ALBUQUERQUE, Aíla Fonseca de; KOGLIN, Arlene. Uso da MTPE no contexto da tradução juramentada: ônus ou bônus?. *Revista Belas Infiéis*, Brasília, v. 13, n. 1, p. 01-25, 2024. e-ISSN: 2316-6614. DOI: 10.26512/belasinfieis.v13.n1.2024.54644
Recebido: 05/07/2024 Aceito: 05/10/2024 Publicado: 12/11/2024

USO DA MTPE NO CONTEXTO DA TRADUÇÃO JURAMENTADA: ÔNUS OU BÔNUS?

APPLYING MTPE TO SWORN TRANSLATION: BONUS OR BURDEN?

Aíla Fonseca de ALBUQUERQUE
Graduanda
Universidade Federal de Pernambuco
Departamento de Letras
Letras – Inglês (Licenciatura)
Recife, Pernambuco, Brasil
http://lattes.cnpq.br/4751889150303620
https://orcid.org/0009-0004-4846-9382
aila.fonseca@ufpe.br

Arlene KOGLIN
Professora adjunta
Universidade Federal de Pernambuco
Centro de Artes e Comunicação
Letras – Inglês (Licenciatura)
Recife, Pernambuco, Brasil
http://lattes.cnpq.br/3897922592505741
https://orcid.org/0000-0003-0711-0811
arlene.koglin@ufpe.br

Resumo: Trata-se de um estudo que teve como objetivo investigar a produtividade da pós-edição de tradução automática (MTPE) aplicada à tradução juramentada de documentos pessoais. Mais especificamente, objetivou-se descrever os desafios enfrentados na pós-edição de certidões de nascimento e de carteiras de identidade no par linguístico português brasileiro/inglês norte-americano, bem como analisar o tempo gasto na MTPE desses documentos em comparação à tradução humana (TH). A metodologia consistiu de TH e de MTPE cronometrada e anotada. Também incluiu a descrição dos elementos e a formatação final, que é característica da tradução juramentada. Os arquivos PDF foram convertidos para Word usando um software de reconhecimento ótico de caracteres (OCR). Os resultados indicaram que a MTPE pode ser mais produtiva que a TH quando não são considerados o tempo gasto no OCR e na formatação. No entanto, como o tamanho da amostra foi limitado a oito documentos e apenas dois tipos de documentos, os resultados não são passíveis de generalização.
Palavras-chave: Tradução juramentada. Pós-edição de tradução automática. Documentos pessoais. Tradução Humana. Esforço temporal.

***Abstract**: This article reports on a study that aimed to investigate the productivity of machine translation post-editing (MTPE) applied to the sworn translation of personal documents. More specifically, it aimed at describing some of the challenges of post-editing birth certificates and ID cards from Brazilian Portuguese into American English, as well as analyzing the time spent on the MTPE of such documents in comparison to manually translating them. The methodology consisted of timed and annotated translation using MTPE and human translation (HT). It also included the description of elements and final formatting that is characteristic of sworn translation. The PDF files were converted to Word using an OCR software. Our results indicated that MTPE might be more productive than HT when time spent on OCR and formatting is not considered. However, since the sample size was limited to eight documents and two types of documents, the results cannot be generalized.*
***Keywords:** Sworn translation. Machine translation post-editing. Personal documents. Human translation. Temporal effort.*

The technology, for better or for worse, is here to stay. (PYM, 2011, p. 5)

From pen and paper to the use of computer-assisted translation (CAT) tools and machine translation (MT), it is evident that technology has revolutionized the field of translation, including one of its many subfields: sworn translation[i]. A recent study by Omazić & Šoštarić (2023) has reported on an increased quality of machine translation output when it comes to sworn translation. But how far has the technology progressed to the point where it can be used advantageously in this field? Is this development sufficient to make machine translation post-editing (MTPE) more productive than human translation (HT)? To answer these questions, it is first necessary to define what sworn translation is and how it is currently practiced in Brazil.

According to the Gov.br website, which provides information about the Brazilian Federal Executive Branch, sworn translation is a type of translation done by a public translator. Mayoral (2000) also defines it as the official translation of documents. In contrast to translators in other fields, a sworn translator has the authority to provide a sworn statement certifying the accuracy and authenticity of the translated documents.

In Brazil, Getúlio Vargas established sworn translation in 1943 through Decree No. 13.609, Article 18, which declared that foreign books, papers, or documents must be accompanied by a version in Brazilian Portuguese in order to have any national validity. This version should be made by translators appointed by the decree. Subsequent presidents have maintained and modified the legislation. To this day, sworn translation continues to bridge Brazil and the world. It should be noted, however, that each country has its own unique approach to this practice, which is governed by specific national laws and regulations.

Despite the prevalence of sworn translation in the market, as indicated by data from National Register of Certified Public Translators and Commercial Interpreters (CNTPIC)[ii], which claims that in 2020 there were over 300 sworn translators hired in Brazil, this field of study still lacks attention. While the market seems to be flourishing, especially as public exams are opened in response to growing demand, this trend is not reflected in academic research in Brazil. The amount of work that is produced and/or published continues to be low (Goederte, 2015; Nascimento, 2006).

As a result, the small number of studies may limit the common knowledge about certain aspects of such an important activity, especially when adding Machine Translation Post-editing (MTPE). For example, the experience of using MTPE may be drastically different for an ID

card, which is characterized by less content, than for a birth certificate, which usually contains numerous stamps, seals, etc., given the specific characteristics and elements of a document (coat of arms, stamps, seals, etc.).

According to Holmes (1972), for a solid knowledge base to exist in a field of translation, theory and practice must have a dialectical relationship in which one provides knowledge for the other. However, this is not the case for sworn translations, which explains the discrepancy between the large number of texts translated and the scarcity of published studies. Therefore, this article and others to come are fundamental to continuing to develop expertise and educational materials for future professionals.

With this in mind, the objective of this study was to investigate the productivity of MTPE applied to the sworn translation of personal documents. More specifically, it aimed a) to describe some of the challenges of post-editing birth certificates and ID cards from Brazilian Portuguese into American English, and b) to analyze the time spent (i.e., temporal effort) on the MTPE of such documents in comparison to manually translating them.

To that end, this article is structured as follows: in the Theoretical Framework section, the key concepts related to sworn translation and MTPE are discussed. The next section describes the methodological aspects (i.e., materials, data collection procedures, and data analysis) of this study. This is followed by a qualitative and quantitative analysis of the data. Finally, the concluding remarks are provided.

Theoretical Framework

Sworn Translation

Translation dates back to 3000 B.C. with the Rosetta Stone. It has also had an immeasurable impact on history. This is evidenced by the translation of the Odyssey and the Bible (Sawant, 2013). It is therefore not surprising that such an ancient activity has evolved into a field with various branches of translation, such as literary translation, multimedia translation, localization, audiovisual translation, and specialized translation.

Specialized translation can be defined as a non-literary type of translation that requires specific knowledge of other fields, such as "science and technology, economics, marketing, law, politics, medicine, and mass media [...]" (Gotti & Šarčević, 2006, p. 9). Additionally, this vast branch includes the translation of legal texts, which itself has its own specificities and typologies (Guedes, 2016).

Legal Translation Studies is concerned with "all aspects of translation of legal texts, including processes, products, and agents." (Ramos, 2014, p. 260). One of these activities is called sworn translation, which refers to the "official, written translation of documents" (Mayoral, 2000, p. 1). It is used to make foreign documents valid and/or authenticated in a given country. However, sworn translation is not limited to legal texts, but to any type of text that would require the authentication of a sworn translator. (Gutíerrez, 2015). For example, social media messages in a foreign language, which may be used as evidence in a court case for the purpose of proving a crime but are not legal texts, will require a sworn translation.

In Brazil, it is not easy to characterize a sworn translation because there are no strict rules regarding the formatting and description of the elements that appear in the documents. In her dissertation, Goedert (2015) collects official normative data for the translation of documents from the *Junta Comercial de Minas Gerais*, but these rules are not always followed by all public translators. Each state of Brazil has its own group of public translators, in which they share their own decisions (especially when it comes to companies and trainees/assistants).

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

Bearing that in mind, using MTPE in sworn translation can be challenging. For instance, if a document is delivered in a format other 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. This is especially true when a translator is not using a translation memory, as was the case in this study.

Finally, before addressing the reported practice applied to sworn translation, it is essential to define and characterize MTPE.

Machine Translation Post-editing (MTPE)

Hutchins and Somers (2002) define machine translation (MT) as computer systems that translate natural languages and may or may not be human-assisted. The power of such systems can be enhanced by translation memories (TM), which are a "type of linguistic database that is used to store source texts and their translations" (Bowker, 2002, p. 92).

The translator's work is improved both logistically and linguistically by using MT in combination with a robust TM. First of all, the time spent on translation is reduced because the main idea behind TM is to reuse previous translations. Additionally, if the memory is up to date with the latest terminology, errors, and extra research can be avoided. Thus, greater productivity and consistency can be achieved by using such a database.

Despite all the advances in machine translation, human intervention is still required to ensure the accuracy and adequacy of the translated text as well as to achieve quality that is similar or equal to human translation, often referred to as full post-editing. Contrary to light post-editing, which is used for texts that are urgently needed and will have only an internal and perishable use, full post-editing requires more corrections, resulting in higher quality. In this case, "the terminology must be consistent and comply with domain-specific requirements. Syntax, spelling, punctuation and other orthographic characteristics as well as formatting must be correct" (Nitzke & Hansen-Schirra, 2021, p. 35).

In the field of sworn translation, there is a need for human intervention in the form of full post-editing, as shown by the recent research of Omazić and Šoštarić (2023, p. 81). Although they used the latest versions of MT in their study on the post-editing of legal texts, the human touch was still necessary in order to achieve the highest level of translation quality.

Mesa-Lao (2013, p. 4) defines post-editing (PE) as "reviewing a pre-translated text generated by an MT engine against an original source text, correcting possible errors, in order to comply with a set quality criterion in as few edits as possible". Although Mesa-Lao uses "reviewing" in his definition, it is worth clarifying that PE is different from revision. "While technology appears central to the practice of post-editing, an interesting feature of revision is that there are very few machine aids specific to this aspect of translation work" (Koponen et al., 2021, p. 3).

There are general and specific guidelines that are designed not only to help out the post-editor produce high-quality target texts but also to save time (Nitzke & Hansen-Schirra, 2021). In terms of final quality and editing needs, there are two well-established distinct categories, namely: light and full PE. However, this distinction is rather superficial for real PE projects as

discussed by Nitzke and Hansen-Schirra. “The guidelines might differ a lot from project to project and every PE project might focus on different quality aspects, e.g., using the correct, pre-defined terminology might be far more important in technical documentation than in post-editing a newspaper article.” (Nitzke & Hansen-Schirra, 2021, p. 30)

Following this line of thought, Translation Automation User Society (TAUS), an independent organization interested in MT, also published their recommendations for MTPE. They were categorized into two levels: a lighter one, involving “good enough” PE, and a more comprehensive or full PE. Considering the legal nature of sworn translation, full PE is deemed the most appropriate, particularly for identification documents that need editing performed with the utmost attention and diligence.

In view of that, what is the reported reality of using MTPE in the sworn translation of documents? In order to answer this question, a targeted search for published work on this practice was conducted. Although it did not yield many results, two studies proved to be particularly valuable, as discussed in the following subsection.

6 *The Practice of Sworn Translation Using MTPE*

Oliveira (2012) reports on the perceptions of an experienced sworn translator regarding the integration of translation technologies in the sworn translation process. The interviewee was selected in the 1980 JUCESP public recruitment examination and had previously worked with technical texts in the Spanish/Brazilian Portuguese language pair, finding MT to be helpful with this type of text. However, as she began to focus her attention on official legal texts, her perspective changed. She did not find it particularly advantageous to use MT for this type of text because it requires weighted decisions, i.e., a specific decision-making process for each situation and/or document. On the other hand, she found it more efficient to create a personal database of previously translated documents in Word and use them as models for future translations.

Nevertheless, that perception should be viewed with caution. For one thing, the interview was conducted by Oliveira in 2012, and the quality of MT output has drastically improved since then. In addition to that, personal documents have a fairly regular pattern of content and vocabulary within a given country. For example, two birth certificates from the same year, but one issued in Pernambuco and the other in Brasília, will have many matches between them. Thus, using MT alone or combined with a TM would avoid the need to translate sentences like “*O referido é verdade e dou fé*” from scratch.

Contrary to this, Oliveira (2012) believes that the pattern found in documents is not enough to establish the fruitfulness of the use of MT and TM in sworn translation. The author explains that the level of individuality and diligence prevents the use of such tools. However, this statement can be questioned, especially when full post-editing is considered. Not all documents will have a large number of exact matches, but this does not mean that those which were processed by MT and/or TM could not save time, especially in the case of long documents with continuous texts, such as income statements for taxes, or common terms in school transcripts, such as "Mathematics" or "Portuguese Language".

Another obstacle to the use of MT, according to Oliveira (2012), is the extreme secrecy that most documents require. In a more recent study on using MT in legal text translation, Omazić and Šoštarić (2023) also address the issue of confidentiality. The authors state that ethical and practical considerations—including data handling policies—as well as confidentiality requirements are factors that need to be considered when integrating MT into translation workflows. Although those issues are essential, they are not considered an obstacle by the authors. They only recommend that "careful attention should be paid to the data-handling policies of the machine translation provider" (Omazić & Šoštarić, 2023, p. 81).

With respect to the confidentiality issue, it should also be noted that CAT tools that have been developed for use in the translation industry already take this into account. According to the website of Smartcat, the cloud-based translation and localization CAT tool we used to conduct our research, all information is stored in the cloud. This protects the translator in case of viruses or theft of their device. Besides that, users cannot access data from other accounts, which is ensured by "three major authentication systems: ADFS, Azure AD, and Okta."

With this information in mind, it is valid to investigate the productivity and impacts of MT in sworn translation today, given the considerable improvements in MT quality as a result of the rise of NMT and AI. These advances include deep learning models that better capture contextual meaning to improve translation fluency and accuracy, and massive data sets that enable more nuanced and culturally aware translations. The next section will then describe the methodological aspects of this study.

Methodology

Data Collection

Data were collected on the Internet using Google and Bing search engines, and they consisted of publicly available personal documents. However, given the challenge of finding examples with appropriate resolutions for optical character recognition (OCR), personal documents belonging to the authors of this paper have also been part of the research.

Along with the concern for image quality, the selection criterion was the choice of two types of personal documents, which are most in demand for sworn translation, i.e., birth certificates and ID cards. The collected documents were organized into two main folders: "Human Translation" and "MTPE", each of which contained two subfolders labeled "ID Cards" and "Birth Certificates".

Material

As proposed by Barros, Babini and Aubert (2010), documents can be divided into five categories: personal documents (which include birth certificates and ID cards); corporate documents; financial-commercial documents; legal documents; and other types of documents.

According to the Gov.br website, an ID card is a Brazilian citizen's identity document. Its purpose is to ensure individuality and identification in various social actions. This official website further defines a birth certificate as the first document with legal validity that gives a person a name, surname, nationality, filiation, and civil and social rights. In other words, it is proof of the existence of a citizen.

To ensure impartiality regarding translation time and speed, four examples of these two documents were used. Since this research was not equipped with a well-fed TM, and therefore the number of matches during the MTPE would be minimal to almost non-existent, the analysis was limited to the results of machine translation. The MT chosen was Yandex, which is Smartcat's free and generic machine translation system, given the authors' experience with this tool.

Procedures

This study included both human translation (HT) and MTPE. The time spent was recorded by a timer and annotated in a note-taking application, all of which are basic Windows software programs. Both the HT and the MTPE were carried out by one of the authors of this article, who has two years of professional experience using the MTPE for sworn translations.

The documents entitled “ID card 1”, “Birth certificate 1, “ID card 2”, and “Birth certificate 2” were manually translated (i.e., HT), whereas “ID card 3”, “Birth certificate 3”, “ID card 4”, and “Birth certificate 4” were subjected to MTPE. Sensitive information protected by law, such as names, dates of birth, places of birth, etc., was not censored until after the translation activities had been completed, as this information is relevant to the quantitative analysis.

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

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

The files were then uploaded to Smartcat with Brazilian Portuguese as the source language and USA English as the target language. Next, the “Use machine translation” option was checked, along with “Use for free with feedback”. No translation memory was added, as the results would not be productive with a small TM. Finally, the documents were machine-translated and post-edited following full post-editing guidelines (cf. O’Brien, 2010).

After each project was completed, the resulting file was downloaded. A new Word document was then created with the following specifications: single spacing, Times New Roman font, size 11, and justified text alignment. The content of the finished file was then pasted into this new document without any formatting (unformatted Unicode text). The goal of this process was to convert the content to plain text, eliminating any extra spaces, bold/italic fonts, or other styles that might distinguish one sentence or word from another. For example, the OCR on “ID card 3” displayed “FEDERATIVE REPUBLIC OF BRAZIL” in an underlined 16-point Arial font in black, while “STATE OF PERNAMBUCO” was displayed in a 21-point Times New Roman font in green. Moreover, the text was spread over two pages with no discernible structure or organization.

Data Analysis

The data were analyzed both quantitatively and qualitatively. To evaluate the productivity of using MTPE, the temporal effort of human translation and MTPE was

compared. Temporal effort, as defined by Krings (2001), is the time spent on post-editing an MT output.

The time spent on OCR and final formatting was also recorded. In addition, MT errors and PE changes were qualitatively analyzed and separated into tables per document.

Analysis and Discussion

To evaluate the productivity of using MTPE in sworn translation, this section presents the analysis of MT output, and the comparison of time spent on HT and MTPE as shown in *Table 1*.

Table 1 — *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

Note. Created by the authors.

As shown in *Table 1*, the total time for MTPE is shorter than for HT. This is particularly evident for 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 took only 6 minutes and 23 seconds for “Birth certificate 3” and “Birth certificate 4”. The time difference between MTPE and HT for ID cards is smaller than for birth certificates, but still relevant: in general, MTPE took 2 minutes and 59 seconds less than HT. Overall, HT took 27 minutes and 41 seconds to translate all documents, and MTPE took 11 minutes and 26 seconds, making HT more time-consuming (57.79%) than MTPE.

This result might suggest that MTPE is more productive than HT and is consistent with the findings of Killman and Castro (2022), who reported on an experiment involving 26 participants that post-edited and translated legal texts. Their results showed that post-editing saves time and significantly improves quality.

However, it is worth mentioning that the OCR can add up to the time of the MTPE. This data can be seen in *Table 2*.

Table 2 — *Time spent on OCR*

MTPE	Duration (min:sec)
ID card 3	04:34
ID card 4	02:58
Birth certificate 3	08:30
Birth certificate 4	02:36

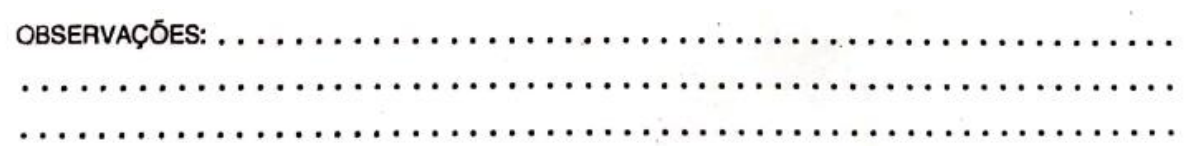
Note. Created by the authors.

Adding the OCR time values 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. Nevertheless, MTPE and OCR still saved 37 seconds in comparison to HT. Although temporal effort was not considerably lower for MTPE when OCR was included, our results are consistent with other experimental studies comparing temporal effort for MTPE and HT (Krings, 2001; O'Brien, 2006a, 2007; Carl et al., 2011; Green, Heer & Manning, 2013, Koglin, 2015).

Several challenges with the source documents and their OCR were responsible for this considerable increase in MTPE time with OCR. First, documents contain security features such as security threads, watermarks, and microprinting to avoid falsification. In the "ID card 3" scan, for example, many of these elements were visible, causing the insertion of false characters in the OCR result file. Moreover, some texts had a vertical configuration, such as "Polegar Direito", which caused poor recognition. Besides that, stamps and small fonts also caused more errors.

In fact, these elements are not the only reason why OCR can be challenging. The Birth Certificate 3 was not issued on official paper, but it was formatted differently. The content was organized using dotted lines (as can be seen in *Figure 1*), which added to the time of OCR and the time of final formatting since the resulting file should preferably be organized as plain text.

Figure 1 — Snippet of Birth Certificate 3. For full information, see Appendix G.



Note. Based on a personal document belonging to one of the authors.

In a company, one possible solution to this challenge would be to separate these two tasks into different teams: one for formatting, and the other for MTPE. Additionally, in the case

of a freelance translator, the professional could ask for digital documents or high-resolution scans only.

Finally, *Table 3* shows the time spent describing and formatting the documents.

Table 3 — *Time spent describing the elements and formatting the resulting file*

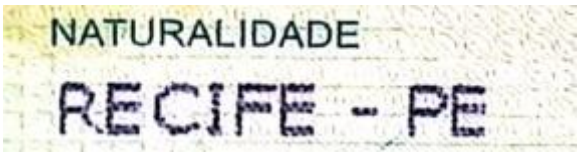
Description/Formatting	Duration (min/sec)
ID card 3	02:52
ID card 4	03:38
Birth certificate 3	03:26
Birth certificate 4	07:35

Note. Created by the authors.

The time spent manually describing elements and formatting were the last activities for using MTPE in this research. Both tasks of describing and formatting took a total of 17 minutes and 32 seconds, compared to 11 minutes and 26 seconds for MTPE across the four documents.

In general, the main difficulty in formatting the resulting SmartCat file was based on the structure of the text found in the documents, considering that the usual formatting for sworn translations is plain and continuous text. For example, the phrases in the ID cards are arranged vertically, e.g., in “ID card 3”, “Birthplace” appears above “Recife - PE”, which is not the preferred order (see *Figure 2*).

Figure 2 — Snippet of ID Card 3. For full information, see Appendix C.



Note. Based on a personal document belonging to one of the authors.

Moreover, “Birth certificate 3” consists of sentences with little to no connectives or punctuation. Similarly, “Birth certificate 4” consists mostly of tables. When converted to plain text, the content did not have a cohesive organization. As a result, reorganizing the text was a time-consuming task.

The total time for describing and formatting the elements, the MTPE, and the OCR was 44 minutes and 36 seconds. As a result, HT was 65.17% more productive than MTPE when performed with the activities listed above. It should be noted, though, that this study used MT

with general domain data rather than specific domain data, which may have affected productivity.

It is also worth mentioning that the post-editing guideline regarding formatting and tags was not followed in our study, since sworn translation has traditionally prioritized textual accuracy over formatting. This is because Brazilian laws in this area date back to 1943, and sworn translation used to be done with pen and paper or a typewriter.

With the advancement of technology, some professionals are choosing whether or not to follow the exact layout of the original document. This flexibility arises from the absence of specific guidelines or regulations by the government regarding formatting standards. In addition, since Brazilian law states that a foreign document is only valid if it is accompanied by a sworn translation, the original document can always be checked to understand its organization and formatting, while the translated text can be used to understand the content.

With this in mind, we will now focus on describing and discussing the MT errors and post-editing changes. To help visualize the translation activity using MTPE in Smartcat, the information for each document has been organized in tables. In the first column, "S" stands for Segment with the enumeration defined by SmartCat. In addition, sensitive information was censored by replacing it with "[censored]". Finally, due to the standardization of the documents, there were identical errors such as the terms "CPF" and "*Válida em todo o território nacional*" that were intentionally omitted from the tables to avoid redundancy. For the sake of brevity, the tables do not include minor grammatical errors such as missing prepositions and punctuation.

The relevant errors and changes in the MTPE of "ID card 3" are shown in *Table 4*.

Table 4 — *MTPE changes 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 EM 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]	DOC OF ORIGIN: BIRTH CERTIFICATE [censored]

S	Source Text	MT	MTPE
		F[censored] CART 7	BOOK [censored] PAGE [censored] 7TH AREA NOTARY OFFICE
22	CPF	CPF	CPF (Taxpayer's ID)

Note. Created by the authors.

In segment 2, "Identification Institute" was omitted along with its name for some unknown reason, so it had to be added manually. This is consistent with the experience of other translators who used MTPE in legal texts from English to Spanish as discussed in Killman and Rodríguez-Castro (2022, p. 64). In their study, 14 out of 26 translators reported a similar experience, i.e., "translations of words and/or phrases were missing from the MT output." Moreover, this alteration falls under the following TAUS guideline: "Ensure that no information has been accidentally added or omitted".

Regarding segment 5, although the MT output makes perfect sense, "Valid Nationwide" is the commonly used phrase in this context; and it can be found in a variety of English-written documents. Still, "Valid throughout the national territory" could be an alternative translation, especially for less-experienced translators. However, choosing the more commonly used phrase over a strictly literal but understandable translation contradicts O'Brien's guideline (2010) of overlooking stylistic and textual issues. In terms of expected outcomes, her guidelines specify high throughput expectations and medium quality expectations. However, since sworn translation involves personal documents for international distribution or legal purposes, quality expectations should be the highest possible.

The MT output also had semantic errors in segments 9, 12, and 15. While "*Expedição*" relates to issuing a document, "expedition" is a synonym for journey or excursion according to the Cambridge Dictionary. "*Filiação*" is a formal way to refer to "parents" in Brazilian Portuguese; therefore, an English equivalent would be "filiation", rather than "affiliation" (which refers to the relationship between companies or organizations). Furthermore, "*Nacionalidade*" relates to one's country of birth, while the same dictionary defines it as "the quality of being real and not influenced by other people". This type of error could have been avoided by integrating a robust translation memory that not only provides the correct term, but also saves time and ensures terminology consistency.

Finally, both segments 20 and 22 contained acronyms such as CN for "*Certidão de Nascimento*" and CPF. CN was post-edited by adding a note with the corresponding term, while CPF had its meaning explicitly stated in brackets.

Table 5 shows the relevant errors and changes in the MTPE of "ID card 4".

Table 5 — *MTPE changes in "ID card 4"*

S	Source Text	MT	MTPE
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 29/08/83

Note. Created by the authors.

After the initial analysis reported in Albuquerque (2023), a new hypothesis emerged regarding the errors in segments 1 and 3. We hypothesized that some of the errors may have been caused by the tags. To test this, we removed the formatting before running the MT. In other words, all content of the OCR of "ID card 4" was selected and then the "Clear all formatting" tool on Word was selected. The file was then uploaded to Smartcat again, and errors were avoided in both segments.

Despite not using the "Clear all formatting" tool when registering the time for *Table 1*, it did not affect the post-editing process because the resulting file was pasted using the option "unformatted Unicode text," as mentioned in the methodology section. Nevertheless, removing the formatting before uploading the file to Smartcat saved a few seconds in the MTPE by avoiding errors, as shown in *Table 6*.

Table 6 — *Time spent on MTPE versus time spent on MTPE after removing all formatting*

MTPE	Duration (min:sec)	MTPE 2	Duration (min:sec)
ID card 3	02:10	ID card 3	01:44
ID card 4	02:53	ID card 4	01:56
Birth certificate 3	03:47	Birth certificate 3	03:10
Birth certificate 4	02:36	Birth certificate 4	03:58

Note. Created by the authors.

By contrast, the second MTPE task resulted in 10 minutes and 48 seconds, compared to the 11 minutes and 26 seconds spent on the first MTPE. These 38 seconds may seem

insignificant on smaller projects, but the cumulative time saved adds up over the course of an 8-hour shift.

It is evident that the majority of the documents experienced a considerable reduction in time spent on MTPE, with the exception of Birth Certificate 4. This specific document was originally formatted with tables. However, when they were removed, the text became disorganized, requiring additional time to correct. For instance, phrases such as “Gêmeo? Não” and “Nome do(a) Gêmeo(a)” were separated into different cells within the table. After its removal, these phrases merged into a single Smartcat segment, resulting in “Gêmeo Nome do(s) gêmeo(s) Não,” which translated to “Twin name o(fs) twin (s) no”.

On the basis of this result, there are two outcomes regarding the post-editing of personal documents. It can be seen that personal documents issued after the standardization implemented by the Brazilian Notarial System (cf. ANOREG, Association of Notaries and Registrars of Brazil, and CRC, Brazilian Information Center of Civil Registry) show a considerable reduction in time and, therefore, it would be productive to use MTPE. On the other hand, documents issued before this standardization may have a smaller reduction in temporal effort due to the greater time spent on formatting, and, in this case, it is up to the post-editor to assess whether it is productive to use post-editing or not.

Another advantage of using MTPE for more recent personal documents is the possibility of eliminating the OCR stage because it is possible to access digital documents in the cloud after the standardization implemented by the Brazilian Notarial System.

As for segment 4 in *Table 5*, “incumbent” means an officer or holder of an office, not the holder of a document. Segment 27 also has a semantic error, while segment 28 seems unusual. Finally, segment 29 shows the substitution of the preposition “of” for the phrasal preposition “as of”, which changes the meaning of stating the starting date of a law that is still in effect.

Table 7 shows the relevant errors and changes in the MTPE of “Birth certificate 3”.

Table 7 — *MTPE changes in “Birth certificate 3”*

S	Source Text	MT	MTPE
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

S	Source Text	MT	MTPE
7	Registro No. [censored] Livro no.	Record No. [censored] book no.	Registry No. [censored] Book no.
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.

Note. Created by the authors.

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". Furthermore, in segment 6, the sentence "*em virtude da Lei etc*" had a missing colon after "*Lei*" in the original file, which was not corrected by the translator in AbbyFineReader (after the OCR). As a result, it was machine-translated as "the etc Act", instead of "Law, etc". A similar situation occurred in segment 22 but with an extra period. This caused the MT system to use "in" before the date, instead of the preposition "on.". However, this segment also had to be changed because of the gender neutrality of the word "parent".

Moreover, "*registro*" was machine-translated as "registry" in segment 2, but this pattern was not kept for segments 7 and 9, where it was machine-translated as "record" and "settlement", respectively. Although the MT solutions were correct options to translate "*registro*", the term was post-edited as "registry" in order to maintain terminological consistency.

Additionally, both segments 11 and 25 had errors in meaning. For some unknown reason, "*do sexo feminino*" was translated as "male", perhaps because "*sexo*" is considered a masculine word in Brazilian Portuguese. In segment 25, "*O referido é verdade*" translated as "This is true" is not clear enough. Also, "*Dou fé*" means "To attest or to certify", not to "believe it" as the MT output suggests. These types of errors could be easily avoided by using a TM.

It is important to recognize that "*O referido é verdade e dou fé*" is a standard phrase for documents and/or public deeds, and is defined by Nascimento (2006, p. 52) as a "culture-bound item", 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" Nascimento (2006, p.

49). In this case, it is also necessary to have knowledge related to documents. So, there are articles, websites, and glossaries written by translators to discuss and help translate. For example, "I certify the above to be true and correct" is an option found in the ProZ.com term search.

The relevant errors and changes in the MTPE of "Birth Certificate 4" are shown in *Table 8*.

Table 8 — *MTPE changes in "Birth certificate 4"*

S	Source Text	MT	MTPE
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

Note. Created by the authors.

In contrast to segment 25 of the *Table 7*, "*dou fé*" in segment 6 has been machine-translated as "I give faith". However, it does not make any sense in the legal or formal use of the English language. Also, "substitute" is not the appropriate word for "*Substituta*" in segment 8. Although both "substitute" and "deputy" are used to refer to someone who fills the role of someone else who is absent, such as a substitute teacher, "deputy" is used in official contexts, in this case, a notary public. Finally, segment 12 has been formatted by replacing the special characters with"/" for better organization, and the acronym "CEP" has been changed to "ZIP Code".

Overall, the main challenge of using MTPE for sworn translation may be the translation of culture-bound items (Nascimento, 2006) and acronyms (i.e. CN for "*Certidão de Nascimento*"). Both of them require more research and post-editing. However, the implementation of a well-fed TM could eliminate or reduce the need to search for their meanings.

Besides that, OCR and formatting can add a considerable amount of time to the MTPE process. For instance, the time for HT in "ID card 1" was four minutes and three seconds, while the total time for MTPE, OCR, and formatting for "ID card 3" was 9 minutes and 36 seconds. Also, each of these three activities had its own challenges as mentioned earlier (resolution; culture-bound items; vertical formatting). Thus, it may not be productive to spend time in three

different softwares (Word, Abby, Smartcat) for a document with as few characters as an ID card. However, it is important to emphasize that the result may be different for, say, 4-page Engineering Undergraduate Transcripts.

Concluding Remarks

This research aimed to analyze whether it is productive (i.e., requires less temporal effort) or not to use MTPE in the sworn translation of personal documents. To do so, two types of documents were selected and the time spent in HT, MTPE, OCR, and formatting was recorded and compared. Our results indicate that MTPE might be more productive than HT for the sworn translation of ID cards and birth certificates when time spent on OCR and formatting is not considered. However, further research is needed because the sample size was limited to eight documents and consisted of only two types of personal documents. Thus, not all variables or aspects could be represented and analyzed, and consequently, the results are not generalizable.

Another factor that may have influenced our results was the limited professional experience of the translator/post-editor of the documents selected for this study. Lack of familiarity with technological tools or even production techniques can have an impact on productivity, as did the "clearing all formatting" option prior to MTPE.

In addition, it would be strongly recommended to expand the data collection with a larger number of participants in order to have a more in-depth understanding of the MTPE process as it applies to sworn translation. As both MTPE and HT were performed by one of the authors of this article, who was a sworn translator assistant in a Brazilian translation company, this perspective does not encompass the entire work of a sworn translator. This activity was formalized as an internship, under Brazilian legislation, and had a limited period of two years. Furthermore, all of the translated work done by an assistant is revised and then signed by a sworn professional.

For further research, some potential aspects that could 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.

The analysis conducted in this particular study also allowed us to gain a better understanding of MTPE as applied to sworn translation. Removing document formatting before uploading it to Smartcat eliminated errors caused by tags and demonstrated the importance of

experience and training in translation technologies. These skills not only prepare translators for the market but also help reduce temporal effort in sworn translation.

Theoretical studies in sworn translation require specific post-editing guidelines to provide a solid foundation for practice. These guidelines would provide educators with explicit guidance to help train translators in developing practical skills, such as describing document elements and using Word tools to significantly reduce formatting time.

Finally, this study may contribute to the training of translators, considering its descriptive nature. Knowledge of the use of technology in translation is crucial to the training of translation students or even novice translators, because of its relevance in the translation industry, its impact on productivity, and the inevitability that these tools will become more prevalent over the years.

REFERENCES

Anoreg/BR divulga Manual de Integração dos Cartórios e alerta sobre prazo de integração – ANOREG. (n.d.). Retrieved July 1, 2024, from https://www.anoreg.org.br/site/manuald-integracaodoscartorios/

Albuquerque, A. F. de. (2023). *The Impact and Challenges of Using Machine Translation Post-Editing in Sworn Translation of Brazilian Personal Documents* [Undergraduate Thesis, Universidade Federal de Pernambuco].

Barros, L. A., Babini, M., & Aubert, F. H. (2010). Terminologia e tradução juramentada: questões de tipologia textual e equivalência terminológica interlinguística Português-francês-italiano. *Filologia e Linguística Portuguesa*, *12*(2), 233-249. DOI: 10.11606/issn.2176-9419.v12i2p233-249

Bowker, L. (2002). *Computer-Aided Translation Technology: A Practical Introduction* (Didactics of Translation). University of Ottawa Press.

Cadastro Nacional de Tradutores Públicos e Intérpretes Comerciais – CNTPIC. (n.d.). Retrieved September 28, 2024, from https://www.gov.br/mdic/pt-br/assuntos/outros-assuntos/uncategorised/cntpic-drei

Carl, M., Dragsted, B., Elming, J., Hardt, D., & Lykke Jakobsen, A. (2011). The Process of Post-Editing: A Pilot Study. *Copenhagen Studies in Language*, (41), 131-142.

CRC integra os Cartórios de Registro Civil de todo o País. (2020, August 27). Blog Do Registro Civil. Retrieved July 1, 2024, from https://blog.registrocivil.org.br/2020/08/27/crc-integra-os-cartorios-de-registro-civil-de-todo-o-pais/

Goedert, A. O. (2015). *Tradução Juramentada e as Modalidades de Tradução: o caso dos históricos escolares* (Dissertação de mestrado, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo).

Gotti, M., & Šarčević, S. (2006). *Insights Into Specialized Translation* (2nd ed.). Peter Lang.

Green, S., Heer, J., & Manning, C. D. (2013). *The Efficacy of Human Post-Editing for Language Translation*. ACM Human Factors in Computing Systems (CHI).

Guedes, R. (2016). Tradução especializada: um mergulho nas especificidades do gênero jurídico. *Tradução em Revista*, *2*(21). DOI: 10.17771/PUCRio.TradRev.28028

Gutiérrez Arcones, D. (2015). Estudio sobre el texto jurídico y su traducción: características de la traducción jurídica, jurada y judicial. *Miscelánea Comillas: Revista de Ciencias Humanas y Sociales*, *73*(142), 141-175.

Holmes, J. S. (2000). The Name and Nature of Translation Studies. In L. Venuti (Ed.), *The Translation Studies Reader* (pp. 172-185). Routledge.

Hutchins, W. J., & Somers, H. L. (1992). *An Introduction to Machine Translation*. Academic Press.

Killman, J., & Rodríguez-Castro, M. (2022). *Post-editing vs. translating in the legal context: Quality and time effects from English to Spanish*. DOAJ (DOAJ: Directory of Open Access Journals). DOI: 10.2436/rld.i78.2022.3831

Koponen, M., Mossop, B., Robert, I. S., & Scocchera, G. (2021). *Translation Revision and Post-Editing: Industry Practices and Cognitive Processes* (1st ed.). Routledge.

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

Koglin, A. (2015). An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors. *The International Journal for Translation & Interpreting Research*, *7*(1), 126-141. DOI: ti.106201.2015.a06

Mayoral, R. (2000). Sworn Translation and Its Functions. *Fédération Internationale des Traducteurs (FIT) Revue, Babel*, *46*(4), 300-331. DOI: 10.1075/babel.46.4.03may

Mesa-Lao, B. (2013). 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.

Nascimento, L. de A. e S. (2006). *Investigating Norms in the Brazilian Official Translation of Semiotic Items, Culture-Bound Items, and Translator's Paratextual Interventions* [Doctoral dissertation, Universidade Federal de Santa Catarina].

Nitzke, J., & Hansen-Schirra, S. (2021). *A short guide to post-editing* (Translation and Multilingual Natural Language Processing, 16). Language Science Press. https://doi.org/10.5281/zenodo.5646896

O'Brien, S. (2006). *Machine-Translatability and Post-Editing Effort: an empirical study using Translog and Choice Network Analysis*. [Doctoral dissertation, Dublin City University, Ireland].

O'Brien, S. (2007). An empirical investigation of temporal and technical post-editing effort. *Translation and Interpreting Studies (TIS), 2*(1), 83–136.

O'Brien, S. (2010). Introduction to Post-Editing: Who, What, How and Where to Next? In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas: Tutorials*, Denver, Colorado, USA. Association for Machine Translation in the Americas.

Oliveira, B. M. de. (2012). Tecnologias da tradução no trabalho de tradutores jurídicos/juramentados: estudo de caso. Abehache, 2(3), 191-205. Retrieved from https://revistaabehache.com/ojs/index.php/abehache/article/view/84/83

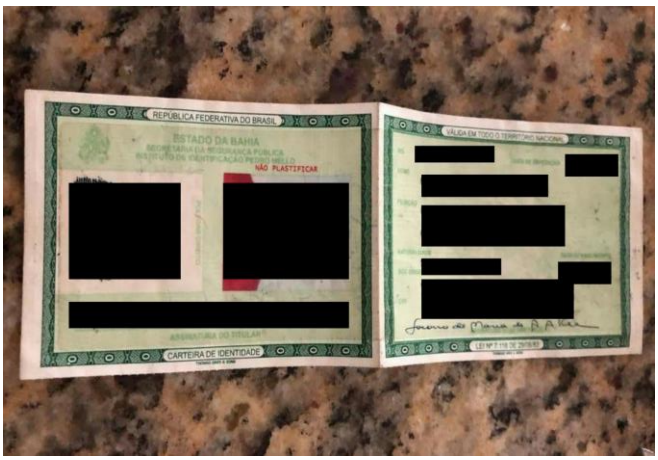
Omazić, M. & Šoštarić, B. (2023). New resources and methods in translating legal texts: machine translation and post-editing of machine-translated legal texts. In L. Kordić. (Ed.), *Language(s) and law* (pp. 71-83). Faculty of Law Osijek.

Ramos, F. P. (2014). Legal Translation Studies as Interdiscipline: Scope and Evolution. *Meta*, *59*(2), 260-277. DOI: 10.7202/1027475ar

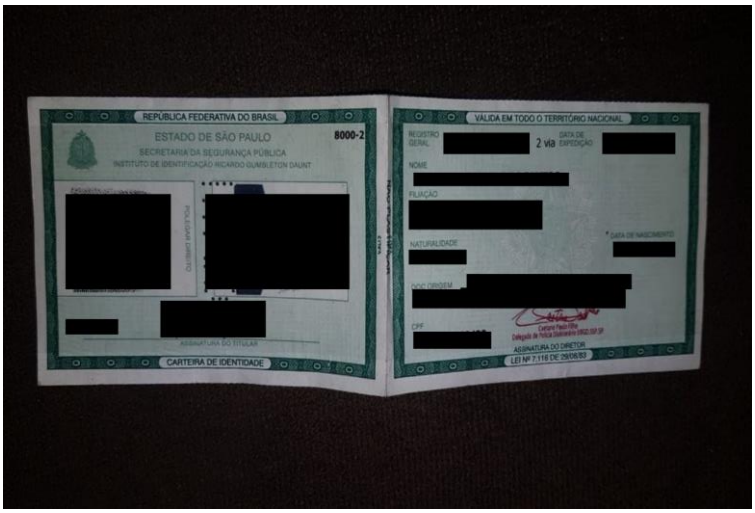
Sawant, D. (2013). History of Translation. *Literary Endeavor, 4*, 109-115.

Smartcat. (n.d.). Why Smartcat is safe to use. Retrieved from https://www.smartcat.com/security/

Appendix A - “ID card 1”

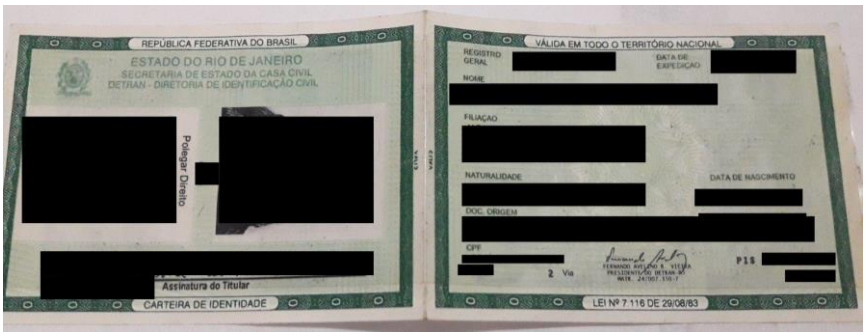


Appendix B - “ID card 2”

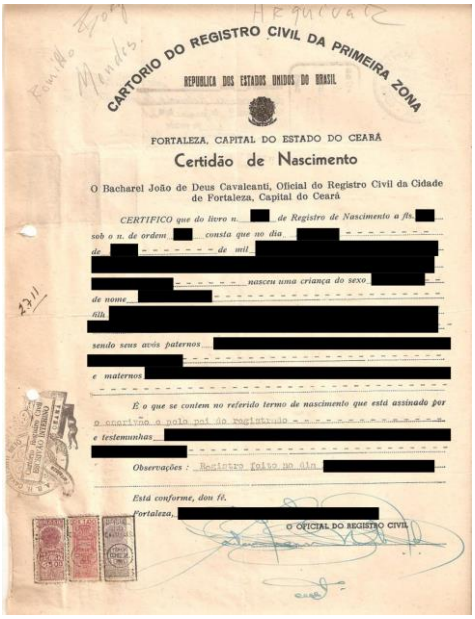


Appendix C - “ID card 3”

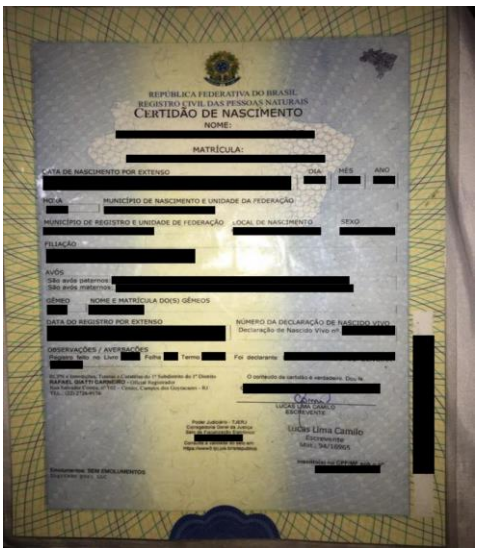
Appendix D - “ID card 4”



Appendix E - “Birth Certificate 1”



Appendix F - “Birth Certificate 2”



ALBUQUERQUE, Aíla Fonseca de; KOGLIN, Arlene. Uso da MTPE no contexto da tradução juramentada: ônus ou bônus?. *Revista Belas Infiéis*, Brasília, v. 13, n. 1, p. 01-25, 2024. e-ISSN: 2316-6614. DOI: 10.26512/belasinfieis.v13.n1.2024.54644

Appendix G - “Birth Certificate 3”

REPÚBLICA FEDERATIVA DO BRASIL

CERTIDÃO DE NASCIMENTO

Romero Longman

Appendix H - “Birth Certificate 4”

REPÚBLICA FEDERATIVA DO BRASIL

OFÍCIO DE REGISTRO CIVIL DAS PESSOAS NATURAIS

MUNICÍPIO E COMARCA DE RIO NEGRO • ESTADO DO PARANÁ

Carmen Lucia Bley Martins

(Oficial)

CERTIDÃO DE NASCIMENTO

[i] This article is based on the study conducted for the undergraduate thesis entitled “The impact and challenges of using machine translation post-editing in sworn translation of Brazilian personal documents” and defended in 2023. This is a revised and rewritten text. It includes theoretical and methodological updates as well as expanded analysis.

[ii] Our translation for “Cadastro Nacional de Tradutores Públicos e Intérpretes Comerciais”.