UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE ARTES E COMUNICAÇÃO

DEPARTAMENTO DE LETRAS

CURSO DE LETRAS INGLÊS (LICENCIATURA)

AÍLA FONSECA DE ALBUQUERQUE

THE IMPACT AND CHALLENGES OF USING MACHINE TRANSLATION POST-EDITING IN SWORN TRANSLATION OF BRAZILIAN PERSONAL DOCUMENTS

Recife

2023

AÍLA FONSECA DE ALBUQUERQUE

THE IMPACT AND CHALLENGES OF USING MACHINE TRANSLATION POST-EDITING IN SWORN TRANSLATION OF BRAZILIAN PERSONAL DOCUMENTS

Article written as a requirement for the attainment of the degree Bachelor in English Language Teaching
Supervisor: Prof. Dr Arlene Koglin

Recife

2023

AÍLA FONSECA DE ALBUQUERQUE

THE IMPACT AND CHALLENGES OF USING MACHINE TRANSLATION POST-EDITING IN SWORN TRANSLATION OF BRAZILIAN PERSONAL DOCUMENTS

Article written as a requirement for the attainment of the degree Bachelor in English Language Teaching

Approved on: /09/2023

EXAMINING BOARD

_____

Prof. Dr. Arlene Koglin (Supervisor)
Universidade Federal de Pernambuco

_____

Prof. Dr. Igor Antônio Lourenço da Silva (Examiner)
Universidade Federal de Uberlândia

Abstract:

Machine Translation Post-Editing (MTPE) and Sworn Translation are both well-known areas in the translation field. However, there is a lack of study regarding the use of MTPE for the translation of personal documents. More precisely, the translation of ID cards and birth certificates from Brazilian Portuguese to American English. In this sense, the objectives of this article were to investigate the productivity of MTPE in Sworn Translation, as well as to describe the overall activity and report on the description of visual elements that appear on documents. The methodology consisted of a timed translation with commentary with MTPE and Human Translation (HT), as well as the description of elements and final formatting that is characteristic of Sworn Translation and the use of an OCR recognition software to convert the PDF files to Word. The results showed that depending on circumstances such as the resolution of the source file, the number of elements to describe, and the lack of use of a Translation Memory (TM), MTPE may be less productive than HT in Sworn Translation. However, this does not mean that this type of technology is not important to translators in this particular field. The results also showed that if the document is in high resolution, and the time for formatting is not taken into consideration, MTPE saves time compared to HT. In this sense, it is a matter of knowing when it is appropriate to apply these tools. And, to do so, the community of translators ought to be knowledgeable and open-minded about using this technology.

Keywords: Machine Translation Post-Editing; Sworn Translation; personal documents.

Resumo:

Machine Translation Post-Editing (MTPE) e Tradução Juramentada são duas áreas reconhecidas no campo da tradução. No entanto, há uma falta de estudos em relação ao uso de MTPE para a tradução de documentos. Mais precisamente, a tradução de carteiras de identidade e certidões de nascimento do português brasileiro para o inglês americano. Nesse sentido, os objetivos deste artigo foram investigar a produtividade do MTPE na Tradução Juramentada, bem como descrever a atividade em geral e relatar a descrição dos elementos visuais que aparecem nos documentos. A metodologia consistiu em uma tradução comentada cronometrada utilizando MTPE e Tradução Humana (TH), assim como a descrição dos elementos e formatação final característica da Tradução Juramentada, além do uso de um software de reconhecimento OCR para converter os arquivos de PDF para Word. Os resultados mostraram que, dependendo de circunstâncias como a resolução do arquivo de origem, o número de elementos a serem descritos e a falta de uma Memória de Tradução (TM), MTPE pode ser menos produtivo do que HT em Sworn Translation. No entanto, isso não significa que esse tipo de tecnologia não seja importante para tradutores deste campo. Os resultados também mostraram que, se o documento estiver em alta resolução e o tempo de formatação não for levado em consideração, MTPE pode exigir menos esforço temporal do que HT. Nesse sentido, trata-se de saber quando é apropriado aplicar essas ferramentas. E, para fazer isso, a comunidade de tradutores precisa estar informada e aberta ao uso desta tecnologia.

Palavras-chave: Pós-edição de Tradução Automática; Tradução Juramentada; documentos pessoais.

1. **INTRODUCTION**

Technology has been revolutionizing the activity of translation. Previously, translators only used simple software in their everyday work activities such as Word and Excel by Microsoft, but nowadays their reality has advanced to powerful Computer-aided Translation Tools, Translation Memory, Cloud-base typing assistants, AI (Artificial Intelligence) translations, and so on. Consequently, it is not surprising that the activity of translation in diverse fields has changed — including Sworn Translation.

Portal Gov.br, a website that provides information relating to the Brazilian Federal Executive, defines Sworn Translation as a translation made by a public translator. It is also defined as the official translation of documents (Mayoral, 2000). Unlike translators from other areas, a sworn translator can validate the content of a translated legal or official document. In Brazil, Getúlio Vargas established the Sworn Translation in 1943 by Decree no. 13.609, Article no. 18, declaring that for books, papers, or documents, to have any national validity, they must be accompanied by a version in Brazilian Portuguese, which should be written by translators appointed by the regulation. The following presidents maintained and amended the legislation, and Sworn Translation remains the bridge between Brazil and the world to this day.

Despite Sworn Translation being a widely popular modality in the market, the same cannot be observed in terms of studies in Brazil, since the number of works produced and/or published is still scarce (Goederte, 2015; Nascimento, 2006). Consequently, the small number of studies may limit the shared knowledge about certain aspects of such an important activity, especially when adding Machine Translation Post-editing (MTPE) to it. For example, given the particular characteristics and elements of a document (coat of arms, stamps, seals, etc), the experience of using MTPE may be drastically different in an ID card, which is characterized by its smaller content, in comparison to a birth certificate that usually contains numerous stamps, seals, etc.

In this sense, this article aims to investigate the productivity of using MTPE in the Sworn Translation of personal documents. More specifically, it aims: to describe some challenges of using MTPE in the translation of birth certificates and ID cards from Brazilian Portuguese to American English and to report on the description of visual elements that appear on documents.

As explained by Holmes (1972), for a solid foundation of knowledge in a translation field to exist, it is necessary that the theory and the practical aspects have a dialectical relationship in which one supplies knowledge to the other. However, this is not the case for Sworn Translation, hence the disparity between the great number of texts translated and the scarce published studies. Consequently, this article and others in the future are fundamental to the continued development of expertise and educational materials for future professionals.

Thus, this article is organized as follows: in section 2, the key concepts relating to Sworn Translation and MTPE will be discussed before moving on to the combination of the two. Then in section 3, the methodological aspects of this study will be described. Next, section 4 presents the qualitative and quantitative analysis of the data. Finally, the final remarks are presented in section 5.

2. THEORETICAL FRAMEWORK

2.1 Sworn Translation

The translation activity goes as far back as 3000 B.C. with the work "Rosetta Stone". It also holds a measureless impact on History, as illustrated by the translation of the Odyssey and the Bible (Sawant, 2013). So, it is not surprising that such an ancient activity would have grown in a field with various branches of translation, such as literary, multimedia, localization, and specialized.

Specialized translation can be defined as a non-literary translation that requires specific knowledge about other areas, such as "science and technology, economics, marketing, law, politics, medicine, and mass media [...]" (Gotti & Šarčević, 2006). Additionally, this extensive branch covers the translation of legal texts, which itself has its own particularities and typologies. (Guedes, 2016).

Legal Translation Studies are concerned with "all aspects of translation of legal texts, including processes, products, and agents." (Ramos, 2014). One of these activities is called Sworn Translation and it refers to the "official, written translation of documents" (Mayoral, 2000). Its purpose is to grant national validity and/or authenticity to foreign documents in a country. However, Sworn Translation is not limited to legal texts, but any type of text in a situation that would require the authentication of a sworn translator. (Gutíerrez, 2015). For

instance, social media messages in a foreign language that could serve as evidence in a court case to substantiate a criminal offense would have to be translated by a sworn translator.

In Brazil, it is not simple to characterize Sworn Translation, since there is not a strict set of rules regarding the formatting and description of elements that appear in the documents. In her dissertation, Goedert (2015) gathers official normative data for translating documents from the Junta Comercial de Minas Gerais, but they are not always followed by all public translators — each area of Brazil has its own group of such translators, in which they share their own decisions (especially when it comes to companies and trainees/assistants).

However, there are some general characteristics that are mandatory. The document must bear a sworn translator's name, signature, and seal, as well as the transcription/mention of all elements contained in a document. For example, a Brazilian birth certificate contains a coat of arms that could (hence the lack of official standardization of the way of describing) be described as "[There appears the coat of arms of the Federative Republic of Brazil]". Other examples of elements would be signatures, stamps, seals, illegible texts, watermarks, text in other languages, and blank spaces. In addition, the text is usually formatted as plain text, with few alterations in typography (i.e. font weight, size, justification).

Thus, using MTPE in Sworn Translation can be challenging. For instance, if a document was issued in a format different than plain text (e.g. in tables), or with inconsistent typography throughout the document, the time spent correcting these aspects would add up to the overall activity. Finally, in order to further reflect on the use of Machine Translation Post-Editing in Sworn Translation, this concept needs to be defined.

2.2 Machine Translation Post-editing (MTPE)

Hutchins and Somers (2002) define Machine Translation (MT) as computer systems that translate natural languages and can be human-assisted or not. The power of such systems can be magnified by Translation Memories (TM), which is a "type of linguistic database that is used to store source texts and their translations" (Bowker, 2002).

When using MT with a robust TM, a translator's work is improved logistically and linguistically. In the first place, the time spent translating is reduced, since the main idea behind TM is reusing previous translations. And, if the memory is up-to-date with the current terminologies, mistakes, and extra research may be avoided. In consequence, more productivity and consistency may be achieved by using such a database.

Even with all the advancements in machine translation, human intervention is still needed to ensure the correctness and adequacy of the translated text. In this sense,

post-editing is a crucial part of the translation activity. Mesa-Lao (2013) defines post-editing as "reviewing a pre-translated text generated by an MT engine against an original source text, correcting possible errors, in order to comply with set quality criteria in as few edits as possible".

Furthermore, Mesa-Lao (2013) affirms that there are different levels of post-editing that range from full post-editing to fast post-editing. The first one aims to modify the output text in a way that a reader would not identify it came from a machine. So, it requires thorough editing, but "in less time than translating from scratch as to not defeat the purpose of using MT output". In the case of fast post-editing, as the name suggests, it requires minimal editing given the context of translation: if it is urgent and/or be used for a limited occasion, for example.

Since this article aims to investigate the activity of Sworn Translation, and therefore the output would be used in official agencies that require the clearest and most correct language, the chosen rules are the guidelines for full post-editing. Mesa-Lao (2013) advises that the source segment should be read before the MT suggestion, so the post-editor can decide if the MT output can be used or not as shown in Table 1:

Table 1 - Conditions for MT output use

The MT output should be used when:	The MT output should not be used when:
1. Large pieces of the sentence/term are correct.	1. Raw MT does not make any sense and it would take longer to post-editing than to translate.
2. The raw MT quality is very high, needing only minor corrections.	2. The post-editor tries to understand the raw MT for minutes but can not.
3. The raw MT contains various errors, but the post-editor types slow, so post-editing would be faster.	3. There are many tagging problems.
4. The MT ouput is correct and understanding.	4. Many changes in the raw MT need to be made.

Reference: adapted from Mesa-Lao (2013).

Additionally, O'Brien (2009) set the following full post-editing guidelines:

> "The message transferred should be accurate; Grammar should be accurate; Ignore stylistic and textuality problems; Ensure that key terminology is correctly translated; Edit any offensive, inappropriate or culturally unacceptable information; All basic rules regarding spelling, punctuation, and hyphenation still apply; For tagged formats, ensure all tags are present and in the correct positions; Throughput expectations: high; Quality expectations: medium."

The guidelines proposed by O'Brien are useful for guiding the work of an MT post-editor in any type of translation, including legal documents. This is especially relevant concerning the assurance of terminology correctness and cultural adaptation. For instance, in documents issued in the U.S., many acronyms can be found (e.g. CA for Court of Appeals, CS for Constitution, and CL for Common Law). If these acronyms are not disclosed and/or translated into the Brazilian context, the understanding of the entire document may be compromised.

Bearing all this in mind, what is the reported reality of using MTPE in the Sworn Translation of documents? To answer this question, a targeted search for published works involving this practice was conducted, and despite yielding fewer results, one study proved particularly valuable as discussed in the following subsection.

2.3 The Practice of Sworn Translation using MTPE

Oliveira (2012) reports on the practice of an experienced sworn translator who was selected by the 1980 JUCESP public recruitment examination. The interviewee had previously worked with technical texts, such as automotive engineering and marketing, and found MT helpful. However, as she started moving her attention to official juridical texts, her point of view changed. For her, this type of text requires weighted decisions, that is a specific decision-making process for each situation and/or document, and therefore could not find the use of MT particularly advantageous. In contrast, she found that creating and using a personal database in Word of previously translated documents as models for future translations is the better option concerning time and productivity.

Nonetheless, it is important to counter-argue that MTPE may be useful in Sworn Translation. Documents have a rather regular pattern in content and vocabulary within a said country; for instance, two birth certificates from the same year, but one issued in Pernambuco and the other in Brasília, will have many matches among them. In this sense, using MT would erase the necessity of translating from scratch sentences like "O referido é verdade e dou fé"

On the other hand, Oliveira (2012) states that the pattern found in documents is not enough to establish the use of MT and TM in Sworn Translation as fruitful. The author explains that the level of individuality and diligence prevent the usage of such tools; which is a statement that may be challenged — especially if full post-editing is taken into consideration. Of course, not all documents will have a great number of exact matches, but that does not mean that those which were processed by MT and/or TM could not contribute to the overall activity, especially when considering long documents with continuous texts like

Income Tax, or common terms found in School Transcripts, such as "Mathematics" or "Portuguese Language".

Finally, Oliveira (2012) also points out that another impediment to using MT would be the extreme secrecy that most documents require. However, according to the Smartcat website, which is a cloud-based translation and localization CAT Tool, all information is saved on the cloud, so the translator is protected in case of viruses, or having their device stolen. Furthermore, users are not able to access data from other accounts, and this is ensured by "three major authentication systems: ADFS, Azure AD, and Okta." On the other hand, it is important to keep in mind that these are words from the company. They obviously would state such information to attract professionals, so these claims have to be questioned or carefully considered. Ultimately, Smartcat was launched in July 2015, so Oliveira's argument could have not included the current service provided by this company and others alike.

Given this information and considering that there have been advancements in MT quality since the publication of the aforementioned research, it is valid to investigate the productivity and impacts of MT in Sworn Translation currently.

3. METHODOLOGY

3.1 Data collection

The data collection was based on documents that were publicly available on the Internet, using Google and Bing search engines. However, given the challenge of finding examples with appropriate resolutions for OCR Recognition, the author of this article also utilized her personal documents.

Along with the image quality concern, the selection criterion was centered on choosing two types of the most requested personal documents for Sworn Translation, as proposed by (Barros, Babini, Aubert, 2010), i.e. birth certificates and ID cards. The authors divided them into five categories: personal documents (which include birth certificates and ID cards); corporate documents; financial-commercial documents; legal documents; and other types of documents.

3.2 Material

According to the Portal Gov.br, an ID card corresponds to the civil identification of the Brazilian people. Its purpose is to ensure individuality and identification in various societal

acts. This official website further defines a birth certificate as the first document with legal validity that grants the name, surname, nationality, filiation, and civil and social rights to a person. In other words, it is proof of a citizen's existence.

In order to ensure impartiality in terms of translation time and speed, four examples of these two documents were utilized. Since this research is not equipped with a well-fed TM, and therefore the number of matches would be minimal to virtually absent during the MTPE, the analysis was limited to the results of Machine Translation. The chosen MT was Smartcat's General Base System, given the author's familiarity with this tool.

3.3. Procedures

This study consisted of both Human Translation (HT) and MTPE with commentaries about their processes carried out by the author of this article. The documents entitled "ID Card 1", "Birth Certificate 1, "ID Card 2", and "Birth Certificate 2" were translated using HT, whereas "ID Card 3", "Birth Certificate 3", "ID Card 4", and "Birth Certificate 4" were subjected to MTPE. Sensitive information that is protected by law, such as names, dates of birth, places of birth, etc, was only censored after completing the translation activities, as this information is relevant to the quantitative analysis.

The first step was to set a stopwatch to separately register the time spent for HT, MTPE, and formatting as well as for describing the contained elements in each document. For HT, the screen was divided into two sections: the right side displayed the document to be translated, while the left side presented an empty Word file.

For MTPE, initially, the documents were converted from PNG/PDF to Word files using ABBYY FineReader PDF as the OCR recognition software. The selection was performed manually, selecting only the text while ignoring elements such as "coat of arms", and seals. All errors were corrected within the platform itself.

Then, the files were uploaded to Smartcat, and Portuguese (Brazil) was selected as the source language and English (USA) as the target language. Next, the option "Use machine translation" was checked, along with "Use for free with feedback". No Translation Memory was added due to the fact that the results using a small-size TM would not be fruitful or even productive. Finally, the documents were machine-translated and post-edited following the full post-editing guidelines (O'Brien, 2009).

After completing each project, the resulting file was downloaded. Subsequently, a new Word document was created with the following specifications: single spacing, Times New Roman font, size 11, and justified text alignment. The content from the completed file was

then pasted into this new document without any formatting (unformatted Unicode text). This process aimed to convert the content into plain text, eliminating any extra spaces, bold/italic fonts, or other styles that could differentiate one sentence or word from another. For instance, the OCR of "ID Card 3" displayed "FEDERATIVE REPUBLIC OF BRAZIL" in an underlined 16-sized Arial font in black color, while "STATE OF PERNAMBUCO" was shown in a 21-sized, Times New Roman font, in green color. Moreover, the text was spread across two pages without any discernible structure or organization.

3.4 Data analysis

Data were analyzed both quantitatively and qualitatively. In order to evaluate the productivity of using MTPE, the temporal effort of Human Translation and MTPE were compared. As defined by Krings (2001), temporal effort consists of the time spent post-editing an MT output.

The time spent with OCR Recognition and final formatting was registered as well. Moreover, the MT errors and PE alterations were analyzed qualitatively and separated into tables per document.

4. ANALYSIS AND DISCUSSION

In order to assess the productivity of using MTPE in Sworn Translation, this section will present the MT output analysis, as well as the time comparison of the time spent on HT and MTPE as shown in Table 2.

Table 2 - Time spent on HT and MTPE

HT	Duration (min/sec)	MTPE	Duration (min/sec)
ID Card 1	04:03	ID Card 3	02:10
ID Card 2	03:59	ID Card 4	02:53
Birth Certificate 1	07:38	Birth Certificate 3	03:47
Birth Certificate 2	12:01	Birth Certificate 4	02:36

As shown in Table 2, the total time for MTPE is shorter than HT. This is particularly evident in birth certificates. For example, HT took a total of 19 minutes and 39 seconds in total for "Birth Certificate 1" and "Birth Certificate 2". In contrast, MTPE only required 6 minutes and 23 seconds for "Birth Certificate 3" and "Birth Certificate 4". The time difference between MTPE and HT in ID cards is smaller, but still relevant: in general, MTPE

took 2 minutes and 59 seconds less than HT. Overall, to translate all documents, HT took 27 minutes and 41 seconds, and MTPE took 11 minutes and 26 seconds – resulting in HT being 57.79% longer than MTPE. So, this average may suggest that MTPE is more productive than HT.

These results might go against the previous research conducted by Oliveira (p. 192, 2012), who stated that technologies relating to MT or CAT tools could not be applied to Sworn Translation given the uniqueness and care required to translate documents. Because if a professional is able to translate faster with MTPE rather than with HT, they might have more time to improve and adapt the translated text, consequently perhaps producing a throughout work.

However, it is worth mentioning that OCR recognition can add up to the time of MTPE as can be seen in Table 3.

Table 3 - Time spent on OCR recognition

MTPE	Duration (min/sec)
ID Card 3	04:34
ID Card 4	02:58
Birth Certificate 3	08:30
Birth Certificate 4	02:36

Adding the time values of OCR recognition to the MTPE total time of the four documents (11 minutes and 26 seconds) resulted in an increase of 16 minutes and 38 seconds, accumulating 27 minutes and 4 seconds. Still, MTPE and OCR Recognition saved 37 seconds in comparison to HT. This considerable increase was caused by several challenges involving the source documents and their OCR Recognition. For example, documents usually are issued using paper currency in Brazil. This type of paper contains security elements such as security threads, watermarks, and microprinting that aim to make counterfeiting more difficult. In the "ID Card 3" scanning, many of these elements were visible, which consequently caused the insertion of mistaken characters. Moreover, some texts had a vertical configuration, such as "Polegar Direito", and this caused poor recognition. Finally, stamps and small-sized fonts also caused more errors.

In fact, the type of paper is not the only reason why OCR Recognition may become challenging. The Birth Certificate 3 was not issued in an official paper but had particular formatting. The content was organized using dotted lines and this added to the time of OCR Recognition and the time of final formatting since the resulting file should be preferably

organized as plain text. Moreover, the document was written using a type of font that is not very readable, and some parts of the text were faded.

The challenging aspect of OCR Recognition and scanning is also pointed out by Oliveira (2012, p. 202): "The time used to scan the documents and correct scanning errors (which occur even with an advanced scanner) would make the process much more time-consuming."[1] In a company, a potential solution for this problem would be the separation of these two tasks into different teams: one for formatting, and the other for MTPE. Additionally, in the case of a freelancer translator, the professional could ask for digital documents or high-resolution scanning only.

Finally, Table 4 shows the time spent on the description and formatting of the documents.

Table 4 - Time spent on description of elements and formatting

Description/Formatting	Duration (min/sec)
ID Card 3	02:52
ID Card 4	03:38
Birth Certificate 3	03:26
Birth Certificate 4	07:35

The time spent manually describing elements and formatting were the last activities for using MTPE in this research. Both tasks of description and formatting totaled 17 minutes and 32 seconds in contrast to the 11 minutes and 26 seconds for MTPE in the four documents.

In general, the main difficulty in formatting SmartCat's resulting file was based on the structure of the texts found in the documents, bearing in mind that the usual formatting for Sworn Translation is plain and continuous text. For instance, the sentences in ID cards have a vertical configuration, e.g. in "ID Card 3", "Birthplace" appears above "Recife - PE", which is not the preferable ordering. Moreover, "Birth Certificate 3" is composed of sentences with little to no connectives or punctuation. In a similar situation, "Birth Certificate 4" is mainly formed by tables. When they were converted to plain text, the content was not organized cohesively. So, the activity of reorganizing the text demanded a considerable amount of time.

The total duration for describing the elements and formatting, MTPE, and OCR recognition was 44 minutes and 36 seconds. As a result, HT ended up being 65.17% more productive than MTPE when performed along with the aforementioned activities. It is worth noting that this study used MT with general-domain data instead of specific-domain data,

---

[1] This quote was translated from Portuguese to English by the author (Oliveira (2012, p. 202)).

which might have affected the results. Bearing this in mind, the next section describes and discusses the MT errors and post-editing alterations.

In order to facilitate the visualization of the translation activity using MTPE in Smartcat, the information was organized in tables for each document. In the first column, "S" stands for Segment with the enumeration defined by SmartCat. Moreover, sensitive information was censored by being replaced with "[censored]". At last, given the standardization of the documents, there were identical errors, which were intentionally omitted from the tables to avoid redundancy. For example, the terms "CPF" and "Válida em todo o território nacional". Moreover, minor grammar mistakes such as missing prepositions or punctuation were not included in the tables due to space constraints of this article's publication.

Table 5 shows the relevant mistakes and alterations on the MTPE of "ID Card 3".

Table 5 - MTPE alterations in "ID Card 3"

S	Source Text	MT	MTPE
2	ESTADO DE PERNAMBUCO / SECRETARIA DE DEFESA SOCIAL / INSTITUTO DE IDENTIFICAÇÃO TAVARES BURIL	STATE OF PERNAMBUCO / SECRETARY OF SOCIAL DEFENSE	STATE OF PERNAMBUCO / SECRETARIAT OF SOCIAL DEFENSE / TAVARES BURIL IDENTIFICATION INSTITUTE
5	VÁLIDA DEM TODO O TERRITÓRIO NACIONAL	VALID THROUGHOUT THE NATIONAL TERRITORY	VALID NATIONWIDE
9	EXPEDIÇÃO	EXPEDITION	ISSUE
12	FILIAÇÃO	AFFILIATION	FILIATION
15	NACIONALIDADE	NATURALNESS	BIRTHPLACE
20	DOC.ORIGEM CN 96706 L131A F245V CART 7AZONA	DOC.ORIGIN CN [censored] L[censored] F[censored] CART 7	ORIGIN DOC.ORIGEM BIRTH CERTIFICATE [censored] BOOK [censored] PAGE [censored] 7TH AREA NOTARY OFFICE
22	CPF	CPF	CPF (Taxpayer's ID)

In segment 2, "Identification Institute" along with its name was omitted for an unknown reason, so it had to be manually added. In addition, even though the MT of segment 5 makes sense, "Valid Nationwide" is the common sentence in this context. Moreover, segments 9, 12, and 15 contained errors in meaning. Finally, segments 20 and 22 both contained acronyms such as CN for "Certidão de Nascimento" that were corrected in PE or with an added comment: "CPF (Taxpayer's ID)".

Moreover, Table 6 shows the relevant mistakes and alterations on the MTPE of "ID Card 4".

Table 6 - MTPE alterations in "ID Card 4"

S	Source Text	MT	Post-Editing
1	REPÚBLICA FEDERATIVA DO BRASIL	Republic FEDERATIVE OF BRAZIL	FEDERATIVE REPUBLIC OF BRAZIL
3	SECRETARIA DE ESTADO DA CASA CIVIL DETRAN - DIRETORIA DE IDENTIFICAÇÃO CIVIL	Secretariat THE STATE OF CASA CIVIL DETRAN-DIRECTORATE OF CIVIL IDENTIFICATION	STATE SECRETARIAT OF CASA CIVIL DETRAN - CIVIL IDENTIFICATION BOARD
4	Assinatura do Titular	Signature the Incumbent	Holder's Signature
27	2 Via	2 way	2nd copy
28	PIS	Pee	PIS
29	LEI NO 7 116 DE 29/08/83	LAW NO 7 116 OF 29/08/83	LAW NO. 7 116, AS OF 08/29/83

It is important to highlight the fact that some mistakes happened because of tags. For example, the errors in segments 1, and 3 would have not happened if the formatting had been correct. Moreover, in segment 4, "incumbent" means an officer or holder of an office, not the holder of a document. Segment 27 also exhibits an error of meaning, while segment 28 seems unusual. Finally, segment 29 showcases the substitution of the preposition "of" by the phrasal preposition "as of", which changes the meaning of stating the starting date of a Law that is still in use.

Table 7 shows the relevant mistakes and alterations on the MTPE of "Birth Certificate 3".

Table 7 - MTPE alterations in "Birth Certificate 3"

S	Source Text	MT	Post-Editing
2	Distrito Encruzilhada	Crossroads District	Encruzilhada District
6	ROMERO LONGMAN, Oficial do Registro de Nascimento e Óbitos da Sétima Zona Judiciária, em virtude da Lei etc.	ROMERO LONGMAN, Birth And Death Registry Officer of the Seventh Judicial zone, under the etc Act.	ROMERO LONGMAN, Birth And Death Registry Officer of the Seventh Judicial Zone, under the Law, etc
7	Registro No. [censored] Livro no.	Record No. [censored] book no.	Registry No. [censored] Book no.

Table 7 - MTPE alterations in "Birth Certificate 3" (Continuation)

S	Source Text	MT	Post-Editing
9	CERTIFICO que no livro de assentamentos de nascimentos, foi feito no dia [censored] do mês de [censored] do ano . [censored]	I certify that in the book of settlements of births, it was made on the [censored] day of [censored] of the year . [censored]	I HEREBY certify that in the book of Birth Registry, it was made on the [censored] day of [censored] of [censored]
11	Do sexo Feminino, Nascido(a) no dia [censored]	Male, born [censored]	female, born on [censored]
22	Declarante O genitor em .[censored]	Declarant The parent in .[censored]	Declarant: the father on [censored]
25	O referido é verdade e dou fé.	This is true and I believe it.	I certify the above to be true and correct.

In segment 2, "Encruzilhada" was translated, but names of places should only be adapted when the term is well established, i.e. "Brazil" for "Brasil", or "Nova Iorque" for "New York". Although maybe the machine interpreted "Encruzilhada" as the noun that means a place where two or more roads meet, instead of the name of a district.

Furthermore, the sentence "em virtude da Lei etc", in segment 6, had a missing colon after "Lei" in the original file that failed to be corrected by the author in AbbyFineReader (after the OCR recognition). In consequence, MT translated it as "the etc Act", instead of "Law, etc". A similar situation happened in segment 22 but with an extra dot. This caused the MT to use "in" before the date, instead of the preposition "on.". However, this segment also needed to be altered because of the gender neutrality of the word "parent".

Moreover, "registro" was translated as "registry" in segment 2, but it was not a choice for the following segments. I.e., the chosen word was "record" in segment 7, and "settlement" in segment 9. So, in order to have consistency throughout the text, both segments were altered with "registry" being the chosen translation.

Additionally, both segments 11 and 25 had mistakes relating to meaning. For an unknown reason, "do sexo feminino" was translated to "male", perhaps because "sexo" is considered to be a masculine word in the Portuguese grammatical gender. In segment 25, "O referido é verdade" translated as "This is true" is not clear enough, since "this" is not defined, meanwhile "The above" emphasizes that the entire content of the document is true. Moreover, "Dou fé" means "To attest or to certify", not to "believe it" as proposed by MT. Those types of mistakes could be easily avoidable when using a TM.

It is important to acknowledge that "O referido é verdade e dou fé" is a standard sentence for documents and/or public deeds, and is defined as "a culture-bound item" by Nascimento (2006), i.e. "Source culture-specific terms or phraseologisms that cannot be

understood by someone who does not have a reasonable grasp of the source culture". In this case, the required knowledge would be related to documents. Consequently, there are discussions and translation solutions in articles, websites, and glossaries written by translators. For instance, "I certify the above to be true and correct" is one option found in the term search of ProZ.com.

Additionally, Table 8 shows the relevant mistakes and alterations on the MTPE of "Birth Certificate 4".

Table 8 - MTPE alterations in "Birth Certificate 4"

S	Source Text	MT	Post-Editing
6	Dou fé.	I give faith.	I attest.
8	Eliane Cristina Leski Matoso Substituta	Leski Matoso Substitute	Eliane Cristina Leski Matoso Deputy
12	Rua Comendador Franco, 18 • Centro • CEP 83.880-000 • Rio Negro • Paraná • Fone: (47) 3642-5015	Rua Comendador Franco, 18 * Centro * CEP 83.880-000 • Rio Negro * Paraná * phone: (47) 3642-5015	Rua Comendador Franco, 18 / Centro / ZIP Code 83.880-000 / Rio Negro / Paraná / Phone: (47) 3642-5015

Unlike the segment 25 of the previous table, "dou fé" was translated as "I give faith" in segment 6. However, this expression does not make sense in the legal or formal use of the English language. Moreover, "substitute" is not the appropriate word for "Substituta" in segment 8. Even though both "substitute" and "deputy" are used to refer to someone who is taking the role of someone else who is absent, e.g. a substitute teacher, "deputy" is used in official contexts, in this case, a Notary Public. Finally, segment 12 was formatted by replacing the special characters for "/" to be more organized and changing "CEP" for "ZIP Code".

Overall, the main challenge of using MT for Sworn Translation might be the translation of culture-bound items as well as acronyms (i.e. CN for Certidão de Nascimento). Both of these terms demanded more research and post-editing. However, implementing a TM might prevent or diminish the necessity of searching for their meanings.

Moreover, the OCR Recognition and formatting can add a relevant amount of time so, in some cases, HT may be more productive than MTPE. For instance, the time for HT in "ID Card 1" was four minutes and three seconds, meanwhile the total time for MTPE, OCR Recognition, and formatting for "ID Card 3" was 9 minutes and 36 seconds, hence each one of these three activities had their own challenges already mentioned (resolution; culture-bound items; vertical formatting). So, it may not be productive to spend time in three different types of software (Word; Abby; Smartcat) for a document with as few characters as

an ID Card. However, it is important to highlight that the result might be different in 4-page Engineering Undergraduate Transcripts, for example. Especially when considering the similar titles shared by some courses like "Calculus I", "Calculus II", etc.
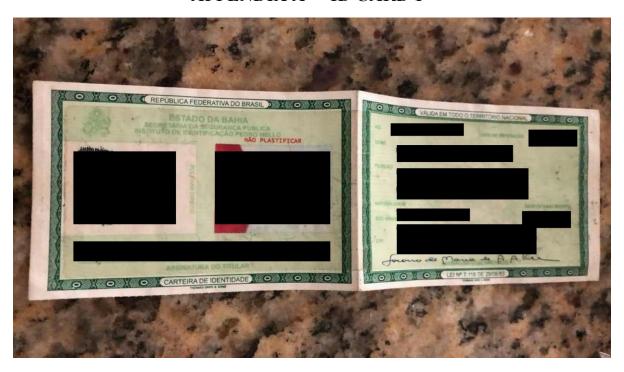
5. FINAL REMARKS

This article aimed to analyze the productivity of using MTPE in the Sworn Translation of personal documents. In order to do so, two types of documents were selected and the time spent in HT, MTPE, OCR Recognition, and formatting were registered and compared. Our results indicate that HT may be more productive than MTPE for the Sworn Translation ID cards and birth certificates. However, further research is needed because the sample size was limited to eight documents. Thus, not all variants or situations could be represented and analyzed, and consequently, the results might not be fully indicative of the Sworn Translation of these specific documents.

Furthermore, the commentaries and translations were performed by the author of this article, who is an assistant of a sworn translator in a Brazilian translation company. This activity is formalized as an internship, under Brazilian legislation, and has a limited time period of two years. So, all of the translated work by an assistant is revised, improved, and signed by a certified professional. Consequently, this perspective does not embrace the entire work of a sworn translator.

So, for further research, some potential aspects that should be addressed are the use of MT combined with TM in Sworn Translation, the analysis of productivity with experienced professionals, and the translation of other types of documents, especially those that contain longer and continuous texts like Income Tax Statements and Transcripts.

Finally, this type of article may potentially contribute to the training of translators, considering its descriptive nature. The open-mindedness and critical thinking relating to the use of technology in translation are crucial to a complete education for translation students or even novice translators, given the relevance in the current market, impact on productivity, and inevitability of these tools being improved as the years go by. Even if the results of using MTPE seem negative or suggest an essential need for human translation, acquiring technological skills allows a researcher or a translator to choose what is best for each situation.

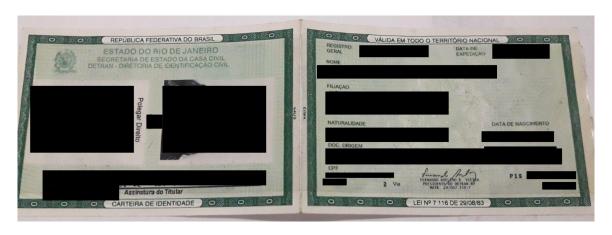# APPENDIX A - "ID CARD 1"
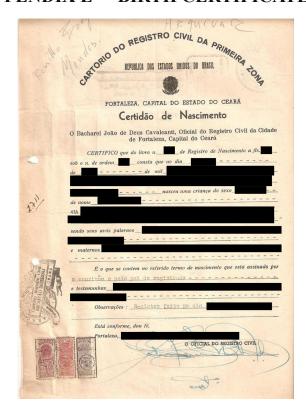

APPENDIX B - "ID CARD 2"

APPENDIX C - "ID CARD 3"

REPÚBLICA FEDERATIVA DO BRASIL

ESTADO DE PERNAMBUCO
SECRETARIA DE DEFESA SOCIAL
INSTITUTO DE IDENTIFICAÇÃO TAVARES BURIL

EC-3

POLEGAR DIREITO

THOMAS GREG & SONS

ASSINATURA DO TITULAR

CARTEIRA DE IDENTIDADE

VÁLIDA EM TODO O TERRITÓRIO NACIONAL

REGISTRO GERAL

DATA DE EXPEDIÇÃO

NOME

FILIAÇÃO

NATURALIDADE

DATA DE NASCIMENTO

DOC. ORIGEM

CPF

THOMAS GREG & SONS

ASSINATURA DO DIRETOR

LEI Nº 7.116 DE 29/08/83

E-57-41.461 - 3023

APPENDIX D - "ID CARD 4"

REPÚBLICA FEDERATIVA DO BRASIL

ESTADO DO RIO DE JANEIRO
SECRETARIA DE ESTADO DA CASA CIVIL
DETRAN - DIRETORIA DE IDENTIFICAÇÃO CIVIL

Polegar Direito

Assinatura do Titular

CARTEIRA DE IDENTIDADE

VÁLIDA EM TODO O TERRITÓRIO NACIONAL

REGISTRO GERAL

DATA DE EXPEDIÇÃO

NOME

FILIAÇÃO

NATURALIDADE

DATA DE NASCIMENTO

DOC. ORIGEM

CPF

FERNANDO AVELINO B. VIEIRA
PRESIDENTE DO DETRAN-RJ
MATR. 24/007.550-7

2 Via

PIS

LEI Nº 7.116 DE 29/08/83

APPENDIX E - "BIRTH CERTIFICATE 1"

APPENDIX F - "BIRTH CERTIFICATE 2"

APPENDIX G - "BIRTH CERTIFICATE 3"

APPENDIX H - "BIRTH CERTIFICATE 4"

REFERENCES

BARROS, L. A.; BABINI, M.; AUBERT, F. H. **Terminologia e tradução juramentada**: questões de tipologia textual e equivalência terminológica interlinguística Português-francês-italiano. Filologia e Linguística Portuguesa, *[S. l.]*, v. 12, n. 2, p. 233-249, 2010. DOI: 10.11606/issn.2176-9419.v12i2p233-249. Disponível em: https://www.revistas.usp.br/flp/article/view/59867. Acesso em: 2 jul. 2023.

GOEDERT, Alessandra Otero. **Tradução Juramentada e as Modalidades de Tradução:** o caso dos históricos escolares. 2015. (219)f. Dissertação (Mestrado em Estudos da Tradução) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015. p. 37-41.

GOTTI, Maurizio; ŠARČEVIĆ, Susan. **Insights Into Specialized Translation**. 2nd ed. Illustrated, p. 9-10. Peter Lang, 2006.

GUEDES, Rosane. **Tradução especializada: um mergulho nas especificidades do gênero jurídico**. Tradução em Revista, [S.l.], 2016. DOI: 10.17771/PUCRio.TradRev.28028.

GUTIÉRREZ ARCONES, David. **Estudio sobre el texto jurídico y su traducción: características de la traducción jurídica, jurada y judicial.** Miscelánea Comillas: Revista de Ciencias Humanas y Sociales, vol. 73, nº 142, 2015, p. 141-175. ISSN 0210-9522. ISSN-e 2341-085X.

HOLMES, James S. **The Name and Nature of Translation Studies.** In: VENUTI, Lawrence (Ed.). The Translation Studies Reader. New York: Routledge, p. 172-185, 2000.

HUTCHINS, W. John; SOMERS, Harold L. **An Introduction to Machine Translation**. London: Academic Press, 1992. p. 2-5.

KRINGS, H. (2001). **Repairing Texts: Empirical Investigations of Machine Translation Post- Editing Processes**. Kent, OH: The Kent State University Press.

MAYORAL, Roberto. **Sworn Translation and Its Functions**. Fédération Internationale des Traducteurs (FIT) Revue, Babel, 46(4), p. 300-331. 2000. Disponível em: https://doi.org/10.1075/babel.46.4.03may. Acesso em 20 de abr. 2023.

MESA-LAO, B. **Introduction to post-editing - The CasMaCat GUI**. In: Proceedings of the Center for Research and Innovation in Translation and Translation Technology Conference: SEECAT Project. Copenhagen Business School, Denmark, 22 May 2013.

NASCIMENTO, Lúcia de Almeida e Silva. **Investigating Norms in the Brazilian Official Translation of Semiotic Items, Culture-Bound Items, and Translator's Paratextual Interventions**. 2006. Tese (Doutorado em Letras/Inglês, Linha de Pesquisa Tradução) – Universidade Federal de Santa Catarina, Florianópolis, 2006. p. 17.

O'BRIEN, Sharon. **Introduction to Post-Editing**: Who, What, How and Where to Next?. 2009. Slides. Disponível em: https://aclanthology.org/2010.amta-tutorials.1.pdf Acesso em: 03 jul. 2023.

OLIVEIRA, Bruna Macedo de. **Tecnologias da tradução no trabalho de tradutores jurídicos/juramentados:** estudo de caso. Abehache, v. 2, n. 3, 2012. p. 191-205. Disponível em: https://revistaabehache.com/ojs/index.php/abehache/article/view/84/83. Acesso em: 30 mar. 2023.

RAMOS, Fernando Pietro. **Legal Translation Studies as Interdiscipline:** Scope and Evolution. Meta, v. 59, n. 2, 2014. p. 260-277. Disponível em: https://doi.org/10.7202/1027475ar. Acesso em: 30 mar. 2023.

SAWANT, Dr. Datta. **History of Translation**. Literary Endeavor, [S.l.], v. 4, p. 109-115, 2013.

SMARTCAT. **Why Smartcat is safe to use**. Disponível em: https://www.smartcat.com/security/. Acesso em: 24 abr. 2023.