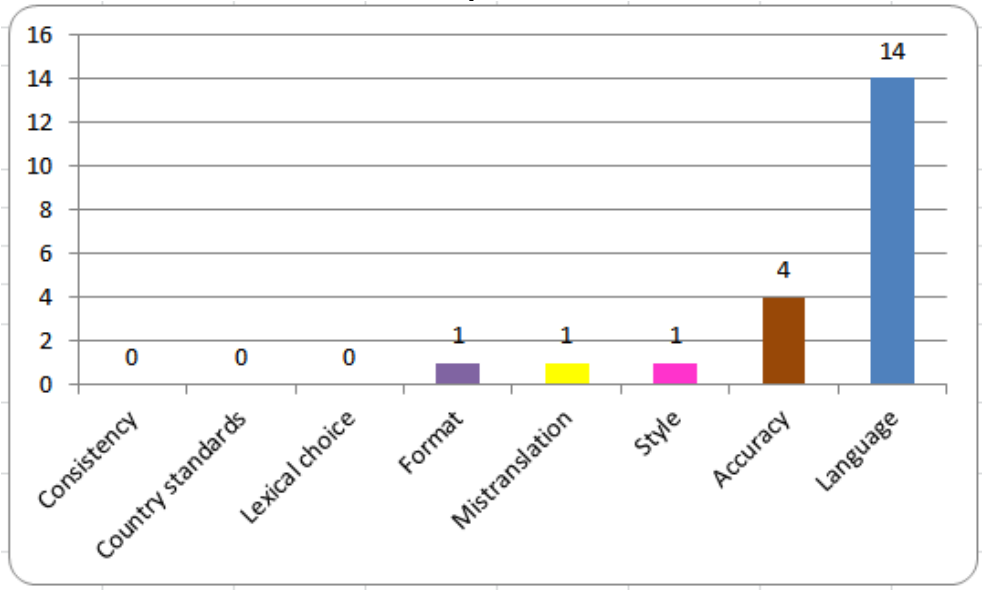


Table F.48: Brazilian Portuguese - Participant 1: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Country standards	1	
Language	4	Language - adverbs: 1 Language - determiners: 1 Language - number: 2
Mistranslation	5	
Total	**10**	

Figure F.49: Brazilian Portuguese - Participant 1: Introduced Errors

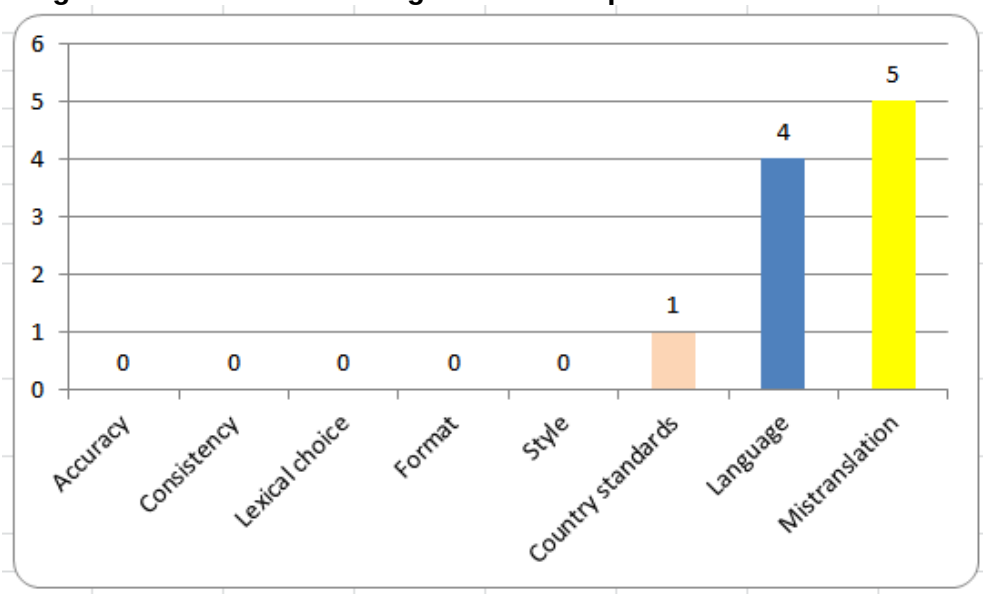
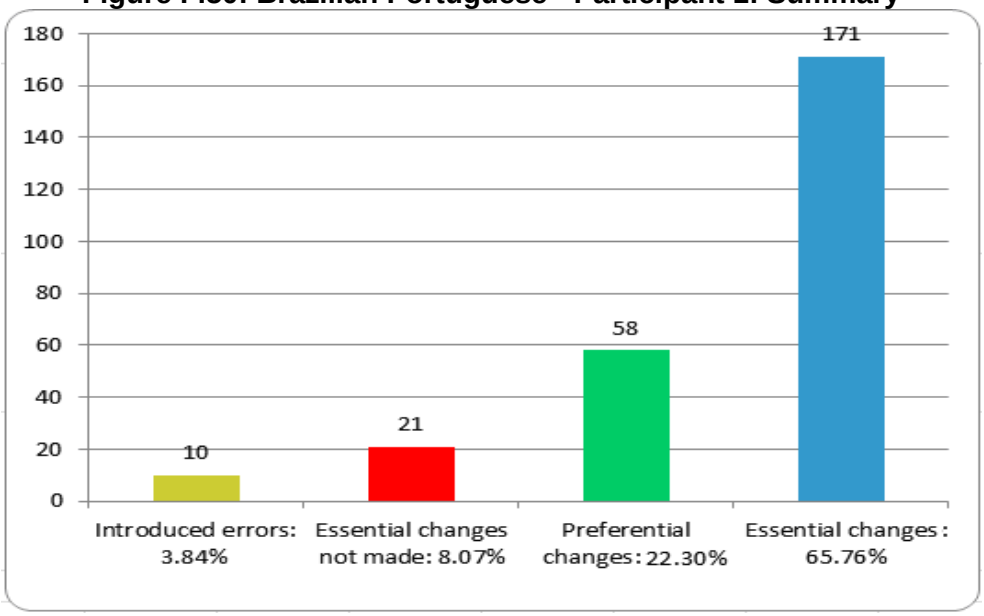


Table F.49: Brazilian Portuguese - Participant 1: Summary

ESSENTIAL CHANGES	171
Preferential Changes	58
Essential Changes Not Implemented	21
Introduced Errors	10
TOTAL	260

Figure F.50: Brazilian Portuguese - Participant 1: Summary



F.13. Brazilian Portuguese - Participant 2

Table F.50: Brazilian Portuguese - Participant 2: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	12	Accuracy - information missing: 10 Accuracy - untranslated text -2
Format	4	
Language	126	Language - adverbs: 1 Language - capitalisation: 21 Language - conjunctions: 2 Language - determiners: 6 Language - gender: 24 Language - number: 21 Language - phrasal ordering: 15 Language - prepositions: 14 Language - pronouns: 3 Language - punctuation: 5 Language - verbs: 14
Lexical choice	2	
Mistranslation	25	
Style	5	
Total	**174**	

Figure F.51: Brazilian Portuguese - Participant 2: Essential Changes

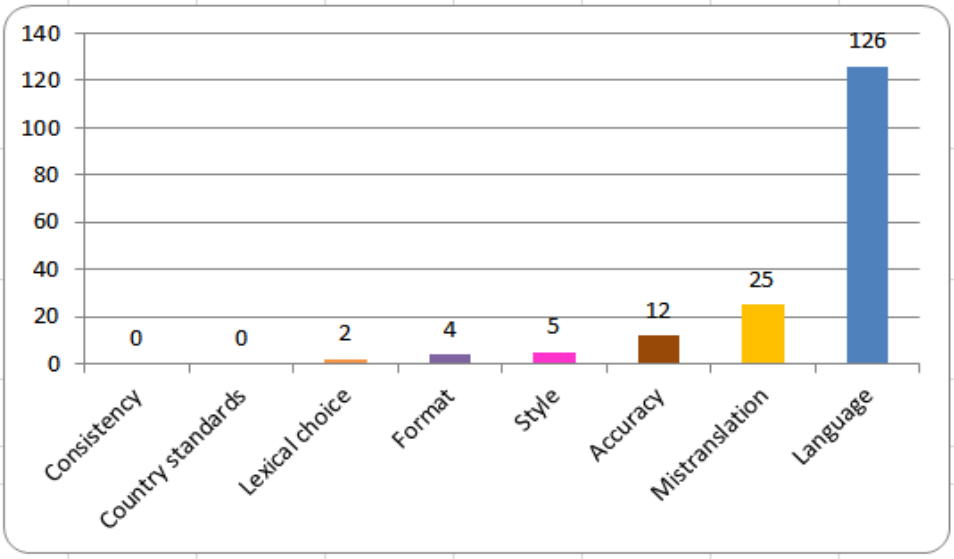


Table F.51: Brazilian Portuguese - Participant 2: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	20	Language - adverbs: 2 Language - determiners: 5 Language - number: 5 Language - prepositions: 3 Language - pronouns: 2 Language - punctuation: 1 Language - verbs: 2
Lexical choice	5	
Style	3	
Total	**28**	

Figure F.52: Brazilian Portuguese - Participant 2: Preferential Changes

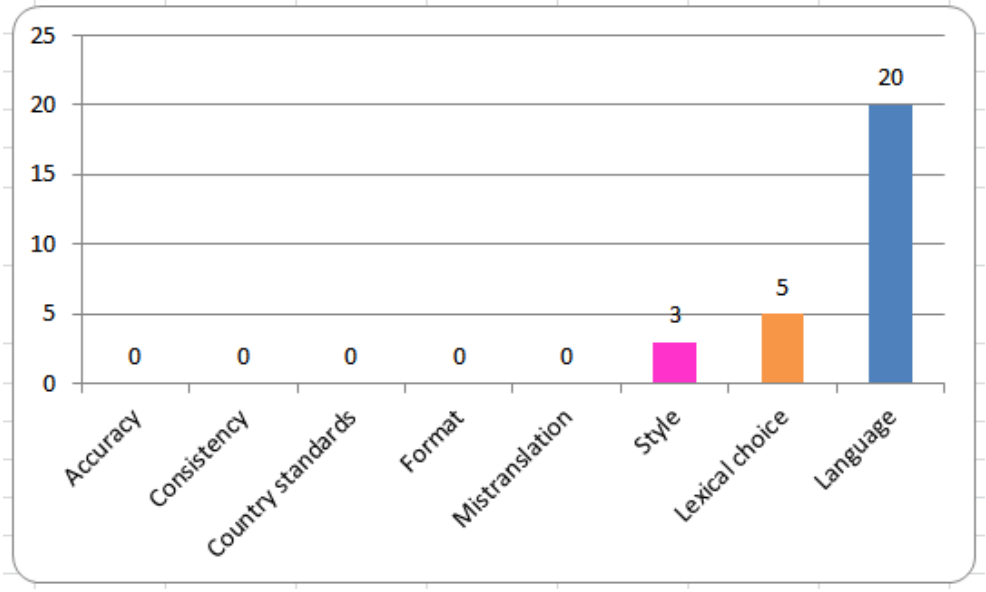


Table F.52: Brazilian Portuguese - Participant 2: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	3	Accuracy - information missing: 3
Format	2	
Language	17	Language - adverbs: 2 Language - capitalisation: 1 Language - conjunctions: 1 Language - determiners: 1 Language - gender: 1 Language - number: 2 Language - phrasal ordering: 2 Language - prepositions: 2 Language - pronouns: 3 Language - verbs: 2
Lexical choice	3	
Mistranslation	6	
Total	**31**	

Figure F.53: Brazilian Portuguese - Participant 2: Essential Changes Not Implemented

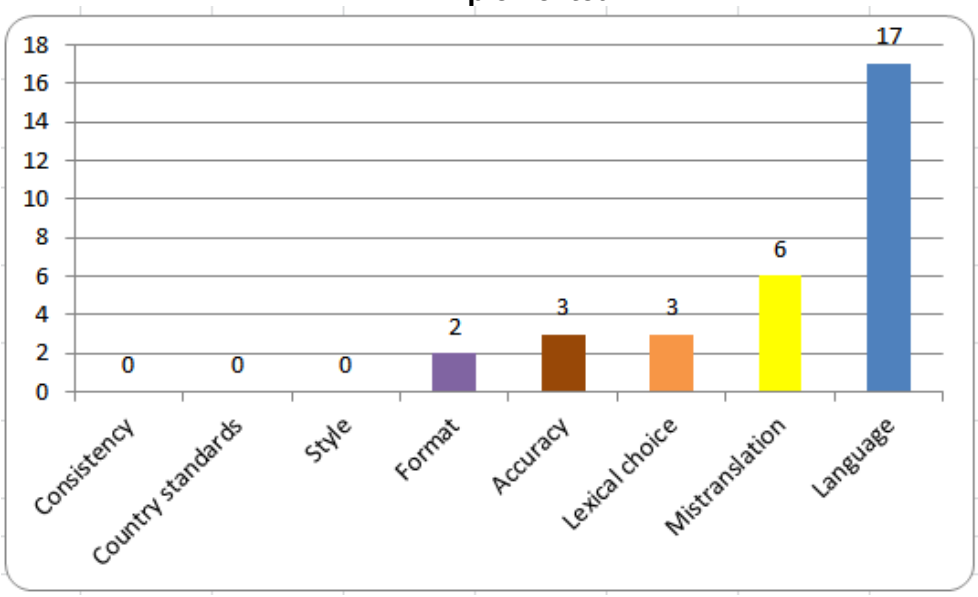


Table F.53: Brazilian Portuguese - Participant 2: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	1	Accuracy - information missing: 1
Language	5	Language - adverbs: 1 Language - determiners: 2 Language - number: 2
Lexical choice	2	
Mistranslation	2	
Total	**10**	

Figure F.54: Brazilian Portuguese - Participant 2: Introduced Errors

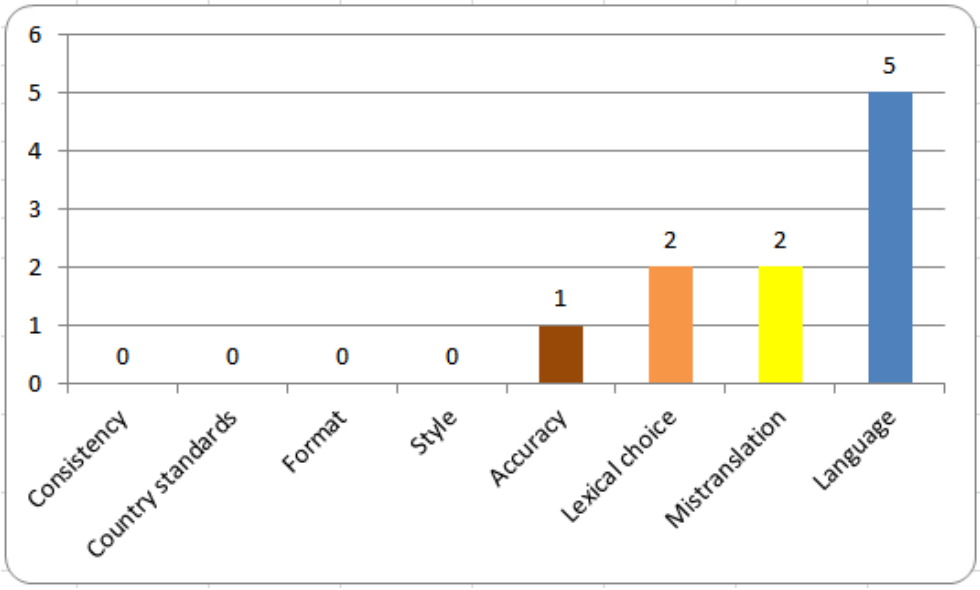
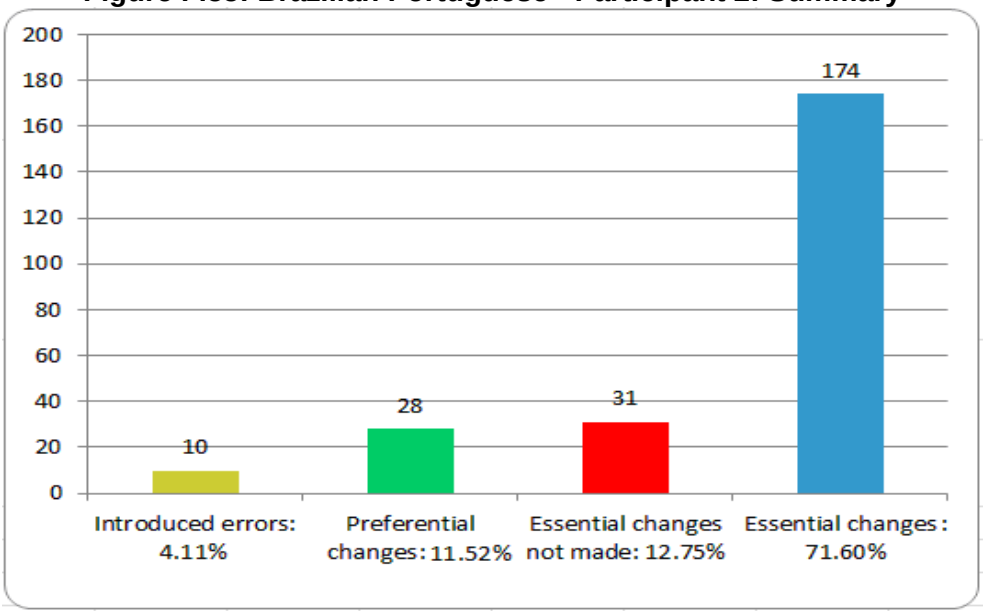


Table F.54: Brazilian Portuguese - Participant 2: Summary

ESSENTIAL CHANGES	174
Preferential Changes	28
Essential Changes Not Implemented	31
Introduced Errors	10
TOTAL	243

Figure F.55: Brazilian Portuguese - Participant 2: Summary



F.14. Brazilian Portuguese - Participant 3

Table F.55: Brazilian Portuguese - Participant 3: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	14	Accuracy - information missing: 13 Accuracy - untranslated text: 1
Format	6	
Language	141	Language - adverbs: 1 Language - capitalisation: 21 Language - conjunctions: 5 Language - determiners: 9 Language - gender: 24 Language - number: 23 Language - phrasal ordering: 16 Language - prepositions: 16 Language - pronouns: 6 Language - punctuation: 5 Language - verbs: 15
Lexical choice	7	
Mistranslation	28	
Style	4	
Total	**200**	

Figure F.56: Brazilian Portuguese - Participant 3: Essential Changes

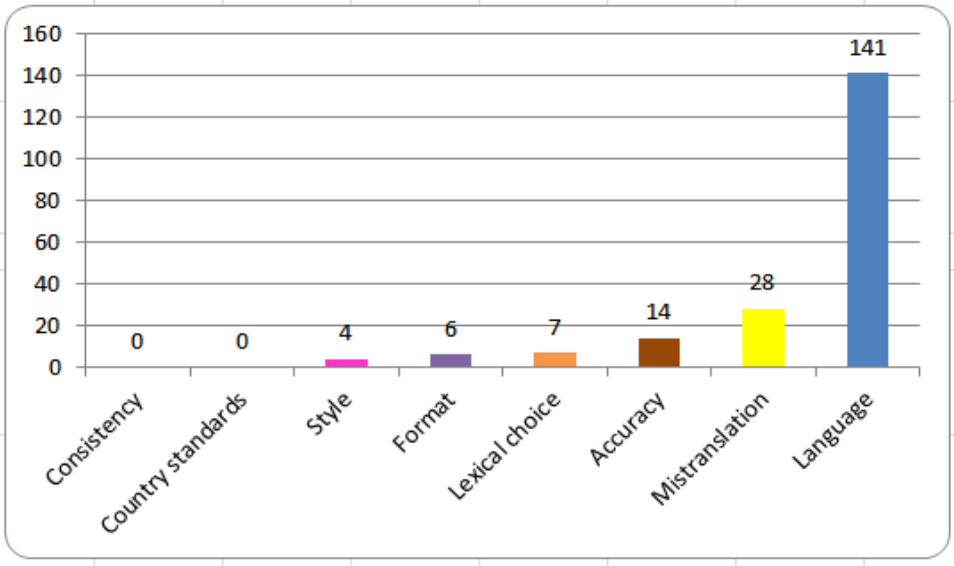


Table F.56: Brazilian Portuguese - Participant 3: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	39	Language - determiners: 11 Language - gender: 2 Language - number: 4 Language - phrasal ordering: 3 Language - prepositions: 7 Language - pronouns: 5 Language - verbs: 7
Lexical choice	19	
Style	6	
Total	**64**	

Figure F.57: Brazilian Portuguese - Participant 3: Preferential Changes

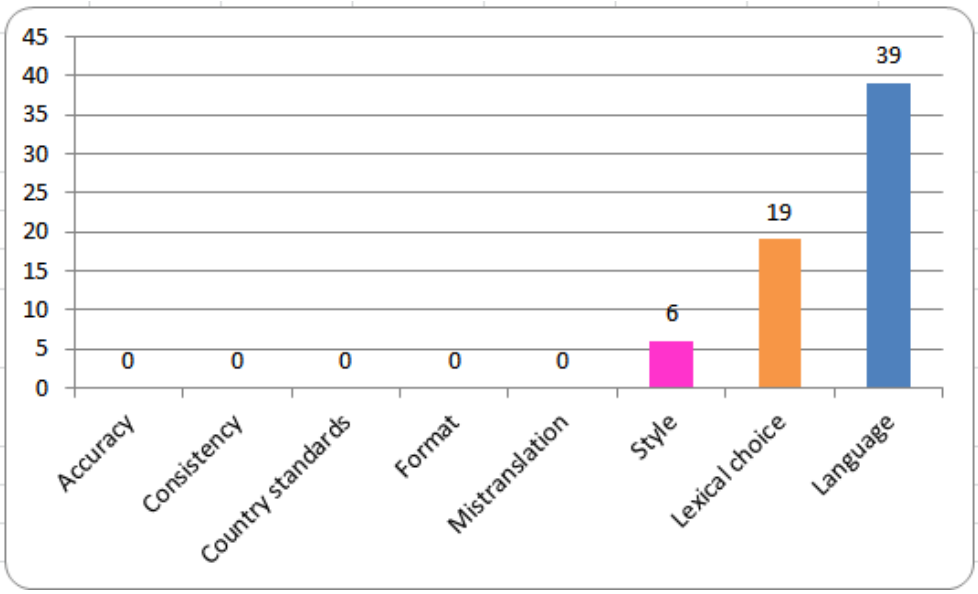


Table F.57: Brazilian Portuguese - Participant 3: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	1	Accuracy - untranslated text: 1
Language	12	Language - adverbs: 1 Language - capitalisation: 3 Language - gender: 3 Language - number: 1 Language - phrasal ordering: 3 Language - prepositions: 1
Mistranslation	1	
Total	**14**	

Figure F.58: Brazilian Portuguese - Participant 3: Essential Changes Not Implemented

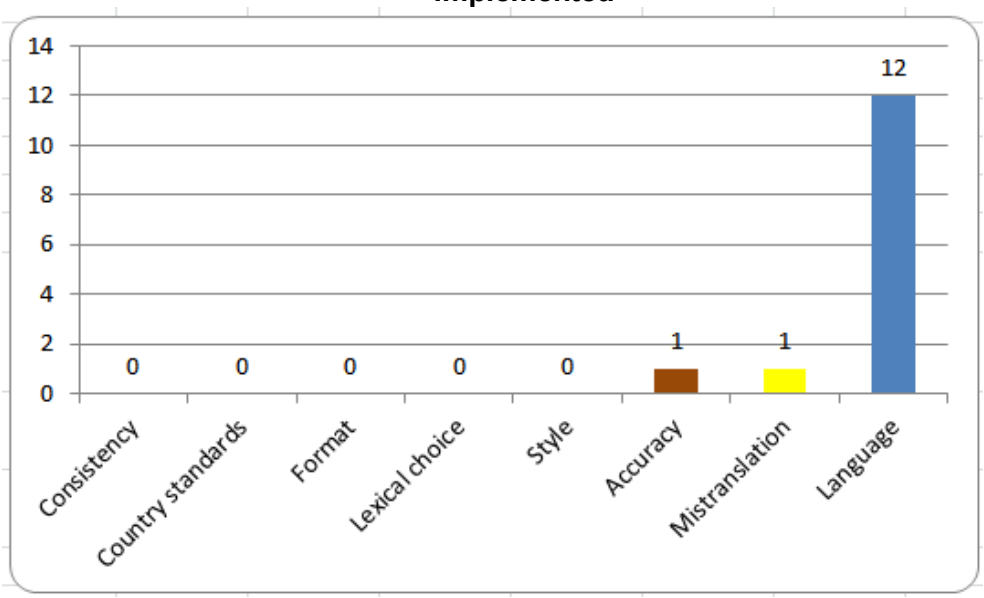


Table F.58: Brazilian Portuguese - Participant 3: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	4	Accuracy - information missing: 4
Language	17	Language - adverbs: 1 Language - capitalisation: 13 Language - number: 1 Language - prepositions: 2
Mistranslation	1	
Total	**22**	

Figure F.59: Brazilian Portuguese - Participant 3: Introduced Errors

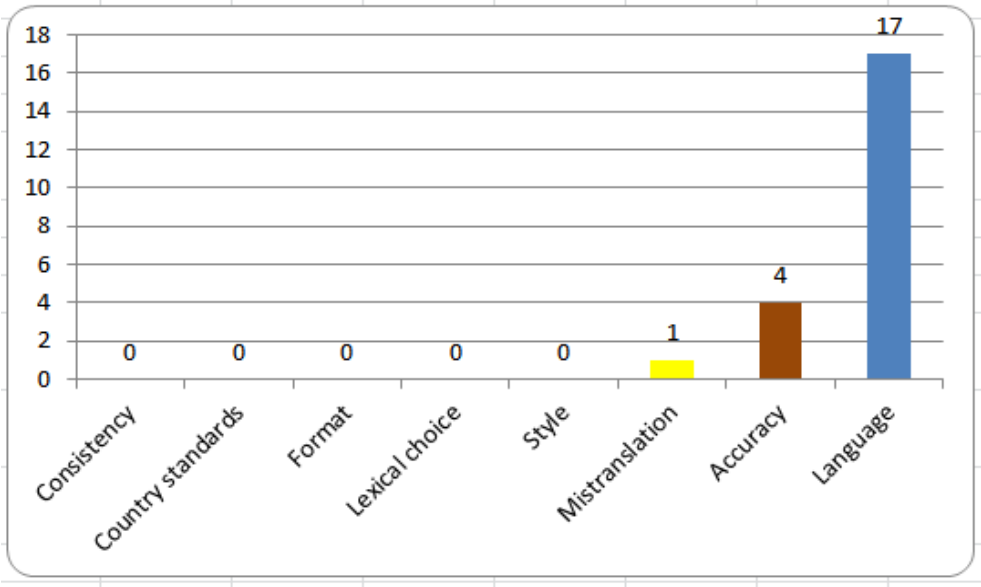
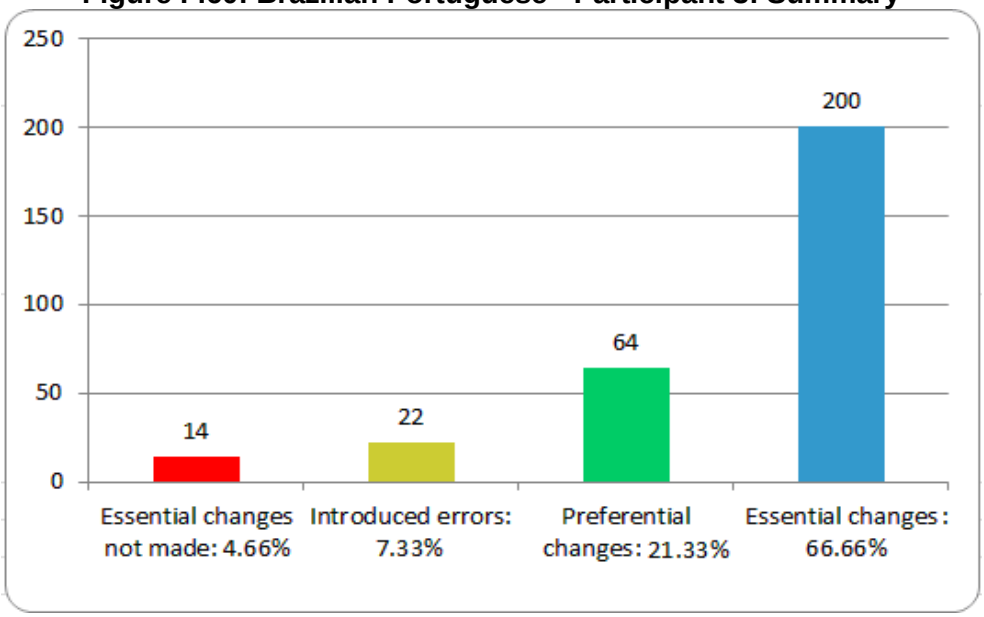


Table F.59: Brazilian Portuguese - Participant 3: Summary

ESSENTIAL CHANGES	**200**
Preferential Changes	**64**
Essential Changes Not Implemented	**14**
Introduced Errors	**22**
TOTAL	**300**

Figure F.60: Brazilian Portuguese - Participant 3: Summary



F.15. Brazilian Portuguese - Participant 4

Table F.60: Brazilian Portuguese - Participant 4: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	8	Accuracy - extra information: 1 Accuracy - information missing: 5 Accuracy - untranslated text: 2
Format	4	
Language	109	Language - capitalisation: 19 Language - conjunctions: 2 Language - determiners: 6 Language - gender: 18 Language - number: 18 Language - phrasal ordering: 14 Language - prepositions: 14 Language - pronouns: 2 Language - punctuation: 3 Language - verbs: 13
Lexical choice	3	
Mistranslation	21	
Style	4	
Total	**149**	

Figure F.61: Brazilian Portuguese - Participant 4: Essential Changes

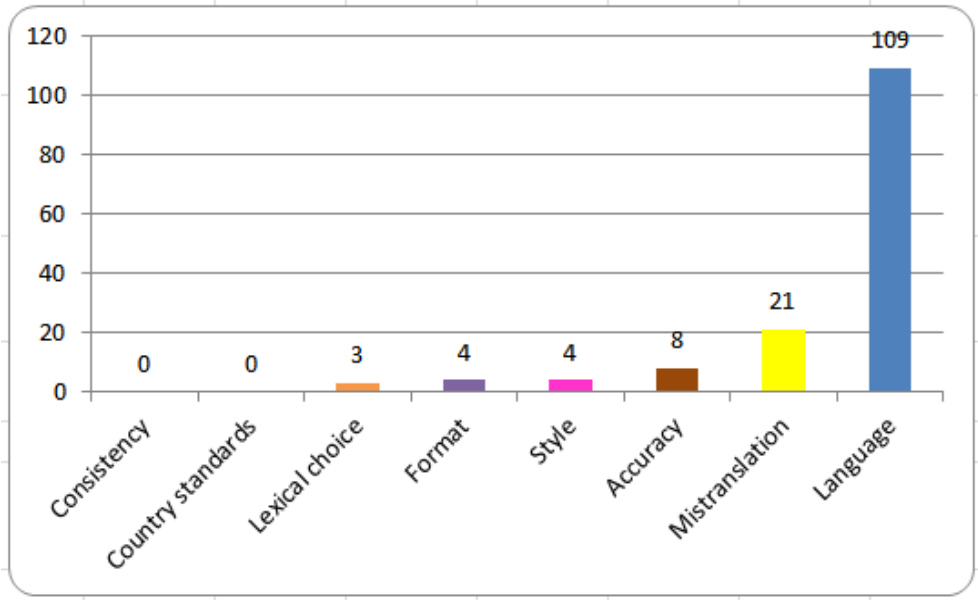


Table F.61: Brazilian Portuguese - Participant 4: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	35	Language - adverbs: 5 Language - capitalisation: 1 Language - conjunctions: 3 Language - determiners: 6 Language - number: 4 Language - phrasal ordering: 1 Language - prepositions: 5 Language - pronouns: 3 Language - punctuation: 2 Language - verbs: 5
Lexical choice	18	
Style	5	
Total	**58**	

Figure F.62: Brazilian Portuguese - Participant 4: Preferential Changes

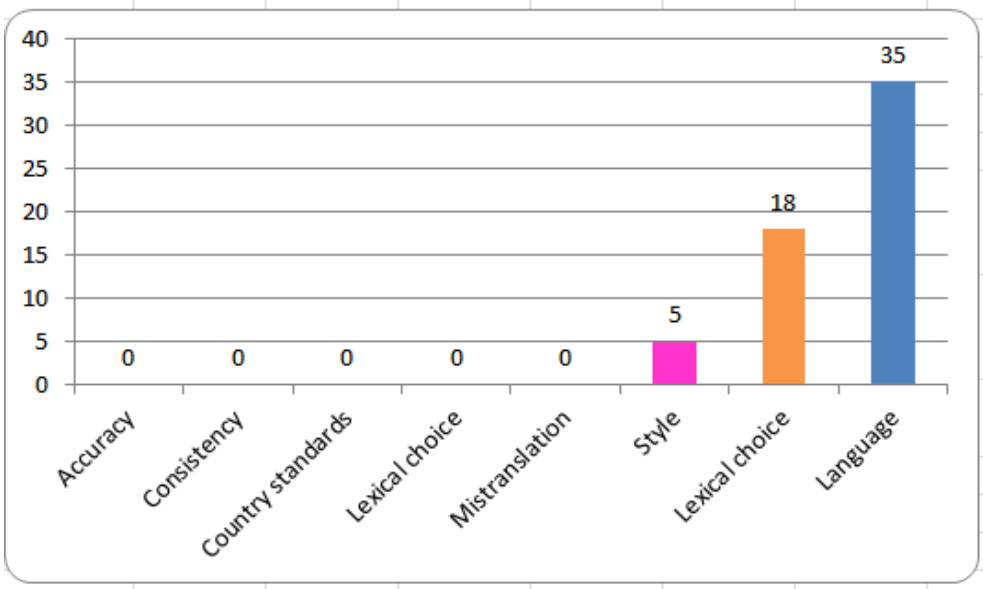


Table F.62: Brazilian Portuguese - Participant 4: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	8	Accuracy - information missing: 8
Format	2	
Language	31	Language - adverbs: 2 Language - capitalisation: 3 Language - determiners: 3 Language - gender: 6 Language - number: 6 Language - phrasal ordering: 4 Language - prepositions: 2 Language - pronouns: 3 Language - punctuation: 1 Language - verbs: 1
Lexical choice	2	
Mistranslation	10	
Total	**53**	

Figure F.63: Brazilian Portuguese - Participant 4: Essential Changes Not Implemented

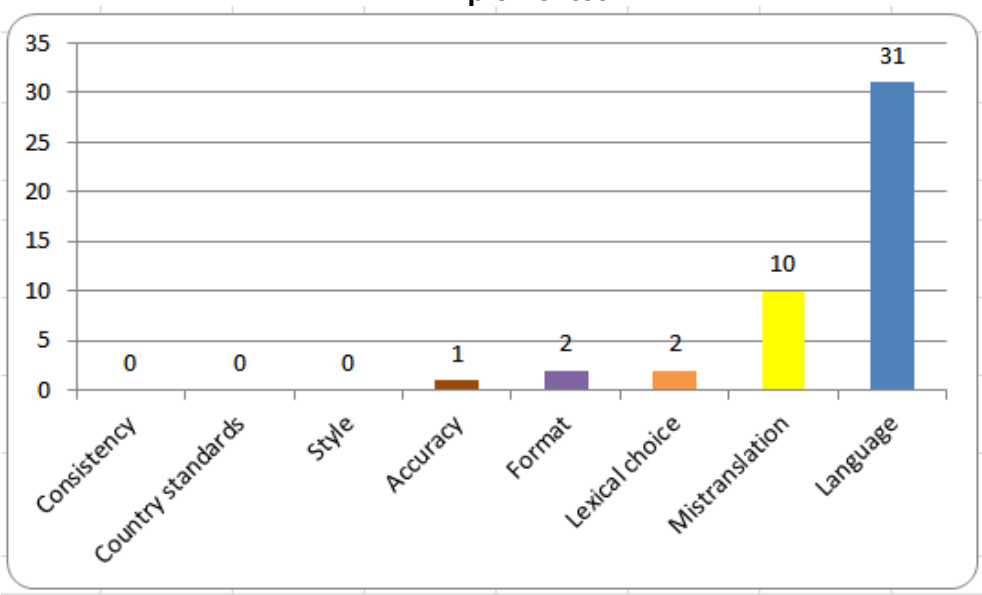


Table F.63: Brazilian Portuguese - Participant 4: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	3	Accuracy - information missing: 3
Format	1	
Language	20	Language - adverbs: 1 Language - capitalisation: 11 Language - determiners: 1 Language - gender: 1 Language - number: 1 Language - phrasal ordering: 1 Language - prepositions: 1 Language - punctuation: 1 Language - verbs: 2
Lexical choice	2	
Mistranslation	1	
Style	1	
Total	**28**	

Figure F.64: Brazilian Portuguese - Participant 4: Introduced Errors

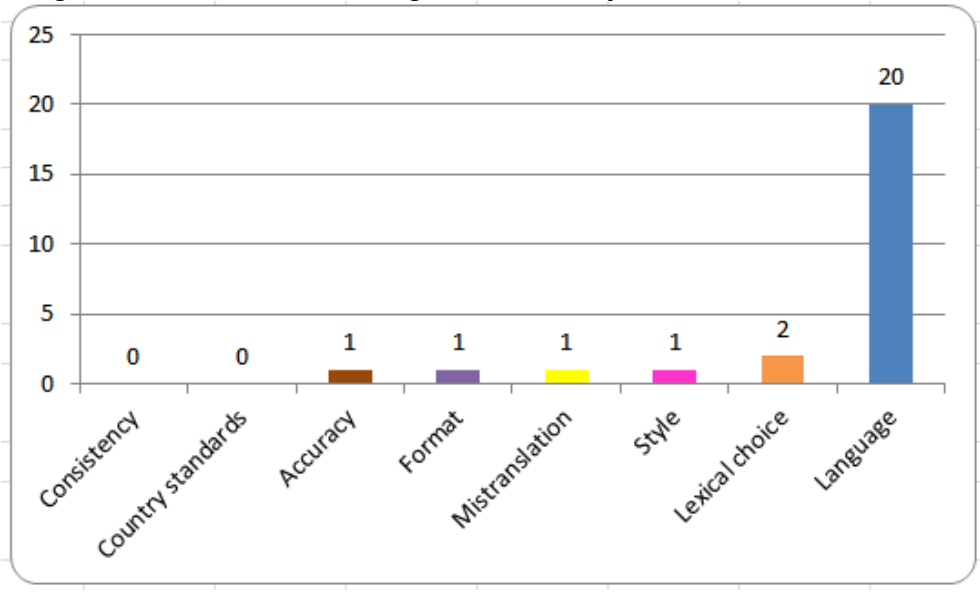
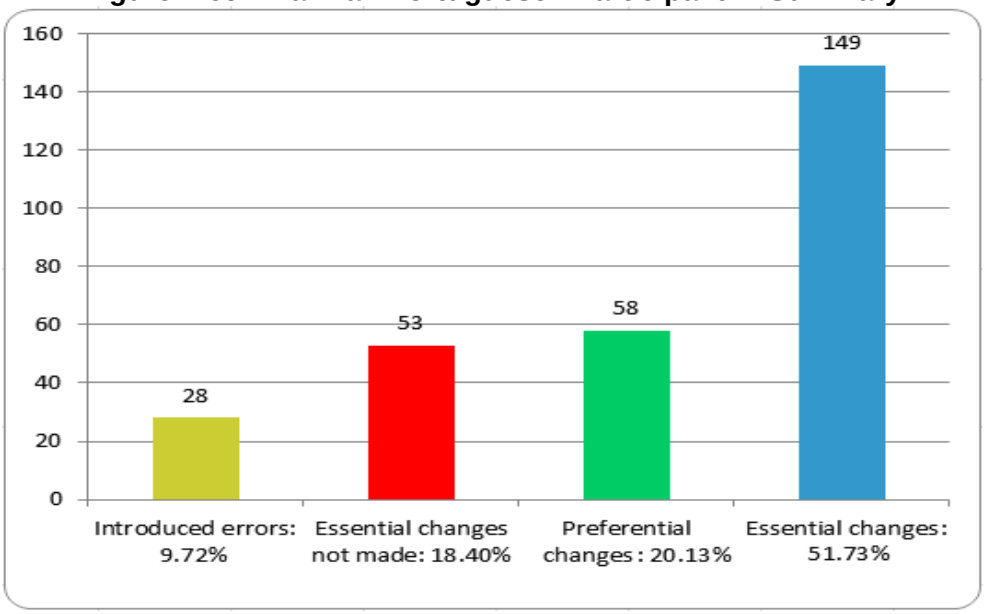


Table F.64: Brazilian Portuguese - Participant 4: Summary

ESSENTIAL CHANGES	149
Preferential Changes	58
Essential Changes Not Implemented	53
Introduced Errors	28
TOTAL	288

Figure F.65: Brazilian Portuguese - Participant 4: Summary



F.16. Brazilian Portuguese - Participant 5

Table F.65: Brazilian Portuguese - Participant 5: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	6	Accuracy - information missing: 4 Accuracy - untranslated text: 2
Format	6	
Language	124	Language - adverbs: 1 Language - capitalisation: 14 Language - conjunctions: 4 Language - determiners: 8 Language - gender: 21 Language - number: 22 Language - phrasal ordering: 15 Language - prepositions: 16 Language - pronouns: 4 Language - punctuation: 3 Language - verbs: 16
Lexical choice	5	
Mistranslation	25	
Style	2	
Total	**168**	

Figure F.66: Brazilian Portuguese - Participant 5: Essential Changes

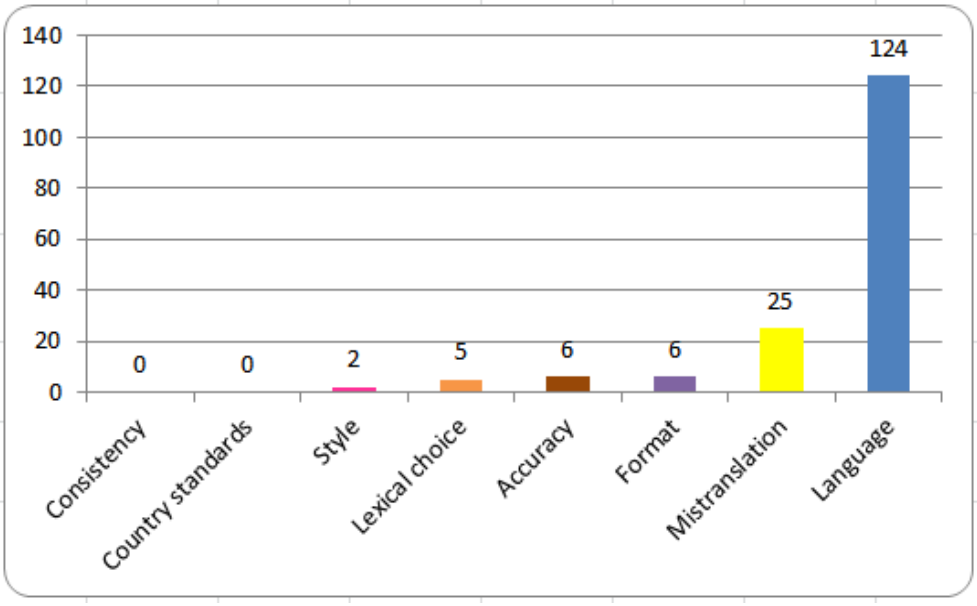


Table F.66: Brazilian Portuguese - Participant 5: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	30	Language - conjunctions: 1 Language - determiners: 6 Language - gender: 2 Language - number: 1 Language - phrasal ordering: 6 Language - prepositions: 2 Language - pronouns: 5 Language - punctuation: 1 Language - verbs: 6
Lexical choice	15	
Style	5	
Total	**50**	

Figure F.67: Brazilian Portuguese - Participant 5: Preferential Changes

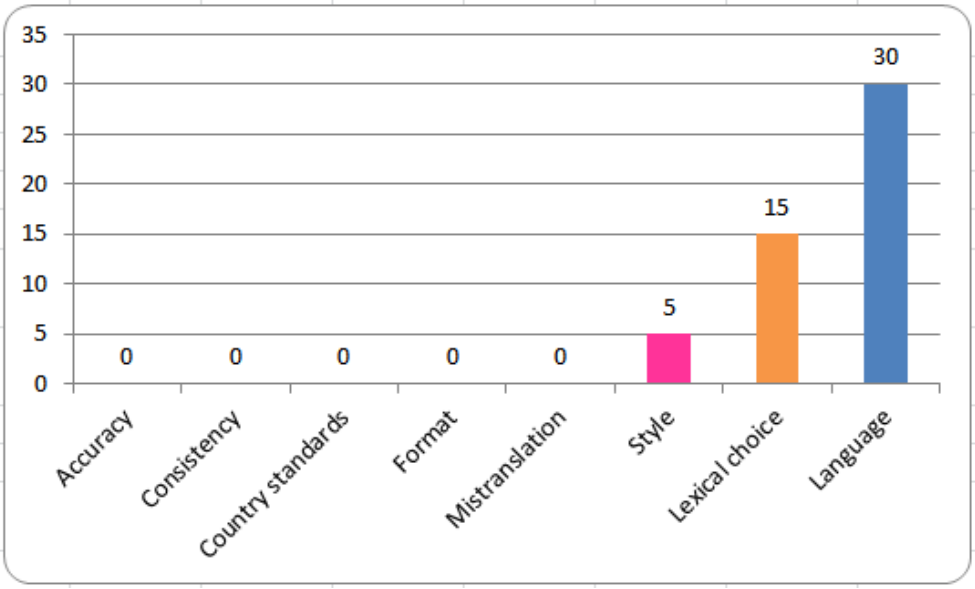


Table F.67: Brazilian Portuguese - Participant 5: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	7	Accuracy - information missing: 7
Language	28	Language - adverbs: 1 Language - capitalisation: 8 Language - determiners: 2 Language - gender: 3 Language - number: 3 Language - phrasal ordering: 2 Language - pronouns: 4 Language - punctuation: 2 Language - verbs: 1
Lexical choice	1	
Mistranslation	5	
Total	**41**	

Figure F.68: Brazilian Portuguese - Participant 5: Essential Changes Not Implemented

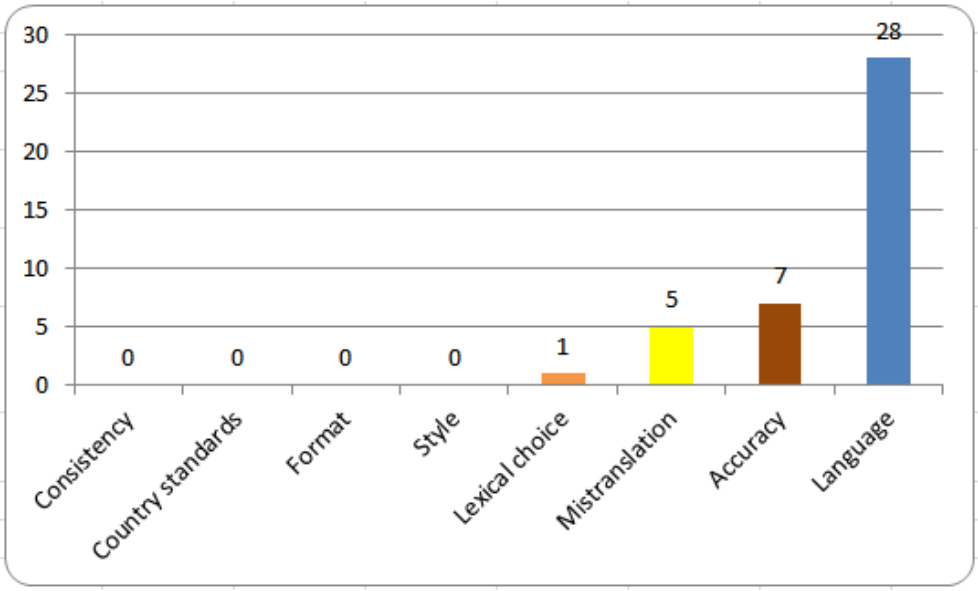


Table F.68: Brazilian Portuguese - Participant 5: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Country standards	1	
Format	1	
Language	7	Language - determiners: 2 Language - number: 3 Language - prepositions: 1 Language - verbs: 1
Mistranslation	1	
Total	**10**	

Figure F.69: Brazilian Portuguese - Participant 5: Introduced Errors

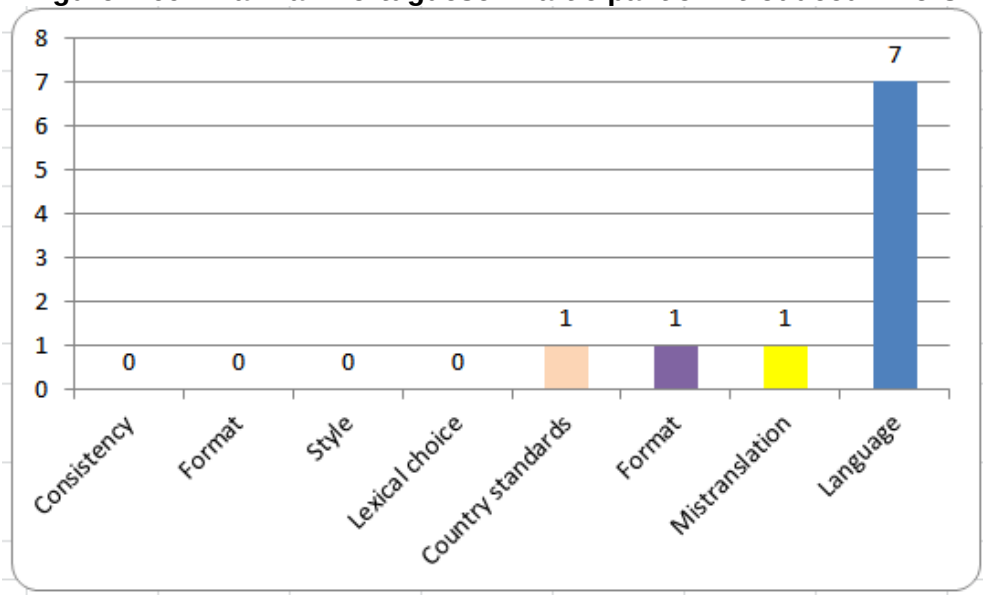
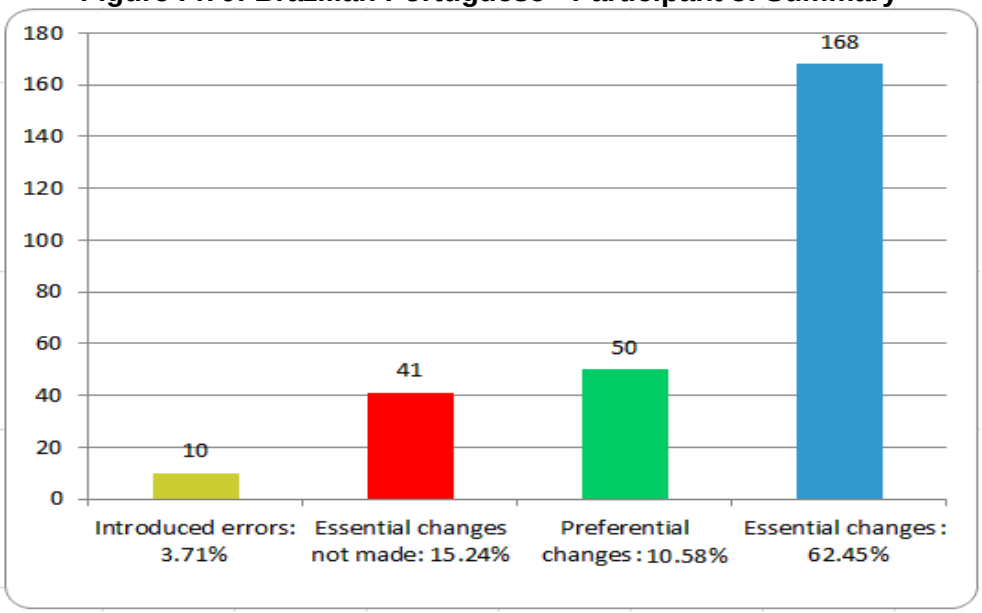


Table F.69: Brazilian Portuguese - Participant 5: Summary

ESSENTIAL CHANGES	168
Preferential Changes	50
Essential Changes Not Implemented	41
Introduced Errors	10
TOTAL	269

Figure F.70: Brazilian Portuguese - Participant 5: Summary



F.17. Brazilian Portuguese - Participant 6

Table F.70: Brazilian Portuguese - Participant 6: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	11	Accuracy - information missing: 10 Accuracy - untranslated text: 1
Format	4	
Language	105	Language - capitalisation: 15 Language - conjunctions: 4 Language - determiners: 5 Language - gender: 22 Language - number: 19 Language - phrasal ordering: 13 Language - prepositions: 10 Language - pronouns: 2 Language - punctuation: 3 Language - verbs: 12
Lexical choice	6	
Mistranslation	24	
Style	2	
Total	**152**	

Figure F.71: Brazilian Portuguese - Participant 6: Essential Changes

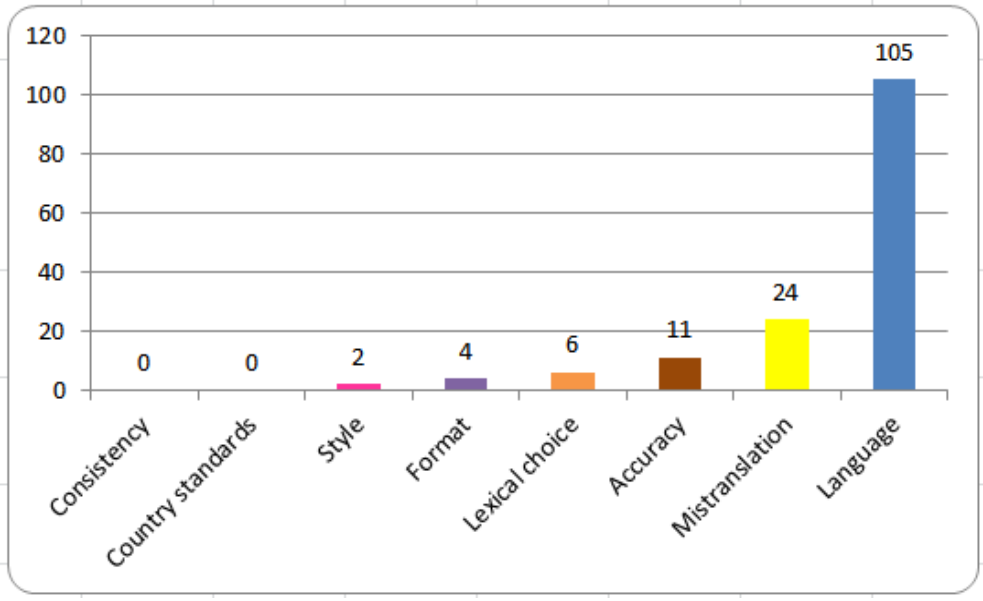


Table F.71: Brazilian Portuguese - Participant 6: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	6	Language - number: 2 Language - phrasal ordering: 1 Language - prepositions: 1 Language - punctuation: 1 Language - verbs: 1
Lexical choice	2	
Style	1	
Total	**9**	

Figure F.72: Brazilian Portuguese - Participant 6: Preferential Changes

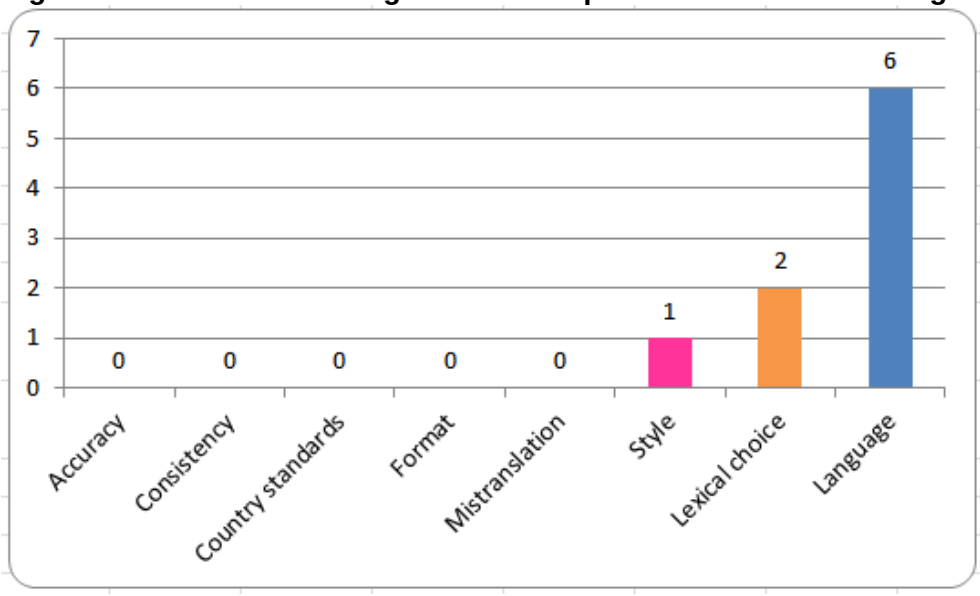


Table F.72: Brazilian Portuguese - Participant 6: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	4	Accuracy - information missing: 3 Accuracy - untranslated text: 1
Format	2	
Language	47	Language - adverbs: 2 Language - capitalisation: 7 Language - determiners: 6 Language - gender: 2 Language - number: 9 Language - phrasal ordering -3 Language - prepositions: 7 Language - pronouns: 6 Language - punctuation: 2 Language - verbs: 3
Lexical choice	1	
Mistranslation	6	
Total	**60**	

Figure F.73: Brazilian Portuguese - Participant 6: Essential Changes Not Implemented

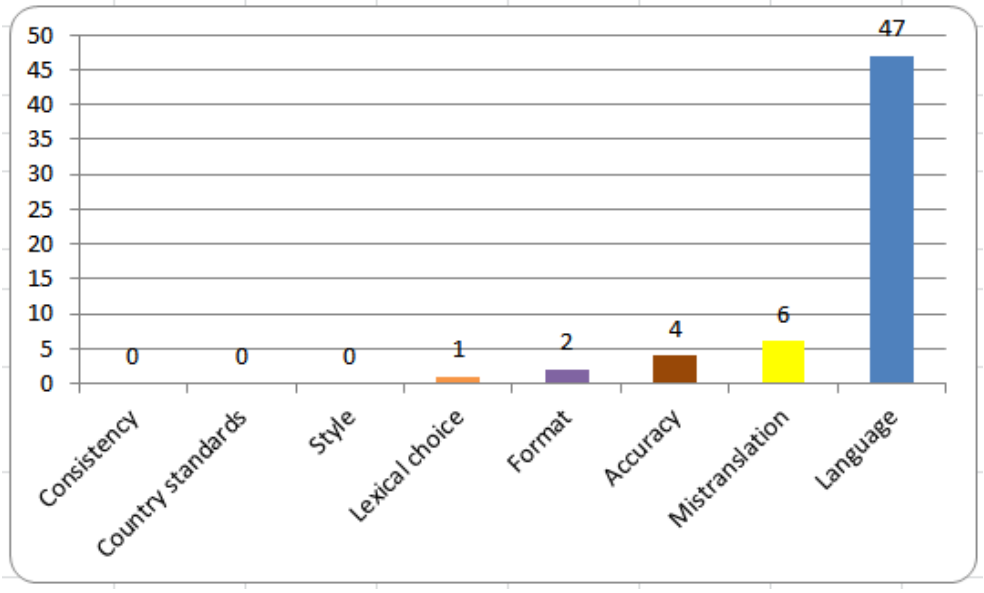


Table F.73: Brazilian Portuguese - Participant 6: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	1	Accuracy - information missing: 1
Language	3	Language - determiners: 1 Language - number: 2
Lexical choice	1	
Mistranslation	2	
Total	**7**	

Figure F.74: Brazilian Portuguese - Participant 6: Introduced Errors

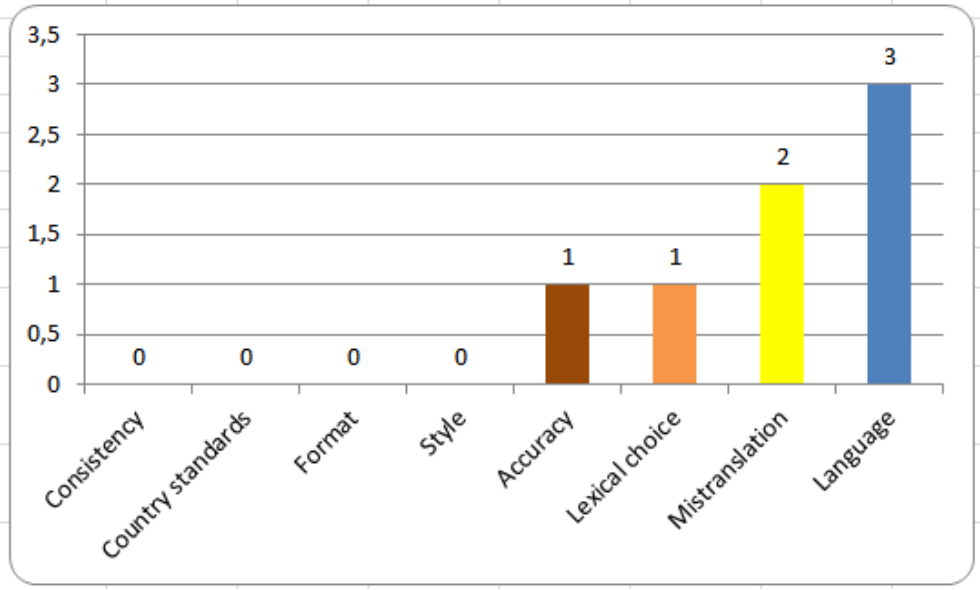
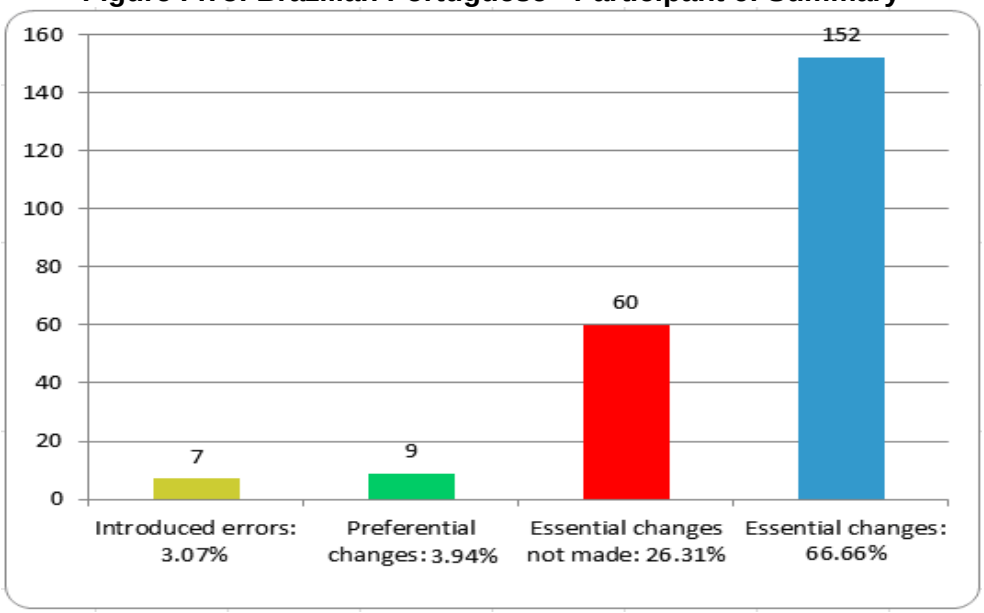


Table F.74: Brazilian Portuguese - Participant 6: Summary

ESSENTIAL CHANGES	152
Preferential Changes	9
Essential Changes Not Implemented	60
Introduced Errors	7
TOTAL	228

Figure F.75: Brazilian Portuguese - Participant 6: Summary



F.18. Brazilian Portuguese - Participant 7

Table F.75: Brazilian Portuguese - Participant 7: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	8	Accuracy - information missing: 6 Accuracy - untranslated text: 2
Format	4	
Language	106	Language - capitalisation: 16 Language - conjunctions: 3 Language - determiners: 7 Language - gender: 16 Language - number: 19 Language - phrasal ordering: 14 Language - prepositions: 10 Language - pronouns: 4 Language - punctuation: 4 Language - verbs: 12
Lexical choice	7	
Mistranslation	16	
Total	**141**	

Figure F.76: Brazilian Portuguese - Participant 7: Essential Changes

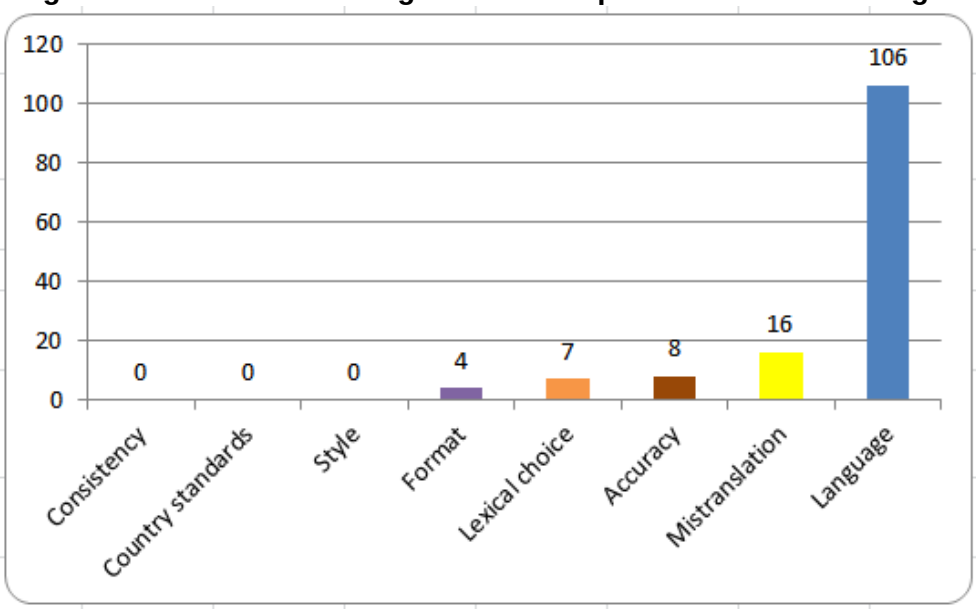


Table F.76: Brazilian Portuguese - Participant 7: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Language	24	Language - adverbs: 3 Language - capitalisation: 1 Language - conjunctions: 1 Language - determiners: 4 Language - number: 2 Language - phrasal ordering: 2 Language - prepositions: 3 Language - pronouns: 4 Language - punctuation: 1 Language - verbs: 3
Lexical choice	10	
Style	3	
Total	**37**	

Figure F.77: Brazilian Portuguese - Participant 7: Preferential Changes

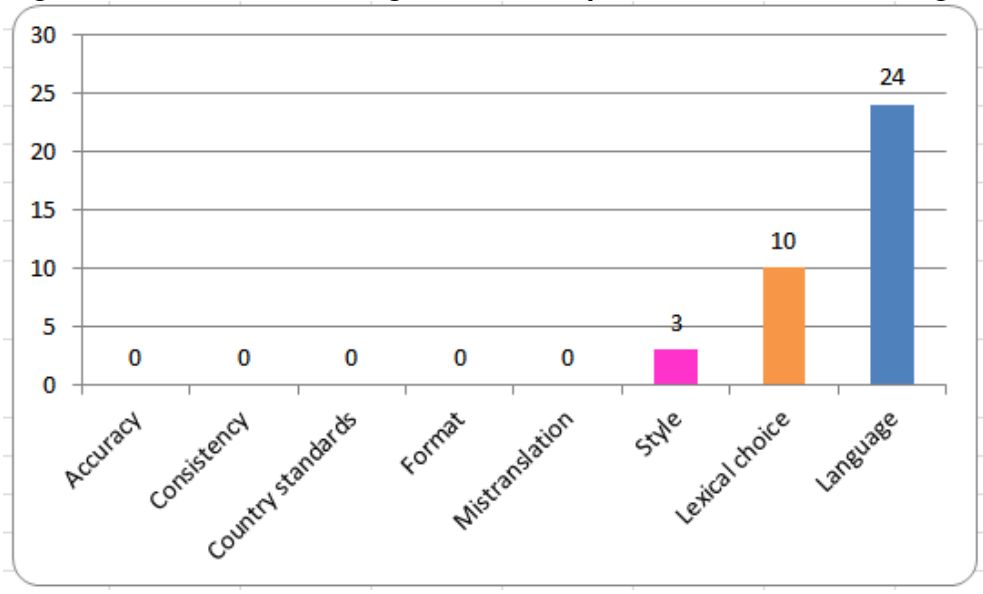


Table F.77: Brazilian Portuguese - Participant 7: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	7	Accuracy - information missing: 7
Format	2	
Language	31	Language - adverbs: 2 Language - capitalisation: 6 Language - determiners: 3 Language - gender: 6 Language - number: 3 Language - phrasal ordering: 4 Language - prepositions: 4 Language - pronouns: 1 Language - punctuation: 1 Language - verbs: 1
Lexical choice	2	
Mistranslation	12	
Total	**54**	

Figure F.78: Brazilian Portuguese - Participant 7: Essential Changes Not Implemented

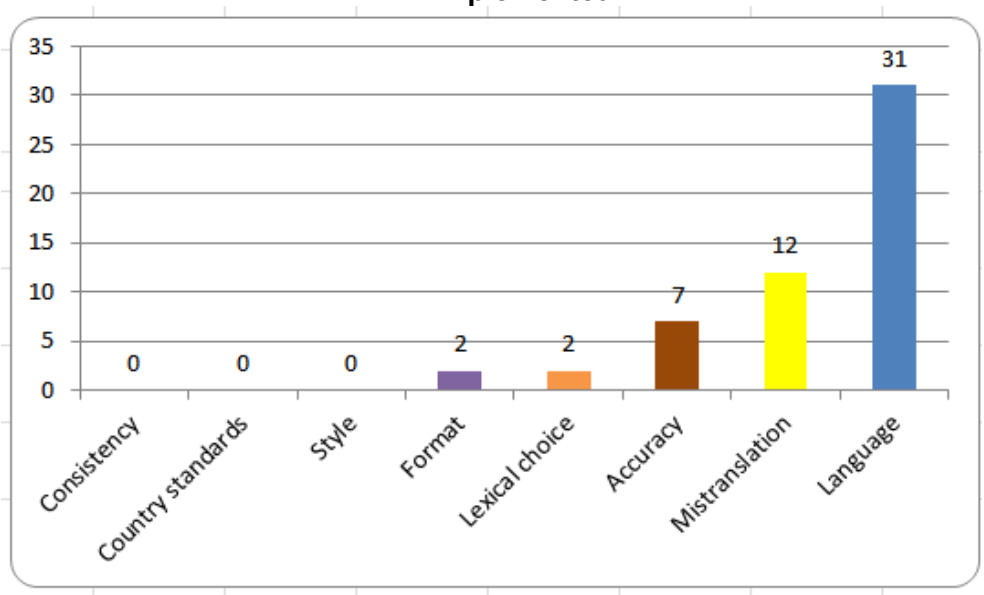


Table F.78: Brazilian Portuguese - Participant 7: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	1	Accuracy - information missing: 6
Language	9	Language - adverbs: 1 Language - capitalisation: 2 Language - determiners: 2 Language - gender: 1 Language - number: 1 Language - prepositions: 1
Mistranslation	9	
Total	**19**	

Figure F.79: Brazilian Portuguese - Participant 7: Introduced Errors

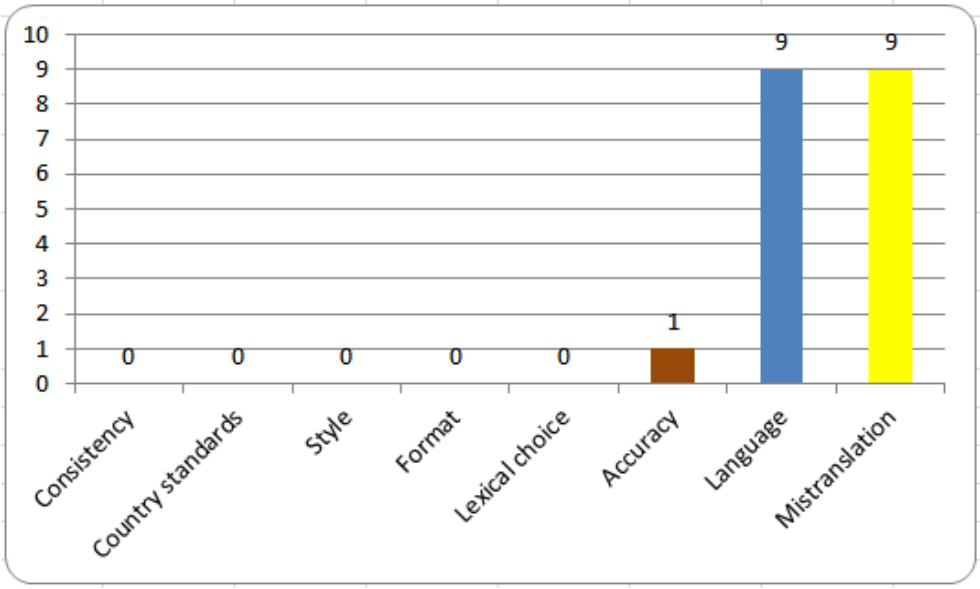
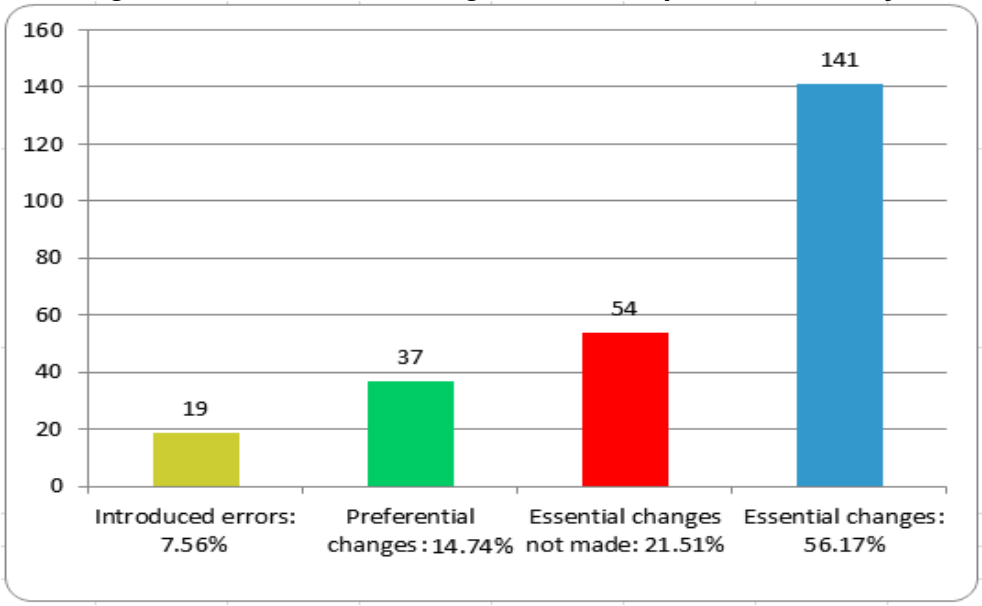


Table F.79: Brazilian Portuguese - Participant 7: Summary

ESSENTIAL CHANGES	**141**
Preferential Changes	**37**
Essential Changes Not Implemented	**54**
Introduced Errors	**19**
TOTAL	**251**

Figure F.80: Brazilian Portuguese - Participant 7: Summary



F.19. Brazilian Portuguese - Participant 8

Table F.80: Brazilian Portuguese - Participant 8: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	8	Accuracy - information missing: 6 Accuracy - untranslated text: 2
Format	4	
Language	136	Language - adverbs: 3 Language - capitalisation: 17 Language - conjunctions: 4 Language - determiners: 7 Language - gender: 23 Language - number: 23 Language - phrasal ordering: 16 Language - prepositions: 17 Language - pronouns: 5 Language - punctuation: 5 Language - verbs: 16
Lexical choice	5	
Mistranslation	23	
Style	2	
Total	**178**	

Figure F.81: Brazilian Portuguese - Participant 8: Essential Changes

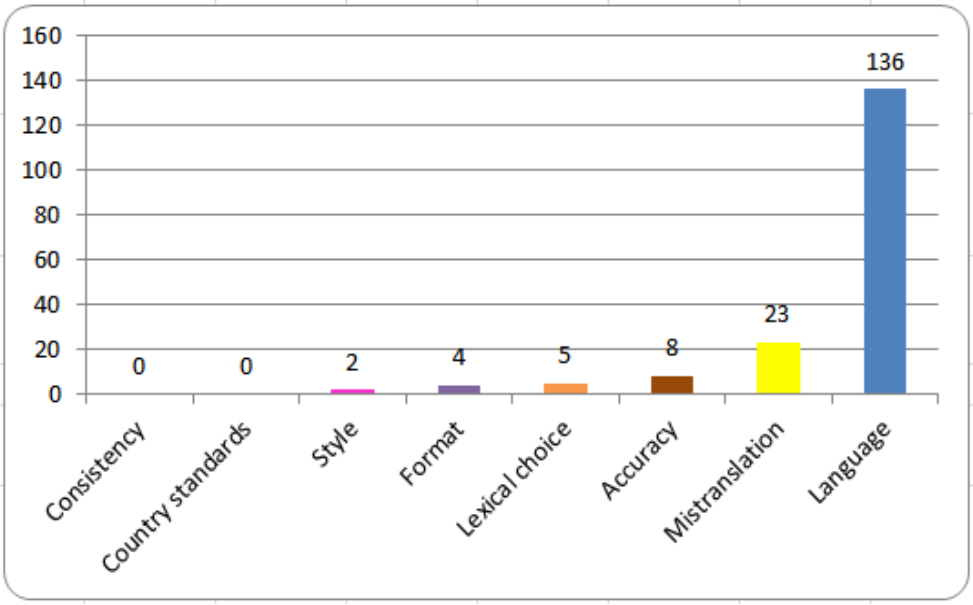


Table F.81: Brazilian Portuguese - Participant 8: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Accuracy	1	Accuracy - extra information: 1
Language	29	Language - adverbs: 1 Language - determiners: 5 Language - number: 6 Language - phrasal ordering: 2 Language - prepositions: 3 Language - pronouns: 2 Language - punctuation: 2 Language - verbs: 8
Lexical choice	4	
Style	1	
Total	**35**	

Figure F.82: Brazilian Portuguese - Participant 8: Preferential Changes

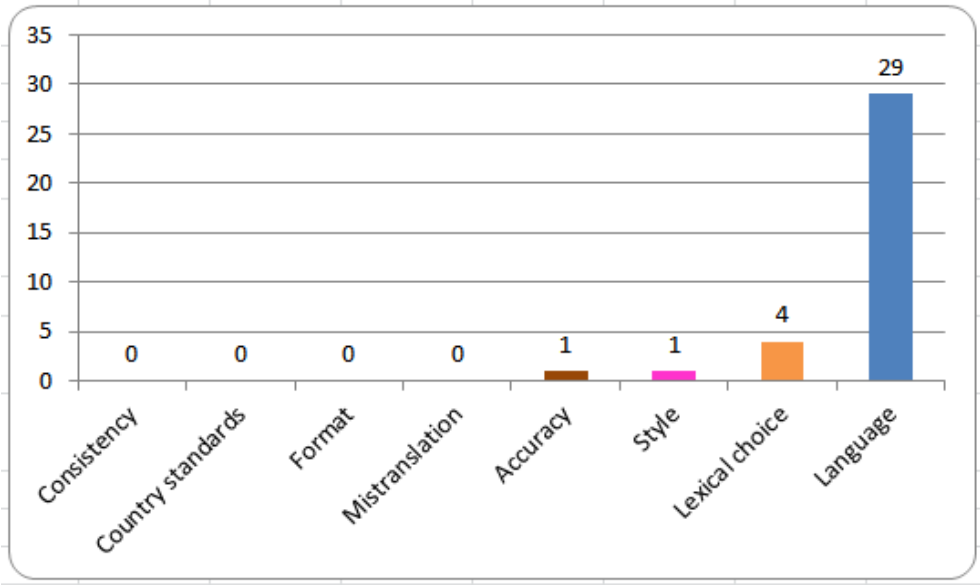


Table F.82: Brazilian Portuguese - Participant 8: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	7	Accuracy - information missing: 7
Format	2	
Language	13	Language - capitalisation: 4 Language - determiners: 2 Language - gender: 3 Language - number: 1 Language - phrasal ordering: 1 Language - pronouns: 1 Language - punctuation: 1
Lexical choice	1	
Mistranslation	6	
Total	**29**	

Figure F.83: Brazilian Portuguese - Participant 8: Essential Changes Not Implemented

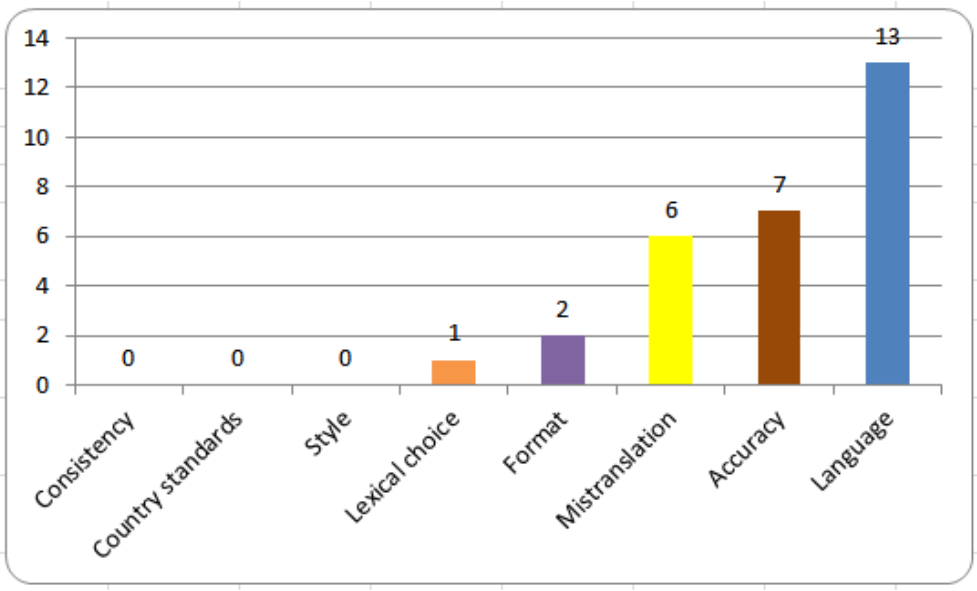


Table F.83: Brazilian Portuguese - Participant 8: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	2	Accuracy - information missing: 2
Language	3	Language - determiners: 1 Language - gender: 2
Mistranslation	3	
Total	**8**	

Figure F.84: Brazilian Portuguese - Participant 8: Introduced Errors

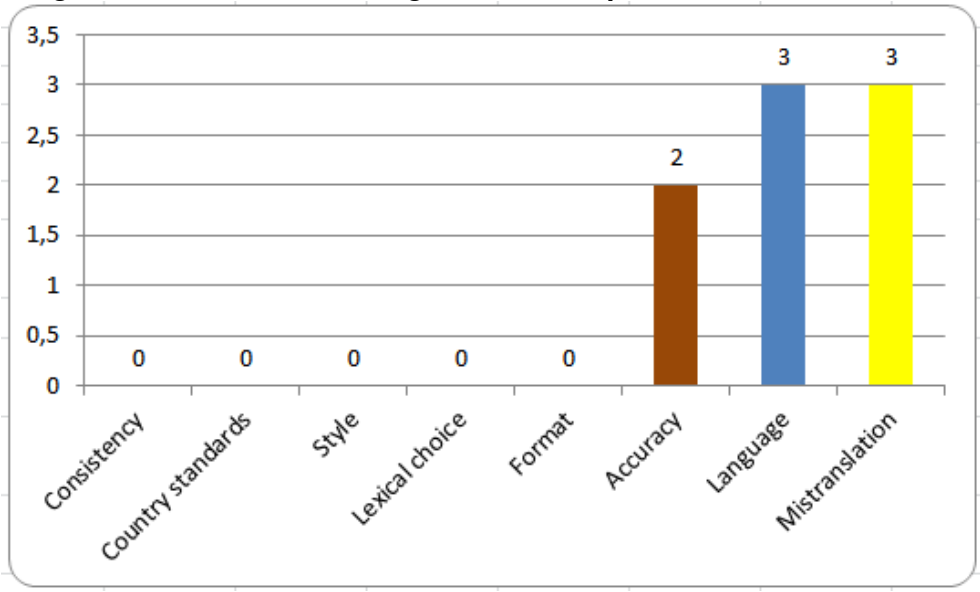
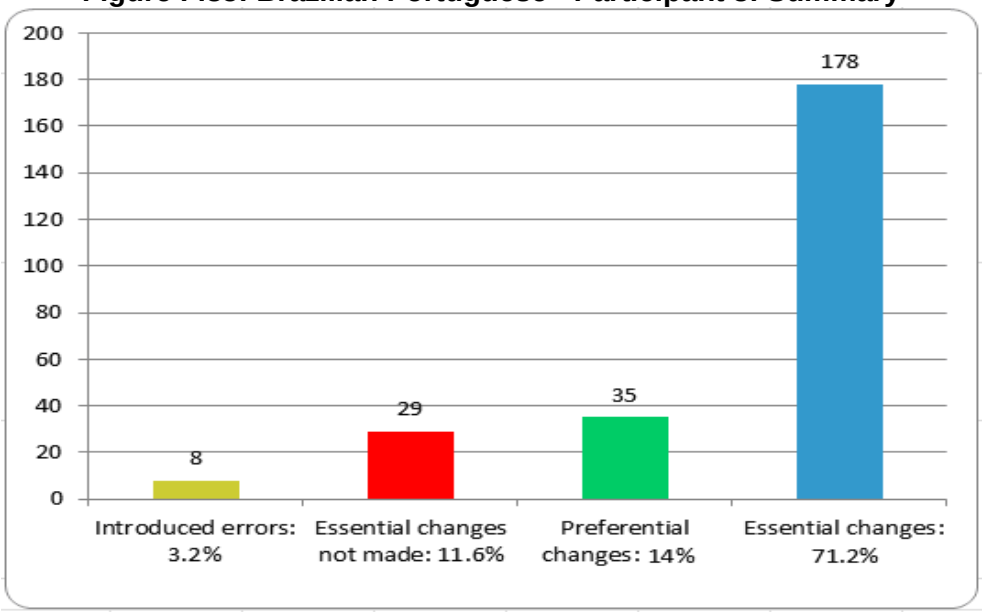


Table F.84: Brazilian Portuguese - Participant 8: Summary

ESSENTIAL CHANGES	178
Preferential Changes	35
Essential Changes Not Implemented	29
Introduced Errors	8
TOTAL	250

Figure F.85: Brazilian Portuguese - Participant 8: Summary



F.20. Brazilian Portuguese - Participant 9

Table F.85: Brazilian Portuguese - Participant 9: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	12	Accuracy - information missing: 10 Accuracy - untranslated text: 2
Format	4	
Language	103	Language - adverbs: 1 Language - capitalisation: 14 Language - conjunctions: 4 Language - determiners: 4 Language - gender: 20 Language - number: 19 Language - phrasal ordering: 12 Language - prepositions: 12 Language - pronouns: 2 Language - punctuation: 3 Language - verbs: 12
Lexical choice	6	
Mistranslation	25	
Style	2	
Total	**152**	

Figure F.86: Brazilian Portuguese - Participant 9: Essential Changes

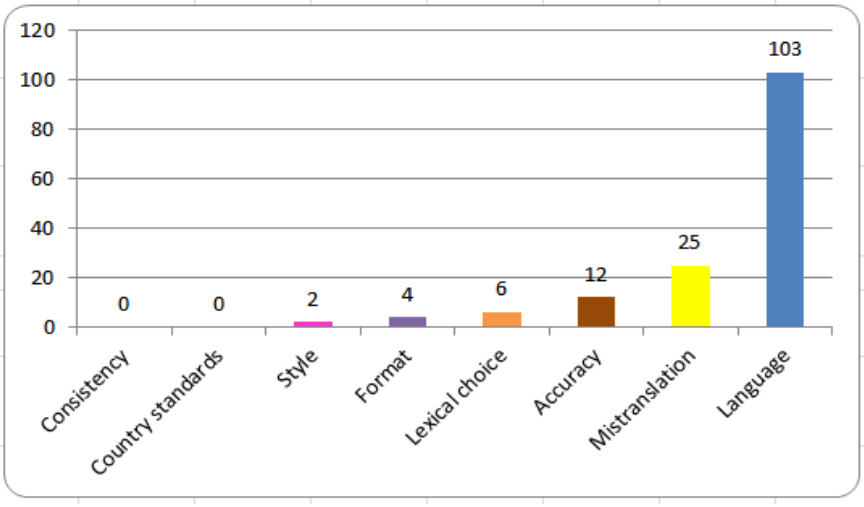


Table F.86: Brazilian Portuguese - Participant 9: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Accuracy	2	Accuracy - extra information: 2
Language	17	Language - determiners: 3 Language - number: 2 Language - phrasal ordering: 2 Language - prepositions: 1 Language - pronouns: 4 Language - punctuation: 2 Language - verbs: 3
Lexical choice	6	
Style	2	
Total	**27**	

Figure F.87: Brazilian Portuguese - Participant 9: Preferential Changes

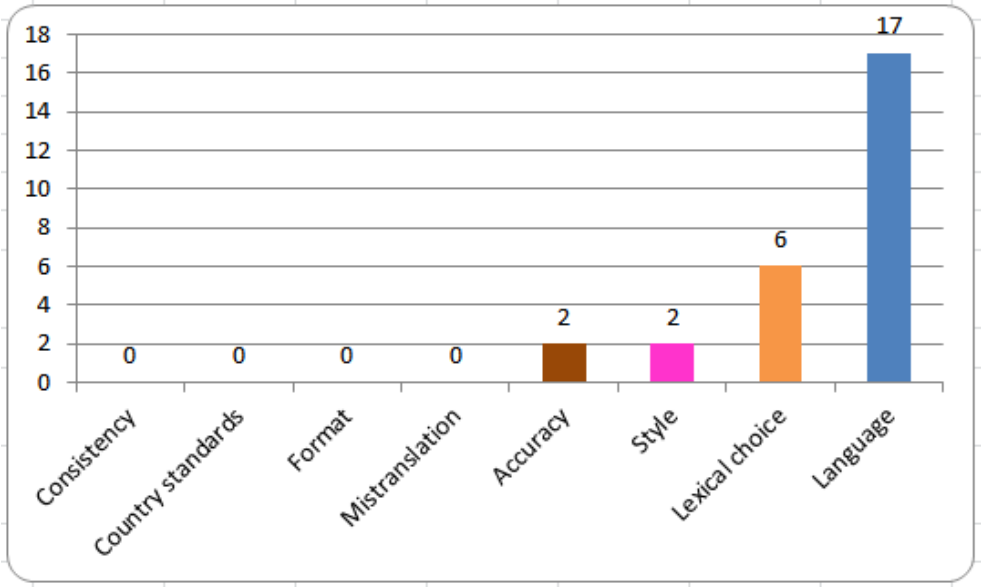


Table F.87: Brazilian Portuguese - Participant 9: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Accuracy	4	Accuracy - information missing: 4
Format	2	
Language	46	Language - adverbs: 1 Language - capitalisation: 8 Language - determiners: 7 Language - gender: 3 Language - number: 9 Language - phrasal ordering: 4 Language - prepositions: 5 Language - pronouns: 5 Language - punctuation: 2 Language - verbs: 2
Lexical choice	1	
Mistranslation	5	
Total	**58**	

Figure F.88: Brazilian Portuguese - Participant 9: Essential Changes Not Implemented

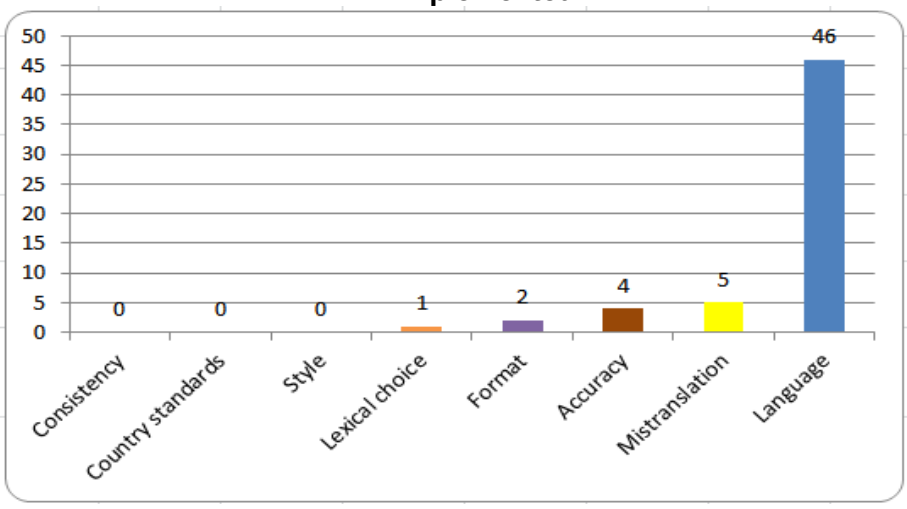


Table F.88: Brazilian Portuguese - Participant 9: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Language	4	Language - determiners: 2 Language - number: 2
Mistranslation	4	
Total	**8**	

Figure F.89: Brazilian Portuguese - Participant 9: Introduced Errors

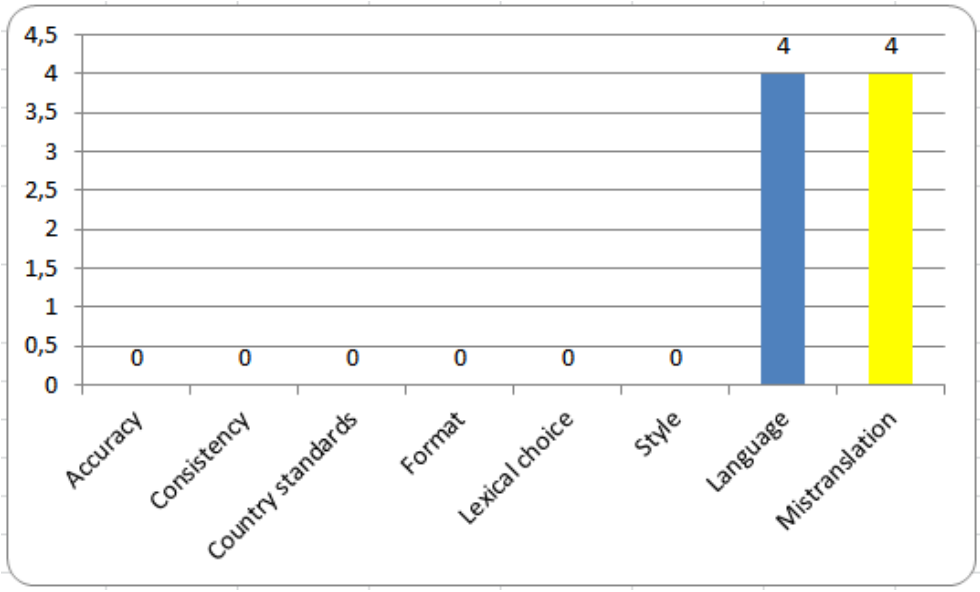
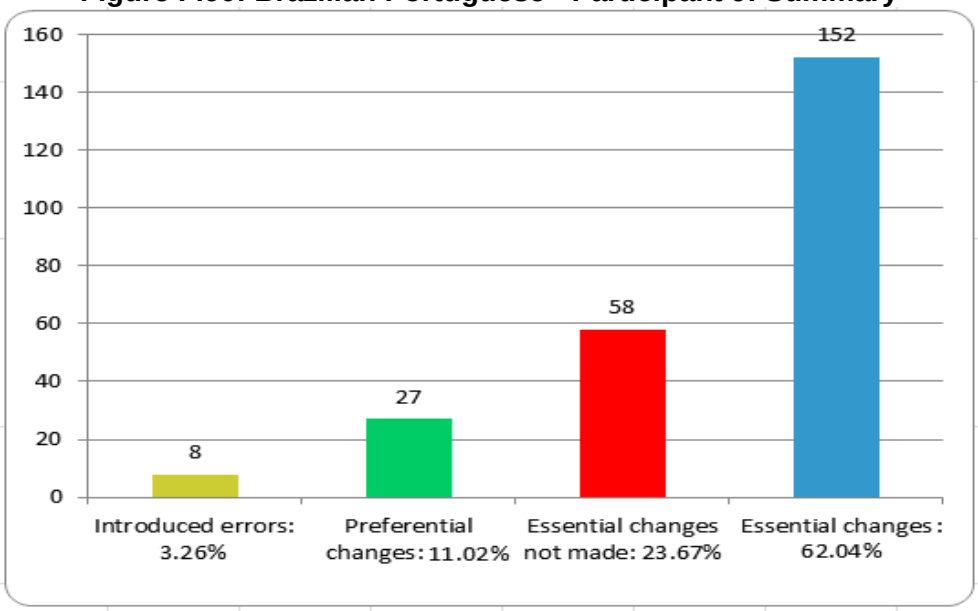


Table F.89: Brazilian Portuguese - Participant 9: Summary

ESSENTIAL CHANGES	152
Preferential Changes	27
Essential Changes Not Implemented	58
Introduced Errors	8
TOTAL	245

Figure F.90: Brazilian Portuguese - Participant 9: Summary



F.21. Brazilian Portuguese - Participant 10

Table F.90: Brazilian Portuguese - Participant 10: Essential Changes

Category	Number of Essential Changes	Subcategories (if applicable)
Accuracy	12	Accuracy - information missing: 10 Accuracy - untranslated text: 2
Format	6	
Language	121	Language - adverbs: 1 Language - capitalisation: 18 Language - conjunctions: 5 Language - determiners: 5 Language - gender: 23 Language - number: 20 Language - phrasal ordering: 14 Language - prepositions: 11 Language - pronouns: 4 Language - punctuation: 5 Language - verbs: 15
Lexical choice	4	
Mistranslation	27	
Style	2	
Total	**172**	

Figure F.91: Brazilian Portuguese - Participant 10: Essential Changes

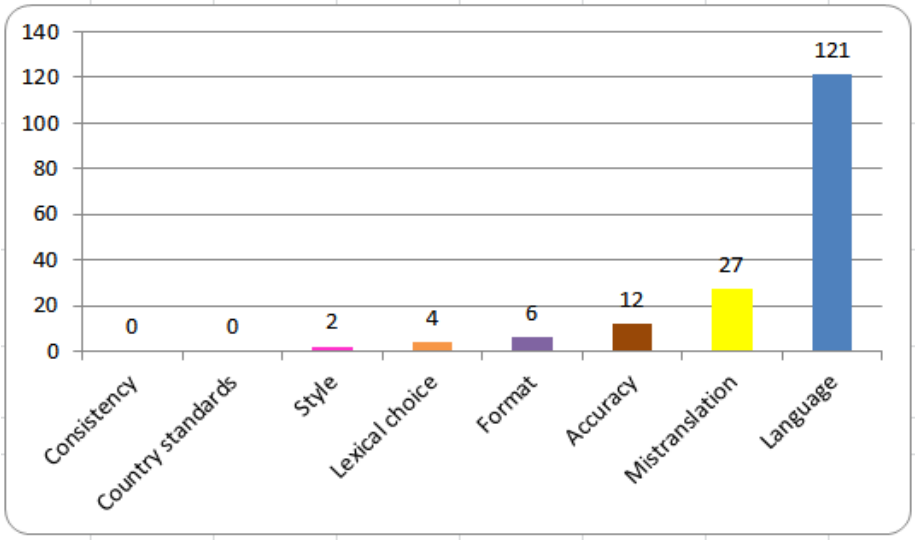


Table F.91: Brazilian Portuguese - Participant 10: Preferential Changes

Category	Number of Preferential Changes	Subcategories (if applicable)
Accuracy	1	Accuracy - extra information: 1
Language	27	Language - conjunctions: 3 Language - determiners: 4 Language - number: 2 Language - phrasal ordering: 3 Language - prepositions: 2 Language - pronouns: 2 Language - punctuation: 1 Language - verbs: 10
Lexical choice	8	
Style	2	
Total	**38**	

Figure F.92: Brazilian Portuguese - Participant 10: Preferential Changes

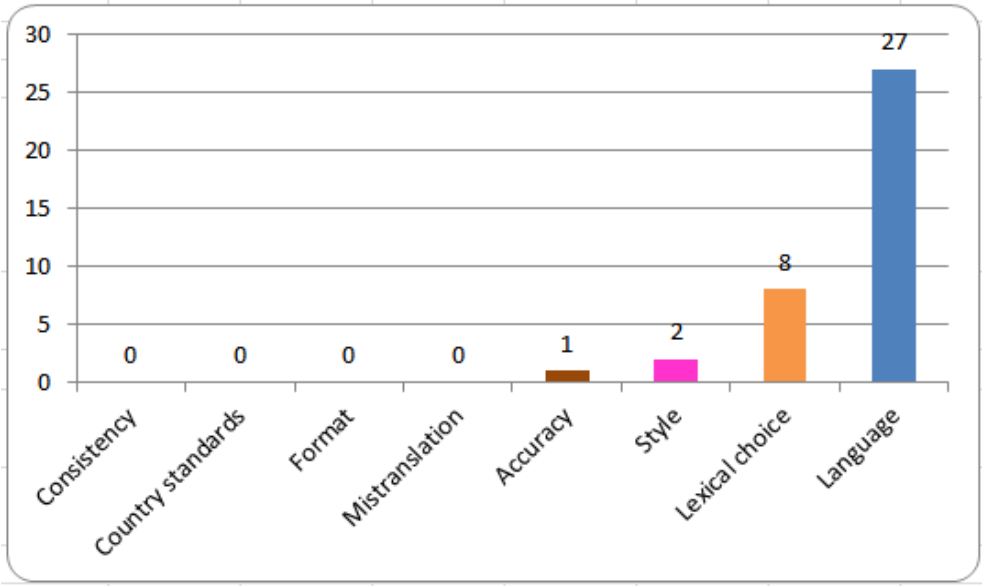


Table F.92: Brazilian Portuguese - Participant 10: Essential Changes Not Implemented

Category	Number of Essential Changes Not Implemented	Subcategories (if applicable)
Language	27	Language - adverbs: 1 Language - capitalisation: 5 Language - determiners: 4 Language - gender: 1 Language - number: 4 Language - phrasal ordering: 2 Language - prepositions: 4 Language - pronouns: 5 Language - verbs: 1
Mistranslation	4	
Total	**31**	

Figure F.93: Brazilian Portuguese - Participant 10: Essential Changes Not Implemented

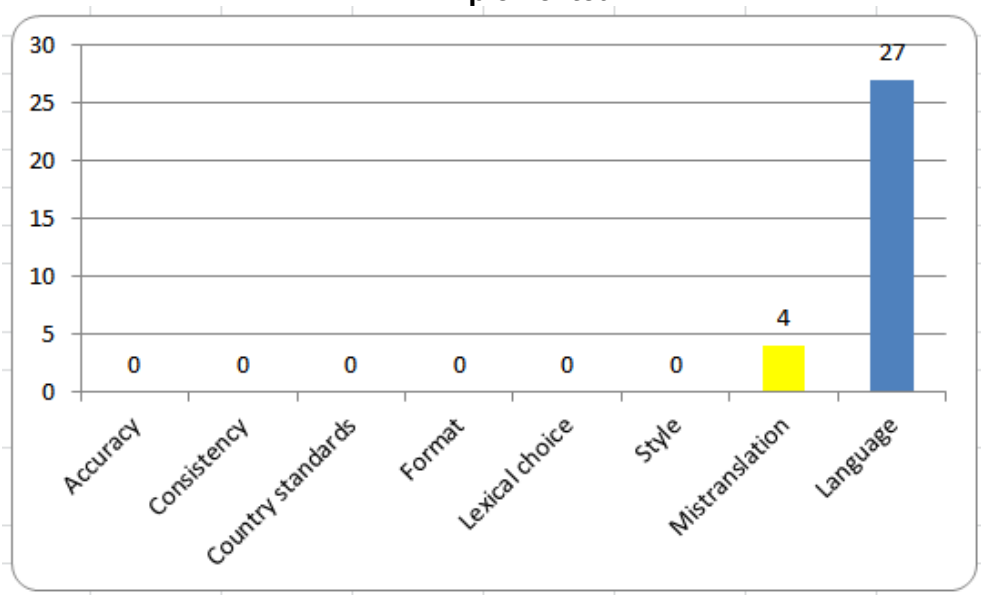


Table F.93: Brazilian Portuguese - Participant 10: Introduced Errors

Category	Number of Introduced Errors	Subcategories (if applicable)
Accuracy	2	Accuracy - information missing: 2
Language	10	Language - capitalisation: 3 Language - determiners: 2 Language - number: 3 Language - phrasal ordering: 1 Language - prepositions: 1
Lexical choice	1	
Mistranslation	2	
Total	**15**	

Figure F.94: Brazilian Portuguese - Participant 10: Introduced Errors

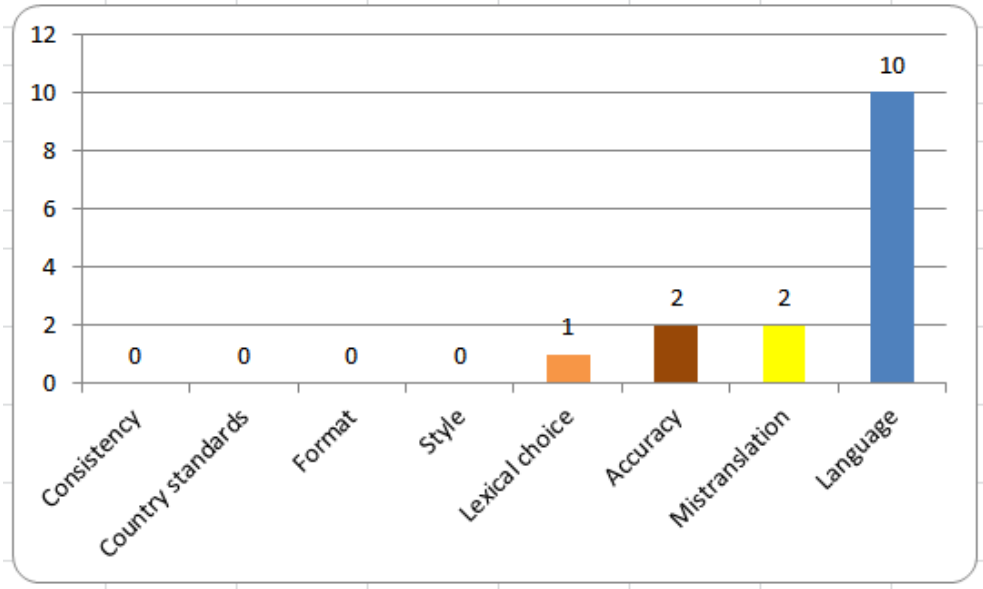
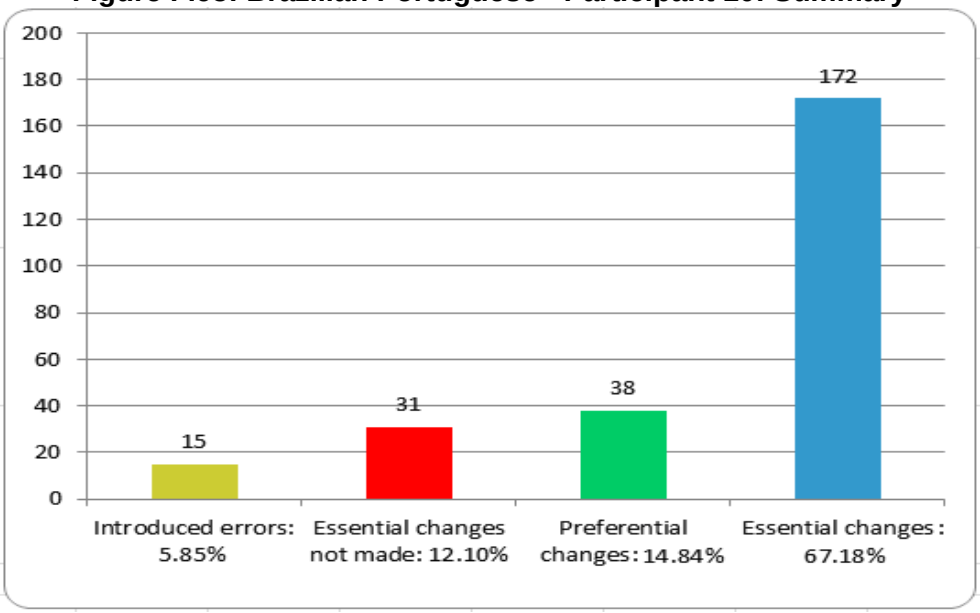


Table F.94: Brazilian Portuguese - Participant 10: Summary

ESSENTIAL CHANGES	172
Preferential Changes	38
Essential Changes Not Implemented	31
Introduced Errors	15
TOTAL	256

Figure F.95: Brazilian Portuguese - Participant 10: Summary



F.22. Summary for all participants combined - Brazilian Portuguese

Figure F.96: Summary for all participants combined - Brazilian Portuguese

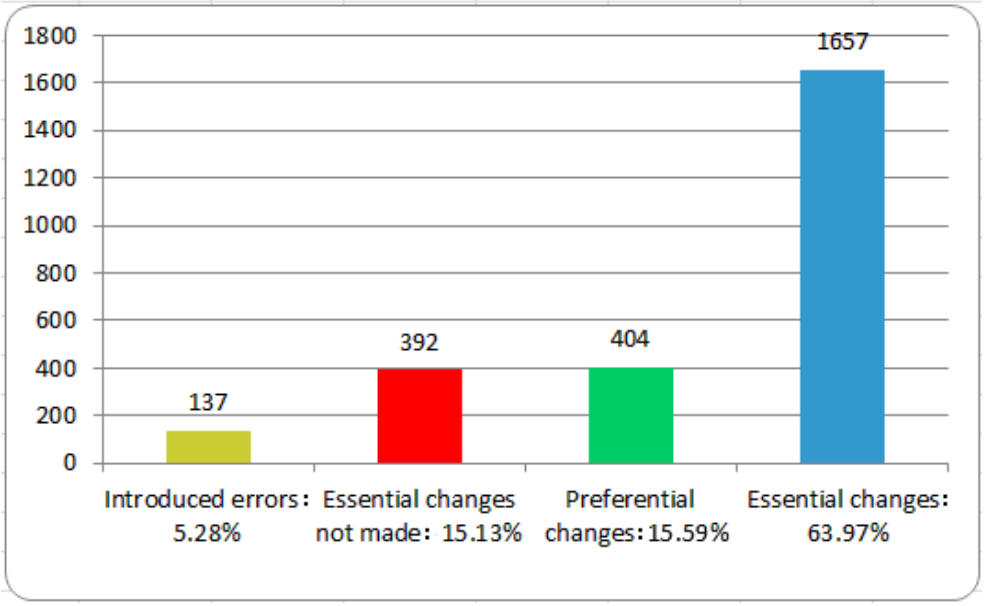
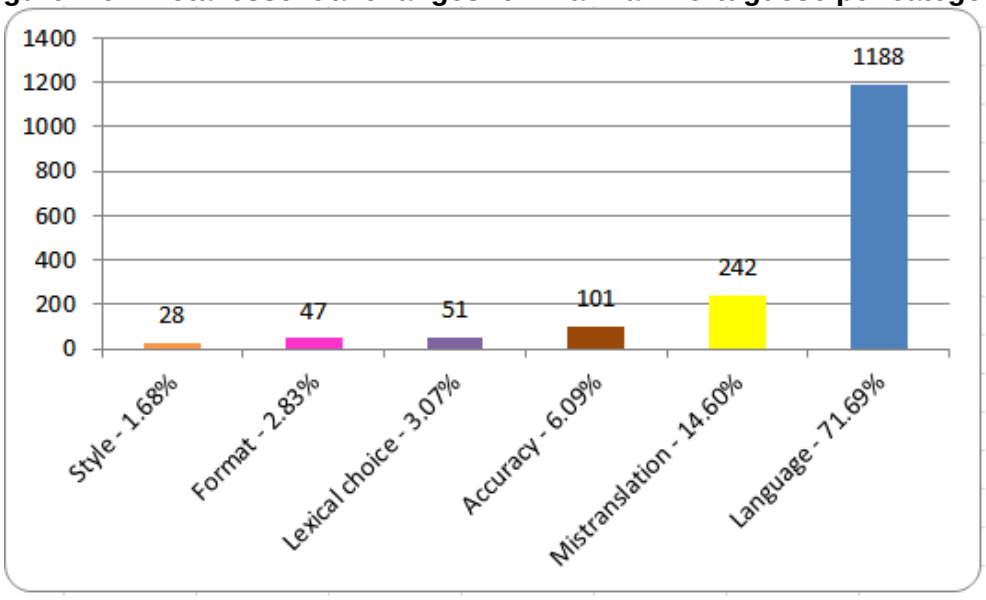


Figure F.97: Total essential changes for Brazilian Portuguese per category



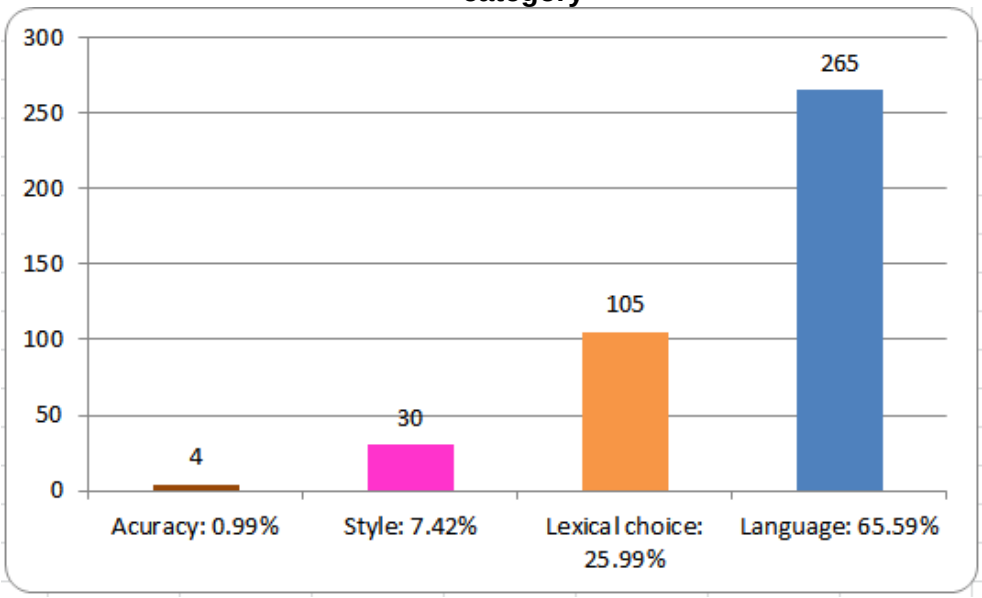
Total of essential changes recorded: 1657

Table F.95: examples of essential changes for Brazilian Portuguese

Source text	Raw MT output	Post-edited text	Comments
The **pop-up** context menu also provides the option to create an additional subfolder	O menu de contexto também fornece a opção de criar uma subpasta **adicionais**	O menu de contexto **pop-up** também fornece a opção de criar uma subpasta **adicional**	Accuracy - Information missing: The term "pop-up", which was missing from the raw MT output, was added by the post-editor after "contexto". Language - Number: The adjective "adicionais" was corrected to "adicional" to agree in number with "subpasta".
To provide a more descriptive name	para fornecer um nome mais descritivo	Para fornecer um nome mais descritivo	Language - Capitalisation: "Para" was not capitalised in the raw MT output, so this was corrected by the post-editor.
How to distribute components so they are equally spaced	Como distribuir componentes de maneira que **são** igualmente espaçados	Como distribuir componentes de maneira que **sejam** igualmente espaçados	Language - Verb: The verb "são" (present of the indicative) was corrected to "sejam" (present of the subjunctive).

Source text	Raw MT output	Post-edited text	Comments
For example, infrequently used components	Por exemplo, **infrequently utilizados os componentes**	Por exemplo, os componentes utilizados com pouca frequência	Accuracy - Untranslated Text: The term "infrequently", which was not translated in the raw MT output, was translated by the post-editor. Language - Phrasal Ordering: The phrasal ordering was corrected.
To finalize the change and update the model, press the {207}Enter{208} key on the keyboard	Para finalizar a alteração e atualizar o modelo, pressione a tecla **Enter{208} {207}** no teclado	Para finalizar a alteração e atualizar o modelo, pressione a tecla {207}Enter{208} no teclado	Format: The order of the tags was corrected.

Figure F.98: Total preferential changes for Brazilian Portuguese per category

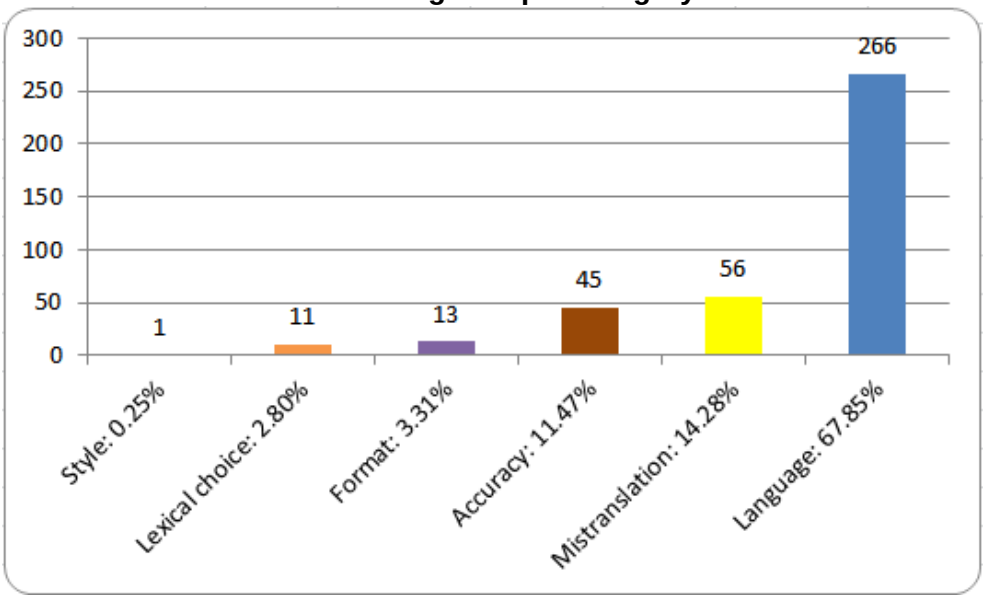


Total of preferential changes recorded: 404

Table F.96: examples of preferential changes for Brazilian Portuguese

Source text	Raw MT output	Post-edited text	Comments
The Factory Options settings are global so that all new documents use the settings.	As configurações de Opções de família são global, para que todos os documentos novos, **utilize as configurações**.	As configurações de Opções de fábrica são globais para que todos os documentos novos **as usem**.	Style: "as configurações" was replaced by the pronoun "as". While this is more elegant, it was not a necessary change. Lexical Choice: The verb "utilizar" was replaced with the verb "usar", which was not necessary either.
The option to turn on or off the visibility of selected drawing layers is also provided.	A opção para ativar ou desativar a visibilidade das camadas do desenho selecionado **também é fornecida**.	**Também é fornecida** a opção para ativar ou desativar a visibilidade das camadas do desenho selecionado.	Language - Phrasal Ordering: The phrasal ordering was unnecessarily changed.
To provide a more descriptive name for the new folder	**para** fornecer um nome mais descritivo para a nova pasta	**A fim de** fornecer um nome mais descritivo para a nova pasta	Lexical choice: "Para" was unnecessarily replaced with "A fim de".
under the new folder you have created	na nova pasta **que você criou**	na nova pasta **criada**	Style: The style was unnecessarily changed.

Figure F.99: Total essential changes not implemented for Brazilian Portuguese per category

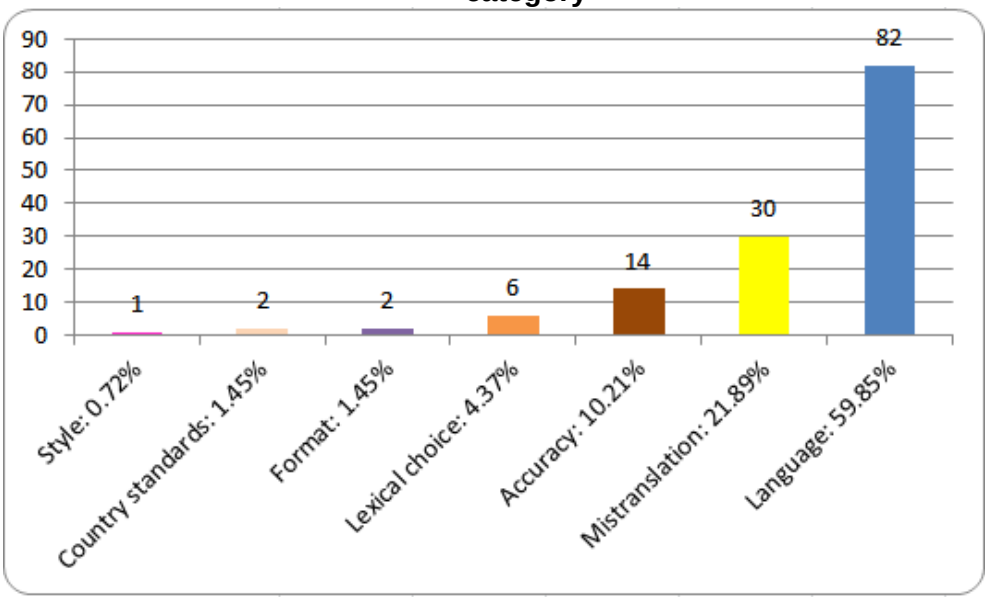


Total of essential changes not implemented recorded: 392

Table F.97: examples of essential changes not implemented for Brazilian Portuguese

Source text	Raw MT output	Post-edited text	Comments
The drawing may be repositioned	O desenho pode ser **reposicionada**	O desenho pode ser **reposicionada**	Language - Gender: "reposicionada" should have been corrected to "reposicionado" to agree in gender with "desenho".
A minimum of 39 MB of free disk space is required on the drive where your temporary files are directed.	Um mínimo de 39 MB de espaço livre em disco é necessário na unidade **onde** seus arquivos temporários são direcionados.	Um mínimo de 39 MB de espaço livre em disco é necessário na unidade **onde** seus arquivos temporários são direcionados.	Accuracy - Information Missing: The preposition "para" should have been added before "onde".
Check boxes enable layer visibility to be toggled on or off	As caixas de verificação Ativar a visibilidade da camada **para ser** ativada ou desativada	As caixas de seleção permitem a visibilidade da camada **para ser** ativada ou desativada	Language - Conjunction: The conjunction "que" should have been added after "permitem". Accuracy - Extra Information: The preposition "para" should have been removed. Language - Verb: The verb "ser" should have been inflected to the present of the subjunctive ("seja").
The components selected in step 2 move to align with the reference component.	Os componentes selecionados na **Etapa** 2 mover para **alinhar** com o componente de referência.	Os componentes selecionados na Etapa 2 movem-se para **alinhar** com o componente de referência.	Language - Capitalisation: "Etapa" should not be capitalised. Language - Pronoun: The pronoun "se" should have been added after "para".

Figure F.100: Total of introduced errors for Brazilian Portuguese per category



Total of introduced errors recorded: 137

Table F.98: examples of introduced errors for Brazilian Portuguese

Source text	Raw MT output	Post-edited text	Comments
Aligns the selected component(s) to the X axis orientation	Alinha o componente selecionado(s) **para** a orientação do eixo X	Alinha o(s) componente(s) selecionado(s) **na** orientação do eixo X	Language - Preposition: The preposition "para", which was incorrect in the raw MT output, was replaced with another incorrect preposition for this context, "em" ("em" + "a" = "na").
{1}May 13, 2010{2}	{1}**pode 13 2010**{2}	{1}**Maio 13, 2010**{2}	Country - Date/Time Format: Although the mistranslation in the raw MT output was corrected (from "pode" to "maio"), the date format used by the post-editor is not the correct date format used in Brazil.
{14}Change the Active Project{15}	{14}**a**lterar o projeto ativo{15}	{14}**A**lterar o **P**rojeto **A**tivo{15}	Language - Capitalisation: Only "Alterar" needed to be capitalised here. Although translators may be asked to follow the same capitalisation as in English in some projects, that was not the case here, and the standard capitalisation rules for Brazilian Portuguese should have been followed.