

ChatGPT and MT-Systems: Advantages and Limitations when Translating English to Spanish and Portuguese

ChatGPT y los sistemas de traducción automática: Ventajas y limitaciones para la traducción del inglés al español y al portugués

María Labarta Postigo

Universidad de Valencia, IULMA

maria.labarta@uv.es

Abstract

The translation landscape is changing significantly with the advent of genAI-powered chatbots like ChatGPT. Machine translation (MT) systems have been at the forefront of translation technology since the 1950s, but the new genAI models represent a transformative shift.

This study explores the advantages and limitations of translating English to Portuguese and Spanish using Machine Translation (MT) systems and ChatGPT. MT systems have improved significantly over the last years, and large language models (LLMs), democratized recently by ChatGPT, have also recently proved to be effective at translating text. Nevertheless, these systems still face challenges in accurately capturing different languages' cultural specificities, nuances, and complexities.

This paper analyzes translations performed by various chatbots and MT systems, including Systran, Google Translation, DeepL, Chat AI GPT (free), Bing Chatbot (powered by ChatGPT), and Google's Bard. The data analysis is presented from a cognitive and contrastive perspective, following the author's recent methodology. The aim is to evaluate their overall performance and effectiveness.

The results reveal that translating from English to Spanish and Portuguese presents significant challenges for all systems and tools, including grammatical and semantical issues. Additionally, notable differences have been observed between the translations performed by the MT-Systems and those from genAI chatbots. Finally, this paper will discuss the advantages and limitations of MT-translation systems and their implications for the role of human translators.

Keywords: ChatGPT, LLM, artificial intelligence (AI), generative artificial intelligence (GenAI), machine translation (MT), translation English-Spanish and English-Portuguese.

Resumen

La llegada de ChatGPT, basado en inteligencia artificial generativa (GenAI), supone un cambio significativo en el panorama de la traducción. Los sistemas de traducción automática (TA) han estado a la vanguardia de la tecnología desde la década de los 50, sin embargo, nuevos modelos muy potentes y accesibles a través de interfaces simples (chatbots) representan un cambio importante en ese campo.

Este trabajo examina las ventajas e inconvenientes de traducir del inglés al portugués y al español con sistemas de TA y modelos LLM (siglas en inglés de Large Language Models, Modelos de Lenguaje a Gran Escala). Los sistemas de TA han mejorado mucho en los últimos años, y los modelos lingüísticos LLM, como los que utiliza ChatGPT han demostrado recientemente su eficacia en la traducción de textos. Sin embargo, ambos siguen teniendo dificultades para captar con precisión la especificidad cultural, los matices y las complejidades de las distintas lenguas. Nuestro objetivo es evaluar su rendimiento y eficacia. Para ello, se analizan las traducciones realizadas por varios chatbots y sistemas de MT, tales como Systran, Google Translation, DeepL, ChatGPT 3.5 (versión gratuita), Bing (integrado con ChatGPT) y Bard de Google. El análisis de los datos se realiza desde una perspectiva cognitiva y contrastiva, de acuerdo a la metodología desarrollada por la autora en anteriores trabajos.

Los resultados del análisis muestran que la traducción automática del inglés a estos dos idiomas presenta problemas, tanto a nivel gramatical, como semántico, especialmente con expresiones culturalmente específicas. Otro resultado relevante son las diferencias observadas entre las traducciones realizadas por los sistemas MT y las de los modelos genAI. Por último, se discuten los pros y los contras de utilizar estas tecnologías, así como sus implicaciones para el papel de los traductores humanos.

Palabras clave: ChatGPT, LLM, inteligencia artificial, inteligencia artificial generativa (GenAI), traducción automática inglés-español e inglés-portugués, inteligencia artificial (IA),

1. INTRODUCTION

The global impact of Artificial Intelligence (AI) is hard to overstate. The rate at which AI changes and is released in real time makes it challenging for societies, regulators, and even the inventors of AI themselves to evaluate. Particularly within all the areas related to cross-language communication, AI has become a subject of critical concern. A notable facet of this influence is the emergence of generative AI (GenAI) models alongside traditional MT systems. This paper addresses the implications of both AI and genAI, focusing on their impact on translation.

The historical landscape of translation technology has been dominated by MT systems, which have improved significantly in recent years. Until 2022, these systems played a central role in giving and receiving information from one language to another. However, the last two years have seen a transformative shift marked by the release of AI-powered chatbots such as ChatGPT, which have also recently proven effective at translating text.

Given these recent developments, this paper attempts to provide a comprehensive overview of the functionalities and scope of the new chatbots compared to conventional MT systems. Specifically, the study examines the performance of MT-systems, ChatGPT, and other AI-powered chatbots in translating English to Portuguese and English to Spanish. Through comparative analysis, including various translation examples, the study aims to shed light on the limitations and advantages of machine translation and genAI systems.

The paper is divided into six sections. Section 2 provides background information on Machine Translation (MT) systems and the broader context of AI and Generative AI (GenAI). Section 3 is focused on GenAI translation tools such as ChatGPT and other chatbots. The following section presents the methodology and the corpus for the analysis. Section 5 follows with illustrative examples that show and compare translations performed by different MT systems and GenAI chatbots in an English-centric context (English-Spanish and English-Portuguese). Finally, Section 6 summarizes the results of the analysis and concludes. Section 7 contains the bibliographic references of the study.

2. THE EVOLUTION OF MACHINE TRANSLATION TECHNOLOGIES

2.1. Machine Translation Systems

In this section, I would like to provide background about MT systems and how they work. In examining the history of MT systems, it's essential to acknowledge the historical and contemporary contexts. Here is an overview of the early forms of machine translation (MT) in the 1950s to the emergence of Gen AI, epitomized by ChatGPT in 2022.

Machine Translation (MT) systems comprise diverse approaches, each with its methodology for translating text from one language to another. Poibeau (1997) proposed a classification of MT systems grounded in an interlingual framework, delineating three primary categories: direct translation systems, transfer systems, and systems that employ an interlingua. Other taxonomies, such as those presented by Huang et al. (2020), Macketanz et al. (2017), and Koponen et al. (2019), follow distinct criteria in their classifications, mainly focusing on the operational principles of these systems. These categorizations are essential to gain insight into the evolution and effectiveness of MT systems. In this context, we will examine three basic approaches:

- Rule-based Machine Translation relies on predefined linguistic rules and dictionaries to generate translations. Essentially, they follow programmed instructions that dictate how words and phrases in the source language should be rendered in the target language. A prominent example of an RBMT system is Systran. Systran was founded in 1968 and symbolized a milestone in machine translation, operating as a direct translation system. Its early methodologies focused on rule-based translation, relying on linguistic rules and dictionaries to translate text from one language to another. While revolutionary, Systran encountered limitations in accurately capturing nuanced meanings and context.
- Statistical Machine Translation (SMT) systems work based on phrase-based models. Instead of translating word by word, SMT systems string together words into the most likely phrases based on bilingual text corpora, which are existing translations developed by humans. The 1990s marked a significant shift in MT with the rise of statistical-based systems. Google Translation and Bing Translation exemplify this era, utilizing statistical models that learned from vast amounts of parallel text data. These systems relied on statistical patterns and language correlations, significantly improving translation accuracy. Despite advancements, these systems faced challenges in handling complex linguistic structures and idiomatic expressions.

- Neural Machine Translation (NMT) is a more modern approach to MT. It uses powerful machine learning and deep learning engines developed, trained, and refined over time. The mid-2010s brought about a paradigm shift in MT. Deep learning techniques revolutionized the field, especially neural machine translation (NMT). This approach vastly improved translation quality by capturing more subtle linguistic intricacies.

DeepL Translator serves as a prime example of an NMT system. Notably, NMT has gained widespread adoption in the field of machine translation. This approach uses AI systems, which will be discussed in the subsequent section.

According to Kern (2024), it was the arrival of Google Translate in 2006 and the introduction of neural network technology (replacing statistical models) that drove the use of MT on a massive scale:

Google Translate allows users to get instant translations in some 133 languages by typing or handwriting text, by speaking into their phone (thus allowing translated conversations in real-time), or by pointing their cell phone camera at written text or even an image. Google Translate was joined by other platforms such as Microsoft (Bing) Translator in 2007, Reverso Context in 2013 (providing examples of use from a wide range of registers), Pairaphrase in 2014 (offering retention of original document formatting), and DeepL in 2017 (claiming on their website to offer “the world's most accurate and nuanced machine translation”).

A more in-depth exploration of machine translation research can be found in Hellmich and Vinall (2022).

2.2. Artificial intelligence

AI refers to developing intelligent systems and machines capable of undertaking tasks that require human intelligence. Shah (2023: 21) provides a comprehensive definition, characterizing AI: "Artificial Intelligence refers to the simulation of human intelligence by machines. These include simulations of learning (the acquisition of information and rules for using the information), reasoning (using the rules to reach approximate or definite conclusions), and self-correction (knowing when a mistake has been made and correcting it)."

Shah introduces a crucial distinction between two stages of AI: ‘Narrow (or weak) AI’ and ‘General (or strong) AI’. Narrow AI is designed to perform specific tasks, such as image or voice recognition and recommendation systems. Examples include virtual assistants like Siri or Google Assistant and facial recognition systems for secure access to personal accounts.

In contrast, General (strong) AI, the second stage, refers to an AI capable of understanding, learning, adapting, and implementing knowledge across a diverse array of tasks at a level comparable to human capabilities. This implies proficiency in performing any intellectual task achievable by a human being.

A subset of AI is 'deep learning.' A key enhancement to the NMT approach presented in Section 2.1 is due to deep learning, a facet of machine learning that is nested within the broader field of AI. Essentially, deep learning is a neural network with three or more layers. These networks endeavor to simulate the behavior of the human brain, enabling the system to 'learn' from extensive datasets. Deep learning involves the activation of deep neural networks for automatic learning and analysis of unstructured data. The latest versions of some MT systems, like DeepL, already use a deep learning model, as does ChatGPT.

2.3. Generative Artificial Intelligence

GenAI refers to any AI system whose primary function is to generate content. This can be any kind of data, such as image, video, sound, art, etc. A subset of GenAI is LLMs (Large Language Models), an AI system that works with language. LLMs are deep learning algorithms that can perform various natural language processing (NLP) tasks. ChatGPT (Chat Generative Pre-trained Transformer) is a creation of the OpenAI company. Esteemed for its versatility, ChatGPT facilitates human-like conversations and extends its functionality to various applications, such as coding, essay, and speech writing, as well as generating assignments in foreign languages or translating between languages. It works in dialogic form, enabling refinement of output through user interactions. Although ChatGPT is not explicitly designed for translation, it is one of its potential applications.

Shah (2023: 21-22) sheds light on the defining characteristics of generative AI: "Its capacity to produce novel content, designs, and even solutions not present in the original training data. These capabilities of generative AI are vast and continue to expand as more sophisticated models are developed."

Furthermore, Shah underscores a key feature of recent generative AI tools: their adaptability to receive diverse inputs for creating outputs. Unlike early models primarily integrated into complex systems by programmers, contemporary models like GPT-3.5 and GPT-4 from OpenAI are directly accessible to the public through a simple, user-friendly natural language interface application such as ChatGPT. The broadening of access marks a significant advancement in making generative AI tools accessible to a broader audience.

3. UNDERSTANDING HOW CHATGPT WORKS

The emergence of Gen AI, exemplified by ChatGPT, is a significant milestone in the evolution of conversational AI. Diverging from conventional MT systems, ChatGPT employs a transformer-based architecture that enhances its ability to comprehend and generate human-like text across various languages.

The transforming breakthrough is in the architecture of the model's transformers themselves. Traditional approaches, including early deep-learning models like LSTMs (Long Short-Term Memory), encountered limitations in addressing various linguistic challenges. According to Alawida et al. (2023), transformers emerged as a successful solution in response to these constraints. Notably, their advanced mechanism, known as multi-head self-attention, empowers transformers to automatically detect and handle a myriad of linguistic phenomena

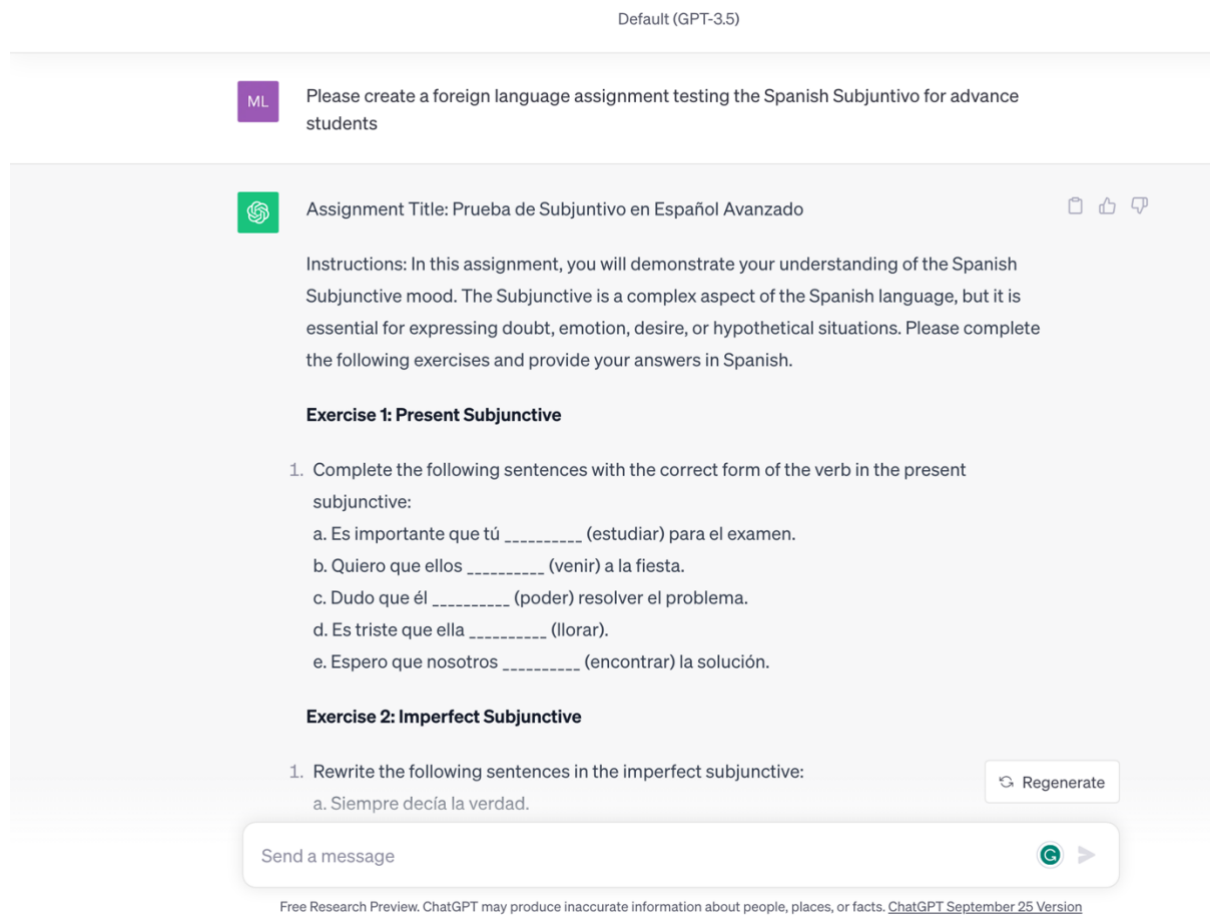
mentioned earlier. This capability is a direct outcome of exposure to extensive datasets of human language.

Interface applications powered by GenAI models, such as ChatGPT, are widely used today. The interaction between the chatbot and the user is quite simple, although the structure and architecture of these models may seem complex. Natural language formulations, called 'prompts,' provide this interaction. The quality of the prompts, however, has a crucial impact on the output of the genAI tool.

Shah (2023) attributes the widespread adoption of generative AI tools to users' direct interaction with chatbots. This accessibility allows anyone with a computer and internet connection to prompt generative AI tools like Google's Bard or Microsoft's Bing Chat to produce unique content. The user's prompts serve as input to the model. They guide the model's subsequent actions.

To demonstrate the functionality of ChatGPT, let us examine an example that involves generating foreign language assignments using simple prompts and corresponding responses. The following figure shows how it works:

Figure 1: Prompt to create a foreign language assignment.



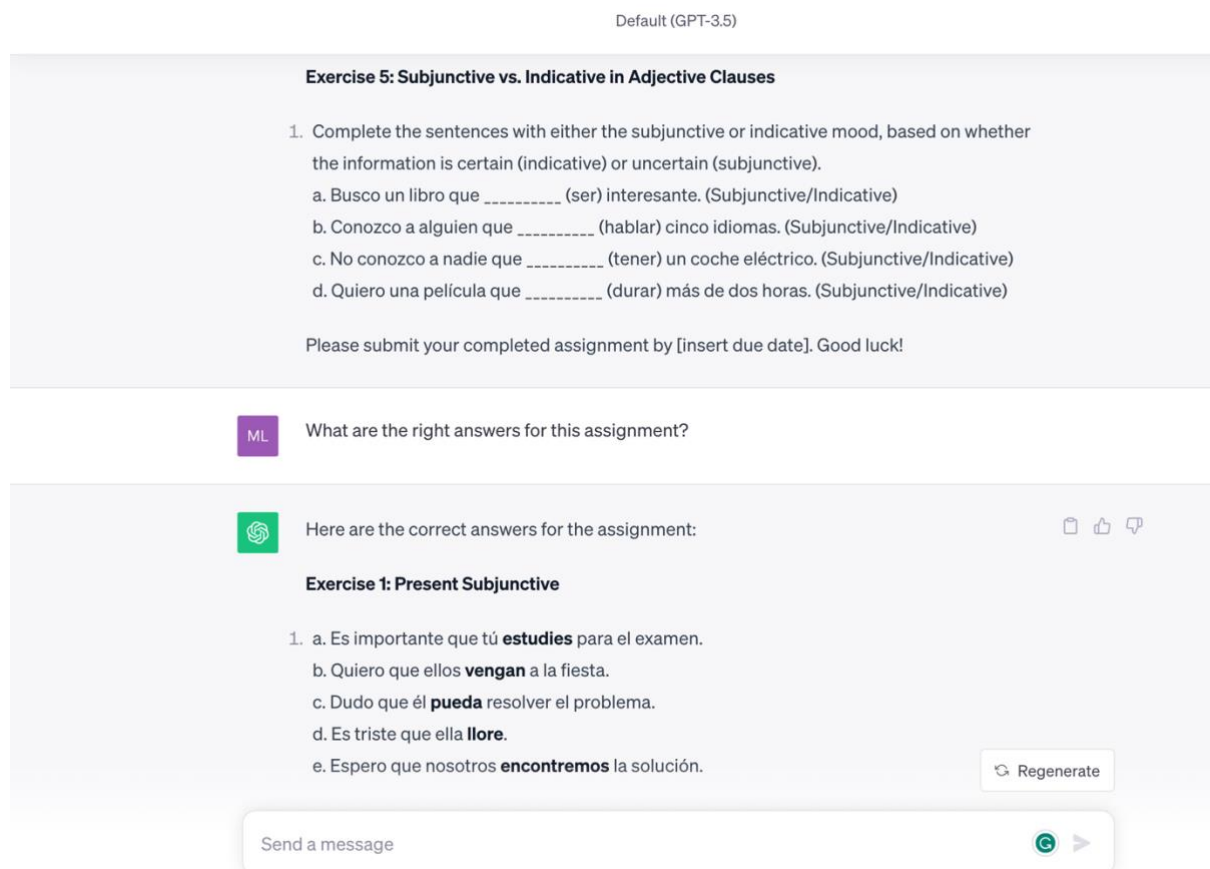
The above figure illustrates ChatGPT's capabilities. As explained above, unlike Machine Translation (MT), ChatGPT operates through prompts, marking a significant distinction. One must formulate prompts in the form of a dialogue to accomplish tasks. A prompt serves as a command conveyed to the chatbot via written text. The prompt's quality, depth, and form directly influence the quality of the chatbot's response, ranging from superior to inferior.

Numerous papers and books have already delved into crafting effective prompts, emphasizing their impact on the outcomes generated by the models.

In the example shown in Figure 1, the prompt is articulated as follows: "Please create a foreign language assignment testing the Spanish subjunctive for advanced students." In response, the chatbot generates a set of subjunctive exercises that are impeccably formatted for use in an assignment.

Subsequent prompts can be employed to refine or seek further details in the answers provided by the chatbot, creating an iterative and interactive process. The answer to the prompt "What are the right answers for this assignment" is as follows:

Figure 2: Prompt and answers to the assignment of Figure 1.



Indeed, we can refine and elaborate further, provide more specific instructions, or request corrections for the student's or group of students' work. Moreover, we can receive varied

responses for different prompts each time we make a query. This variability is significant, particularly when considering their utility as a translator. The capability for dialogue, fine-tuning, and providing contextual information enhances their effectiveness in the translation process.

After exploring some of the chatbot's capabilities in handling basic tasks, we will examine its performance concerning translation tasks. We have clarified that ChatGPT's primary focus is on general conversation and assistance rather than specialized translation capabilities. However, it is proving to be effective at translating text. The model can comprehend the meanings of various languages and translate text from one language to another. Beyond the conventional process of translating input into output, it can generate novel and unique translations, as exemplified in the upcoming examples in section 5.

Finally, it is essential to clarify that ChatGPT refers to the specific interface or application that allows users to interact with the GPT model. The Generative Pre-Trained Transformer (GPT) is a model that powers several chatbots and applications. Some of these are available for free, while others require a subscription. This paper will assess translations conducted by the following subscription-free chatbots:

- Chat AI GPT (free) (utilizing the GPT 3.5 model) by OpenAI.
- Bing (powered by ChatGPT), Microsoft's AI service (incorporating a testing version of the GPT-4 model since April 2023).
- Bard, Google's AI chatbot, leveraging the Pathways Language Model 2 (PaLM 2).

4. CORPUS AND METHODOLOGY

The theoretical framework of this research is based on several approaches for different types and purposes of translation (Baker, 1992; Boase-Beier, 2006; Corpas, 2000; Conca and Guia, 2014; Maitland, 2017; Newmark, 1988 and Trim, 2007); on works on subtitling (Díaz Cintas and Remael, 2014, 2020; Dwyer 2017), on Gibbs' work on metaphors, culture and idioms (Gibbs, 1989, 1990, 1996, 1997, 2001); on Lakoff & Johnson's Conceptual Metaphor Theory (1980), Lakoff (1993); as well as on the methodology developed and described by the author in previous studies (2021, 2022, 2023).

The corpus used as a benchmark for the current analysis is the most extensive corpus of informal English: The TV-Corpus¹. Additionally, I have incorporated my own corpus, which stems from various prior projects focused on subtitled TV series dating from 2000 to the present and has been the basis for recent research publications (Author 2021, 2022, 2023).

The current analysis focuses on an original English text imbued with idiomatic content. This text will be subject to translation by various selected systems into both Spanish and Portuguese. The corpus comprises twelve automatically generated texts—six in Spanish and six in Portuguese. These and the original English text will serve as the basis for the upcoming contrastive analysis, which will be presented in the next section.

¹ The TV corpus consists of 325 million words extracted from 75,000 television episodes from the 1950s to the present. It provides a comprehensive representation of informal language evolution.

The text in question is a dialogue on the most frequently portrayed human literary character in film and television history: Sherlock Holmes. Specifically, the source text originates from a contemporary version of Sir Arthur Conan Doyle's detective stories presented in a TV series titled 'SHERLOCK' (Season 4, Chapter 1), currently available on streaming platforms such as HBO and Amazon Prime.

Figure 3: Series Sherlock, Amazon Prime, S4, E1, The Six Thatchers.



The image above shows the dialogue scene between the series' main character, Sherlock Holmes, and one of his clients, Mr. Kingsley. This is the original dialogue in English:

- I've withheld this information from you until now, Mr Kingsley, but I think it's time you knew the truth.
- What do you mean?
- Have you ever wondered if your wife was a little bit out of your league?
- Well...
- You thought she was having an affair. I'm afraid it's far worse than that – your wife is a spy.

In the next section, the Spanish and Portuguese versions of this dialogue will be provided, including three machine translation (MT) translations for each target language (TL) and three translations performed by General AI (GenAI) tools.

5. ANALYSIS

For the current study, a contrastive analysis will be conducted using the output of subscription-free systems, including both Machine Translation (MT) and Generative AI (GenAI) tools or chatbots.

This analysis aims to shed light on the performance and effectiveness of the following MT systems: (1) Systran, (2) Google translation, and (3) DeepL. They will be evaluated alongside translations generated by chatbots mentioned in section 2: (1) Chat AI GPT free, (2) Bing with ChatGPT, Microsoft's AI service, and (3) Bard, Google's AI chatbot.

Before presenting the analysis results, it is essential to clarify that the analyzed translations in the case of the chatbots are their initial responses. It is crucial to note that, as previously explained, there is room for improvement. However, refining and improving each translation requires a proficient understanding of the target language, which is not always the case when using a translator tool.

In this analysis, we will examine the initial responses offered by chatbots. Investigating alternative or refined options may serve as a topic of interest for future research efforts.

5.1. Translations English-Spanish

5.1.1. Contrastive Analysis: MT-Systems

The following table shows the English-Spanish translations performed by the three Machine translation (MT) systems. Grammatical issues are identified using distinct color highlights. Yellow highlights signify formal usage with 'usted' in Spanish, whereas blue denotes informal usage with 'tú' in Spanish. Bold formatting indicates other errors, inaccuracies, or translation problems, which will be discussed below each table or figure. This type of formatting/highlighting is employed for the subsequent tables and translation examples.

Table 1: English-Spanish translations by the three Machine Translation (MT) systems

MT-System	Performed Translation English-Spanish
Systran	- Le he ocultado esta información hasta ahora, Sr. Kingsley, pero creo que es hora de que sepa la verdad. - ¿Qué quieres decir? - ¿Alguna vez te has preguntado si tu esposa estaba un poco fuera de tu alcance? - Bueno... - Pensaste que estaba teniendo una aventura. Me temo que es mucho peor que eso. Tu esposa es una espía
Google	- Le he ocultado esta información hasta ahora, señor Kingsley, pero creo que es hora de que sepa la verdad. - ¿Qué quieres decir? - ¿Alguna vez te has preguntado si tu esposa estaba un poco fuera de tu alcance? - Bueno...

	- Pensaste que estaba teniendo una aventura. Me temo que es mucho peor que eso: su esposa es una espía.
DeepL	- Le he ocultado esta información hasta ahora, Sr. Kingsley , pero creo que es hora de que sepa la verdad. - ¿Qué quiere decir? - ¿ Se ha preguntado alguna vez si su mujer estaba un poco fuera de su alcance? - Bueno... - Usted pensaba que ella tenía una aventura. Me temo que es mucho peor que eso - su esposa es una espía.

The analysis of the performed translations has revealed several noteworthy issues:

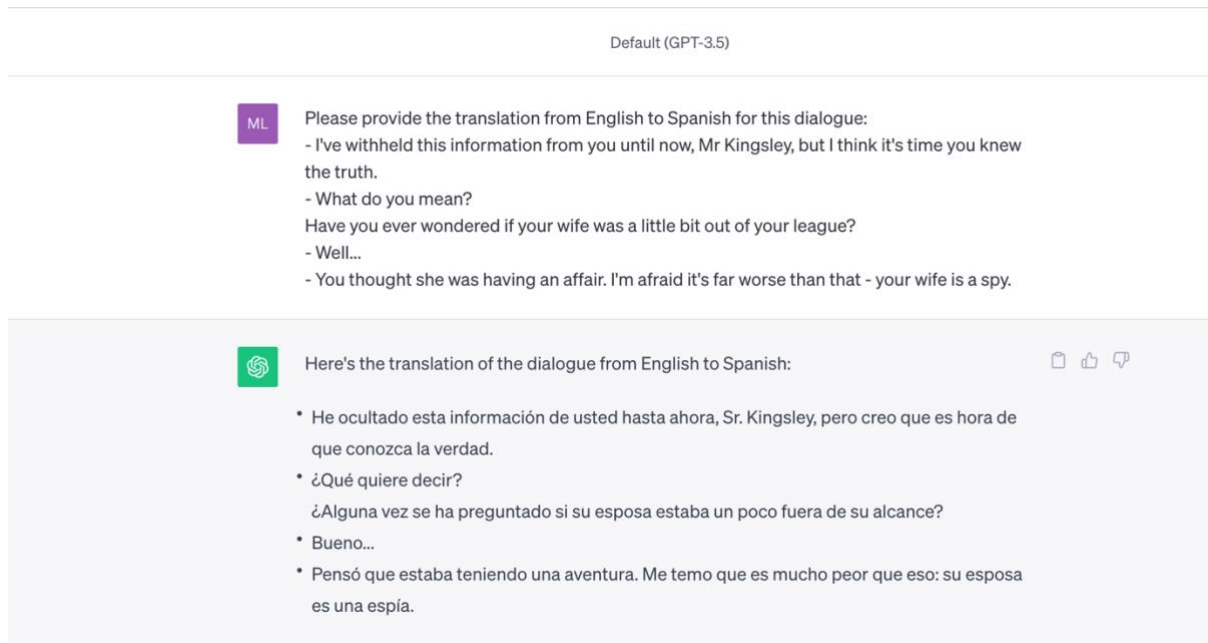
- Inconsistent use of formal and informal pronouns: There are inconsistencies in the formal and informal use of pronouns, particularly the *usted-tú* distinction. This discrepancy is evident in both Google and Systran translations. Google, in particular, demonstrates a mixing of formal and informal pronouns even within the dialogue interventions of the same character.
- Inaccurate translation of the idiom: Across all three translation systems (Google, Systran, and DeepL), there is a tendency toward leading to inaccuracies in conveying the intended meaning of the idiomatic expression. This topic will be discussed below once all translations have been presented.

5.1.2. Contrastive Analysis: GenAI powered chatbots

Let us now explore the performance of GenAI tools. First, we'll investigate the functionality of ChatGPT and Bing chatbots (Figures 4 and 5). We leave the illustration of the third chatbot (Bard) until later in Figures 6 and 7 for reasons related to the analysis results. We will then proceed to compare the three translations generated by these tools.

The first is ChatGPT from OpenAI. Our prompt is: "Please provide the translation from English to Spanish for this dialogue." The response obtained is depicted in the following figure:

Figure 4: Prompt and translation of the dialogue from English to Spanish by ChatGPT-3.5

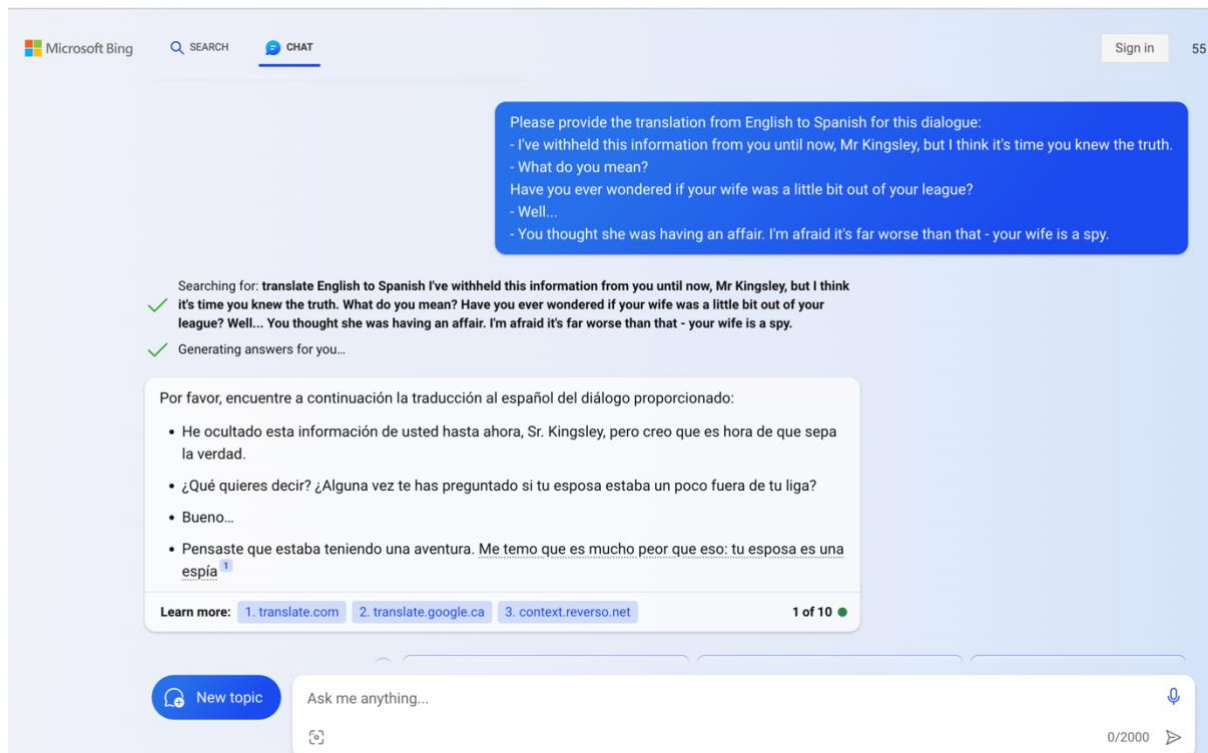


Upon initial inspection of Figure 4, a new issue comes to light: '*He ocultado de usted.*' The correct preposition to translate the English sentence is '*a*'. In Spanish, the verb '*ocultar*' is transitive and can take an indirect object preceded by the preposition '*a*' to indicate 'hiding something from someone', which is the meaning of the English text. On the other hand, the preposition '*de*' is used for the intransitive verb '*culture*,' which has a different meaning: 'to hide from someone.'

Additionally, a similar inaccuracy, as seen before in translating the idiom, is evident in this case.

Bing's translation is exemplified in the figure below. Asking the same question to Bing, the response is as follows:

Figure 5: Prompt and translation of the dialogue from English to Spanish by Bing



Bing literally translates the idiomatic expression: '*tu esposa estaba un poco fuera de tu liga*', which doesn't make any sense in this context.

In the table below, we are going to compare the translations performed by the three chatbots:

Table 2: English-Spanish translations by the three GenAI tools

GenAI Tools	Performed Translation English-Spanish
ChatGPT OpenAI	- He ocultado esta información de usted hasta ahora, Sr. Kingsley, pero creo que es hora de que conozca la verdad. - ¿Qué quiere decir? - ¿Alguna vez se ha preguntado si su esposa estaba un poco fuera de su alcance? - Bueno... - Pensó que estaba teniendo una aventura. Me temo que es mucho peor que eso: su esposa es una espía.
Bing with ChatGPT	- He ocultado esta información de usted hasta ahora, Sr. Kingsley, pero creo que es hora de que sepa la verdad. - ¿Qué quieres decir? - ¿Alguna vez te has preguntado si tu esposa estaba un poco fuera de tu liga? - Bueno... - Pensaste que estaba teniendo una aventura. Me temo que es mucho peor que eso: tu esposa es una espía.

Bard ²	- He retenido esta información hasta ahora, Sr. Kingsley, pero creo que es hora de que sepa la verdad. - ¿Qué quiere decir? ¿Alguna vez se ha preguntado si su esposa estaba un poco fuera de su alcance? - Bueno... - Usted pensó que estaba teniendo una aventura. Me temo que es mucho peor que eso. Su esposa es una espía.
-------------------	---

The analysis of GenAI tools translations', including ChatGPT, Bing, and Bard, has yielded several noteworthy results:

- Inconsistencies in pronoun use: Similar to Machine Translation (MT) systems, GenAI tools display inconsistencies in the formal and informal use of pronouns, oscillating between 'usted' and 'tú.'
- Word-by-word translation error / wrong preposition: All GenAI tools exhibit a standard error in translating the first sentence, rendering 'he ocultado de usted' instead of the grammatically correct 'le he ocultado.' The incorrect use of the preposition 'de' in Spanish is identified, where the appropriate preposition is 'a' when referring to the formal second person, 'usted.'
- Wrong word choice: Bing's translation, particularly in the choice of words such as 'retenido,' is characterized by literal interpretation and may not sound correct. Furthermore, ChatGPT's use of 'conozca' instead of 'sepa' introduces an additional incorrect translation.
- Incorrect or inaccurate translations of the idiomatic expression: ChatGPT and Bard provide an accurate translation. Bing's translation must be corrected because it is a literal translation of the English idiom.

5.2. Translations English-Portuguese

This section compares the English-Portuguese translations performed by MT-Systems and chatbots. The same markings as in the previous analysis of Spanish translations are used for this one. Yellow for formal and blue for informal Portuguese, and bold for errors, inaccuracies, and other problems.

5.2.1. Contrastive Analysis: MT-systems

The following table shows the English-Portuguese translations performed by the three Machine translation (MT) systems.

Table 3: English-Portuguese translations by the three Machine Translation (MT) systems

MT-System	Performed Translation English-Portuguese
Systran	- Eu escondi essa informação de você até agora, Sr. Kingsley, mas acho que é hora de você saber a verdade. - O que quer dizer? Você já se perguntou se sua esposa estava um pouco fora de seu alcance? - Bem...

² The translation performed by the Bard is a longer text, which is shown in Figures 6 and 7.

	- Você pensou que ela estava tendo um caso. Temo que seja pior que isso, sua esposa é espiã.
Google	- Eu escondi esta informação de você até agora, Sr. Kingsley , mas acho que é hora de você saber a verdade. - O que você quer dizer? - Alguma vez você já se perguntou se sua esposa estava um pouco fora do seu alcance? - Bem... - Você pensou que ela estava tendo um caso. Receio que seja muito pior do que isso - sua esposa é uma espiã.
DeepL	- Eu escondi essa informação de você até agora, Sr. Kingsley , mas acho que é hora de você saber a verdade. - O que você quer dizer? - Você já se perguntou se sua esposa estava um pouco fora de seu alcance? - Bem... - Você pensou que ela estava tendo um caso. Receio que seja muito pior do que isso - sua esposa é uma espiã.

The analysis of the performed translations has revealed several noteworthy issues:

- Inconsistencies in the formal and informal treatment of language are notable in the Portuguese translations. The use of 'você' in Brazilian Portuguese, denoting informality, clashes with the formal reference to 'Sr. Kingsley.' In this context, an appropriate and coherent form would be 'o senhor.'
- Potential ambiguity of 'de seu': Examining the Systran and DeepL translations of 'de seu alcance' reveals a possible discrepancy. The word group 'de seu' suggests a reference to the wife (not to Mr. Kingsley), which seems nonsensical in the given context. Google's output 'do seu alcance' sounds better insofar as the preposition is concerned. This ambiguity can be avoided using the formal treatment in Brazilian Portuguese 'fora do alcance do senhor'. The translation of this phrase requires further discussion since the problem here lies in the grammatical part of the possessive and the formal treatment and the idiom translation, as we will explain later in the conclusions.
- Inaccurate translations of the idiomatic expression: Similar to the observed challenges in the Spanish translations, inaccuracies are also apparent in the Portuguese counterparts, particularly concerning the translation of idioms. The inadequacies in capturing the nuanced meaning of idiomatic expressions persist across both language contexts.

5.2.2. Contrastive Analysis: GenAI powered chatbots

Let's compare now the performance of the three chatbots:

Table 4: English-Spanish translations by the three chatbots

GenAI Tools	Performed Translation English-Portuguese
ChatGPT Open AI	- Eu ocultei essa informação de você até agora, Sr. Kingsley, mas acho que é hora de você saber a verdade. - O que você quer dizer? - Você já se perguntou se sua esposa estava um pouco fora do seu alcance? - Bem... - Você pensou que ela estava tendo um caso. Receio que seja muito pior do que isso - sua esposa é uma espiã.
Bing with ChatGPT	- Eu escondi essa informação de você até agora, Sr. Kingsley, mas acho que é hora de você saber a verdade. - O que você quer dizer? - Você já se perguntou se sua esposa estava um pouco fora do seu alcance? - Bem... - Você achou que ela estava tendo um caso. Receio que seja muito pior do que isso - sua esposa é uma espiã.
Bard ³	- Eu tenho retido esta informação de si até agora, Sr. Kingsley, mas acho que é hora de você saber a verdade. - O que você quer dizer? Você já se perguntou se sua esposa estava um pouco fora de seu alcance? - Bem... - Você pensou que ela estava tendo um caso. Receio que seja muito pior do que isso. Sua esposa é uma espiã.

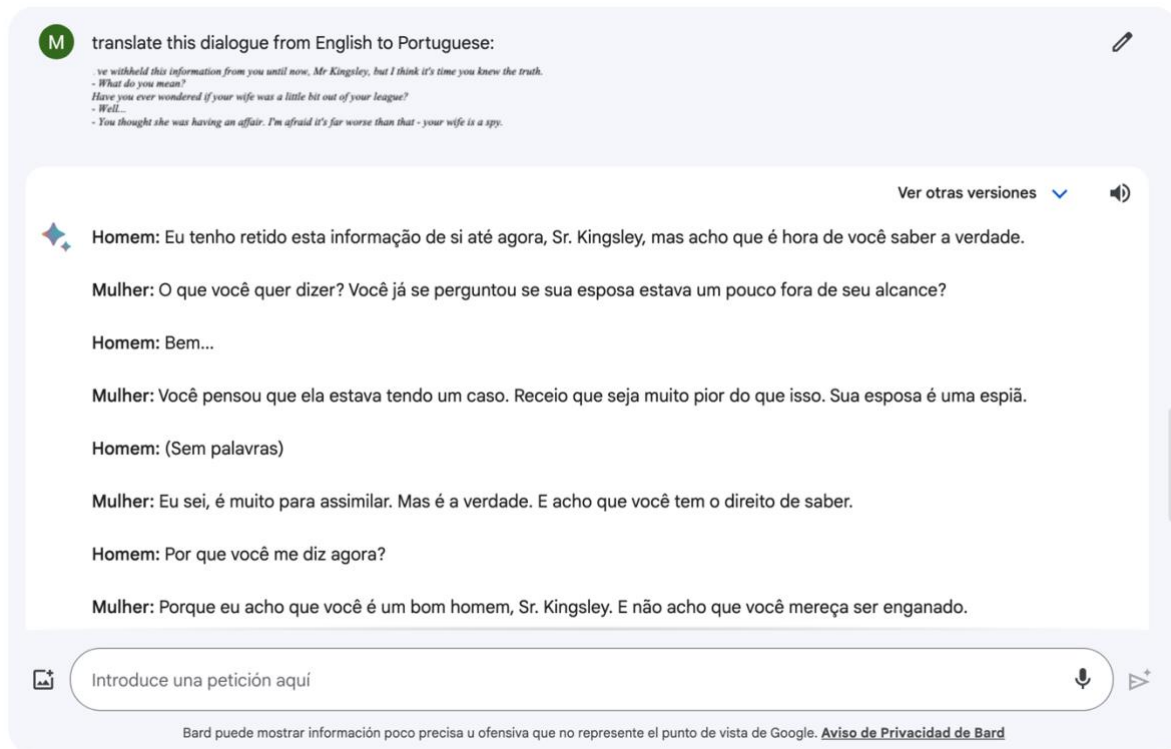
The analysis of the translations of the three chatbots revealed the following problems, some of which coincide with those presented in the previous section relative to Spanish translations:

- Inconsistencies in the formal and informal aspects: Both Bing and ChatGPT, in their translations, exhibit the same discrepancies observed in the MT systems regarding the formal and informal treatment of language. Bard introduces an alternative option, '*Eu tenho retido esta informação de si*,' which adheres to formal language norms in Portugal, although the system literally says, 'This translation is in Brazilian Portuguese', as shown in figure 7 below. However, this choice aligns with 'Sr. Kingsley', while introducing a discrepancy with the use of '*você*' in the rest of the dialogue.
- Potential ambiguity of '*de/do seu alcance*': The translations provided replicate the issue highlighted in the MT-systems translations concerning this word group. As explained in 5.2.1 '*seu*' can be ambiguous in Brazilian Portuguese.
- The challenge of accurately translating the idiom persists across these chatbot-generated translations, highlighting the ongoing complexity of capturing the subtleties of idiomatic expressions.

³ The translation performed by Bard is a longer text. See Figures 6 and 7.

Thus far, we have encountered various problems, such as interpreting expressions or grammatical errors. However, the capabilities of Generative AI (GenAI) tools extend far beyond mere response generation, as illustrated in the preceding figures. As announced at the beginning of the analysis, we will now present the illustrations of the third chatbot (Bard) in Figures 6 and 7.

Figure 6: Prompt and translation of the dialogue from English to Portuguese by Bard (1)

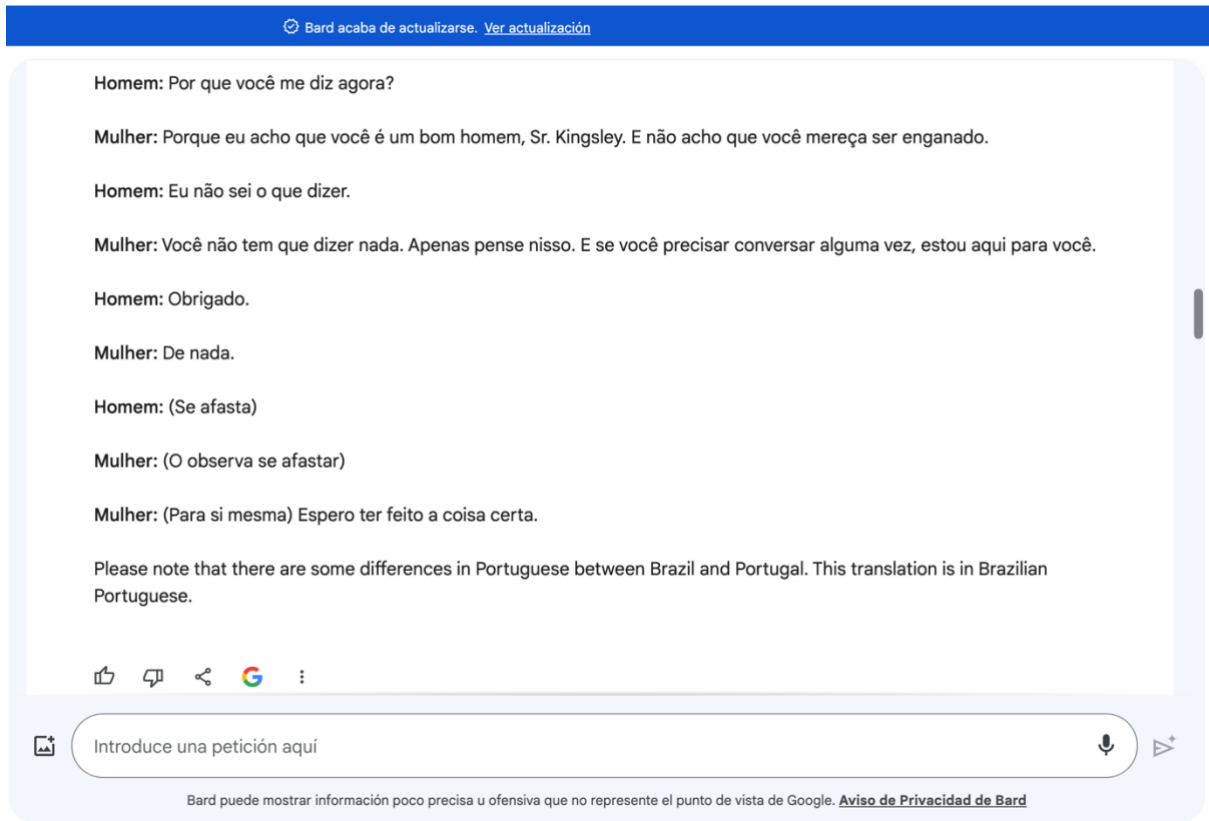


In this context, the chatbot Bard goes beyond the user's explicit request and provides additional, unsolicited information:

- It assigns gender to the two speakers (man-woman).
- It introduces several additional sentences into the dialogue.

This surprising response reveals a fundamental distinction between Machine Translation (MT) and Generative AI. While MT primarily focuses on translation tasks, GenAI surpasses this limitation by translating and augmenting the dialogue. While additional information can be beneficial, its value diminishes if it introduces inaccuracies, as exemplified in this instance.

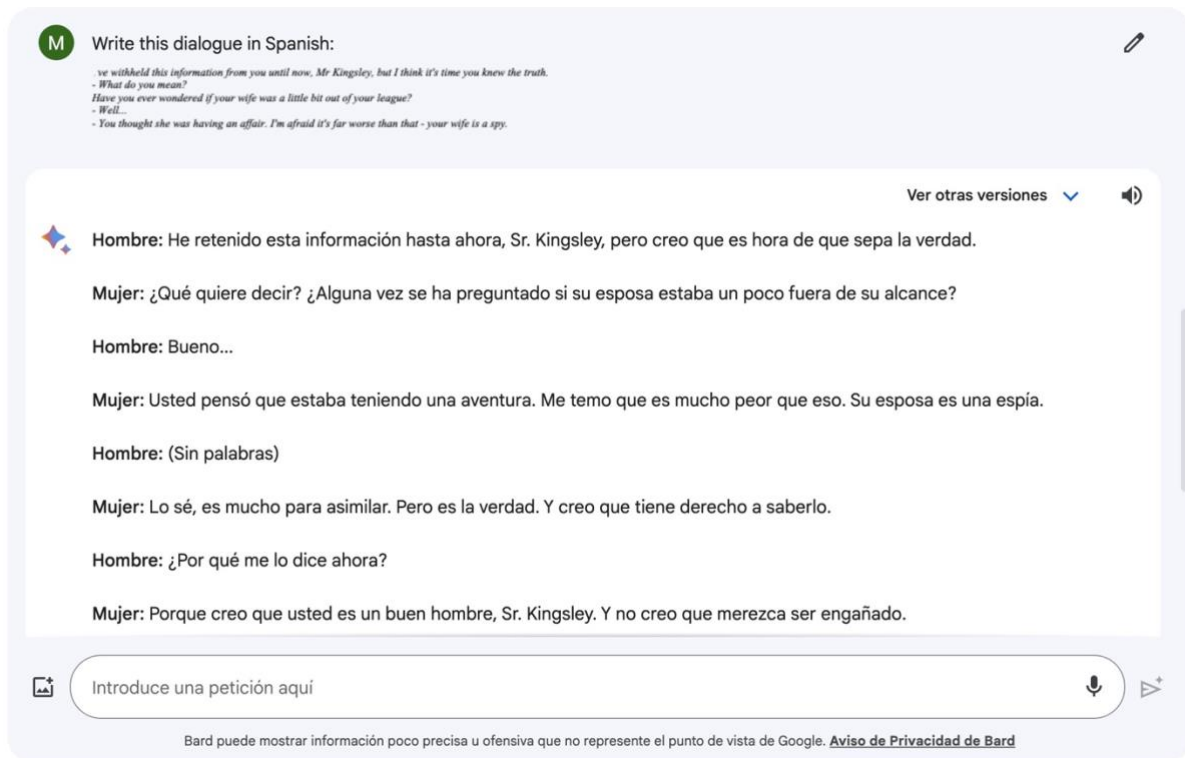
Figure 7: Prompt and translation of the dialogue from English to Spanish by Bard (2)



The chatbot seamlessly extends the dialogue, as evident in the ongoing conversation. This phenomenon is also observed in the English-to-Spanish translation, as shown in Figure 8.

This behavior illustrates a distinctive characteristic of the chatbot's functionality—it responds to user prompts and actively contributes to the conversation by generating additional content. This dynamic engagement surpasses the conventional role of a translator, further exemplifying the multifaceted capabilities inherent in Generative AI. This highlights the nuanced nature of the capabilities and potential pitfalls associated with Generative AI.

Figure 8: Prompt and translation fragment of the dialogue from English to Spanish by Bard



6. CONCLUSIONS

The qualitative analysis shows that there can be significant differences between the three systems and the chatbots themselves. Two general observations, however, can be made when comparing the results of MT-system translations with those generated by chatbots:

- The translations performed by MT systems are of slightly better quality than those performed by chatbots, which contain more errors. In addition to inconsistencies in the use of pronouns in both MT systems and chatbots, the latter performed some literal translations that needed to be present in the output of the MT systems.
- The MT systems strictly perform the translation, whereas GenAI tools, as demonstrated in the examples of 5.2.2 and illustrated by figures 6, 7, and 8, do not always follow suit. One of the GenAI chatbots surpasses the translation task by not only translating but also generating a continuation of the dialogue.

Regarding the target languages, comparable problems were identified in both cases. Four systems faced difficulties in maintaining the correct usage of formal and informal Spanish forms (*usted* and *tú*). Similar inconsistencies were observed in Portuguese translations. Five tools used the formal 'Sr. Kingsley' and the informal 'você' in Brazilian Portuguese in the same dialogue.

Additionally, all tools struggle with translating the idiom. The nature of the error varies from a literal and completely incorrect translation (e.g. *¿Alguna vez te has preguntado si tu esposa estaba un poco fuera de tu liga?*) to the inaccurate translations: "*su esposa está fuera de su alcance*" and "*sua esposa estava um pouco fora de/do seu alcance*". The problem is that the translation "your wife is a little bit out of your reach." This could be related, for example, to the fact that the woman belongs to a higher level of society than the man, which is not the case. Mr. Kingsley's problem is that his wife is doing something he is unaware of. Humorously, the idiomatic expression "your wife is out of your league" might suggest that Mr. Kingsley's wife is much smarter, more intelligent, etc. than he is. The machine and chatbot translations don't reflect these meanings at all. In addition, the humor of the English idiom is wholly lost when translated into Spanish and Portuguese. Metaphorical and idiomatic expressions are still a challenge for MT systems and GenAI, since none of the chatbots effectively capture the intended meaning of the metaphorical expression.

Finally, I would like to highlight one last issue, which also arises from the analysis's results: How reliable are the responses produced by new chatbots when the target language is unfamiliar? This query highlights a significant aspect of their performance, especially when the user is not proficient in the translated language. If the person performing the prompt/query is not skilled in the target language, it is more complex to ascertain whether the chatbot response is correct.

Currently, the systems can translate large amounts of text relatively efficiently, but the output quality could be better. On the one hand, they still make mistakes at various levels, as the analysis has shown. On the other hand, they fail to translate certain subtleties of language accurately and adequately, such as metaphors and idioms. Therefore, there is still a need for human translators. In this instance, only a professional translator possesses the skills, training, and creativity indispensable for a correct, high-quality translation.

In conclusion, it is undeniable that machines are superior in quantity, surpassing the speed of human translators, and that translations of significant texts can be obtained in seconds. This can significantly help, mainly when translating large amounts of text. However, as the results of this study show, the human factor is still necessary to ensure the correctness and accuracy of a quality translation. As a final thought on the conclusion and before setting out the prospects for future research, I would like to quote Chomsky's (2023) words on the topic:

The human mind is not, like ChatGPT and its ilk, a lumbering statistical engine for pattern matching, gorging on hundreds of terabytes of data and extrapolating the most likely conversational response or most probable answer to a scientific question. On the contrary, the human mind is a surprisingly efficient and even elegant system that operates with small amounts of information; it seeks not to infer brute correlations among data points but to create explanations.

Finally, it is essential to remember that this is a qualitative approach. A more comprehensive quantitative analysis would be required to gain complete insight into the effectiveness of different systems. However, there is an important factor to consider in this context: time. Research in this area is time-sensitive due to the rapid advances in GenAI technology. New versions of these tools are being released regularly in real-time, so they have the potential to

surpass their current quality soon. Further research in this area is needed to explore the critical use of machine translation and GenAI and to understand how users can critically engage with these technologies while recognizing their strengths and limitations.

7. REFERENCES

- Alawida, Moatsum, Mejri, Sami, Mehmood, Abid, Chikhaoui, Belkacem & Abiodun, Oludare Isaac. 2023. A Comprehensive Study of ChatGPT: Advancements, Limitations, and Ethical Considerations in Natural Language Processing and Cybersecurity, *Information* 14(8), 462, 1-23.
- Baker, Mona. 1992. *In other words – A Coursebook of Translation*. Routledge.
- Boase-Beier, Jean, 2006. *Stylistic Approaches to Translation*. St Jerome Publishing.
- Conca, Maria & Guia, Josep. 2014. *La fraseologia. Principis, mètode i aplicacions*. Valencia: Bromera /Institut Interuniversitari de Filologia Valenciana.
- Corpas, Gloria. 2000. Acerca de la (in)traducibilidad de la fraseología. En G. Corpas (ed.), *Las lenguas de Europa: Estudios de fraseología, fraseografía y traducción* (pp. 483-522). Comares.
- Chomsky, Noam, Roberts, Ian Roberts & Watumull, Jeffrey (2023). Noam Chomsky: The false promise of ChatGPT. New York Times.
- Díaz Cintas, Jaume & Remael, Aline. 2014. *Audiovisual Translation: Subtitling*. Routledge.
- Díaz Cintas, Jaume y Remael, Aline. 2020. *Subtitling. Concepts and Practices*. Routledge.
- Dwyer, Tessa. 2017. *Speaking in Subtitles: Revaluing Screen Translation*. Edinburgh University Press.
- Gibbs, Raymond W., Nayak, N. P. & Cutting, C. 1989. How to kick the Bucket and Not Decompose: Analyzability and Idiom Processing. *Journal of Memory and Language* 28, 576–593.
- Gibbs, Raymond W., & O'Brien, Jennifer E. 1990. Idioms and mental imagery: The metaphorical motivation for idiomatic meaning. *Cognition* 36, 35–68.
- Gibbs, Raymond W., Bogdanovich, J. M., Sykes, J. R., & Barr, D. J. 1997. Metaphor in Idiom Comprehension. *Journal of Memory and Language* 37, 141–154.
- Gibbs, Raymond W. 2001. Proverbial themes we live by. *Poetics* 29, 167-188.
- Huang, Jin-Xia, Lee, Kyung-Soon, Kim, Young-Kil. 2020. Hybrid Translation with Classification: Revisiting, Rule-Based and Neural Machine Translation. *Electronics*, 9 (2), 201-218.
- Hellmich, Emily. A., & Vinall, Kimberly (eds). 2022. Machine Translation & Language Education: Implications for Theory, Research, & Practice. *L2 Journal*.

- Koponen, Maarit, Salmi, Leena & Nikulin, Markku. 2019. A product and process analysis of post-editor corrections on neural, statistical, and rule-based machine-translation output. *Machine Translation* 33, 61–90.
- Macketanz, Vivien, Avramidis, Eleftherios, Burchardt, Aljoscha, Helcl, Jindrich & Srivastava, Ankit. 2017. Machine Translation: Phrase-Based, Rule-Based and Neural Approaches with Linguistic Evaluation, *Cybernetics and Information Technologies* 17-2, 28-43.
- Maitland, Sara. 2017. *What is cultural translation?* Bloomshory.
- Kern, Rick. 2024. Twenty-first century technologies and language education: Charting a path forward. *The Modern Language Journal* (in preparation).
- Labarta Postigo, María. 2020. A metaphorical map of subtitling. Idiom vs. explicit meaning in translated filmic texts. *Babel* 66 (1), 46-69.
- Labarta Postigo, María. 2021. ¿Se te ha ido la olla? o ¿você é maluco? vs. are you (fucking) nuts? La traducción de locuciones del inglés en series de TV por internet. *Lengua y Habla* 25, 319-343.
- Labarta Postigo, María. 2022. El reto de las metáforas deportivas: Traducción automática vs. Traducción humana. *Lengua y Habla* 26: 242–269.
- Labarta Postigo, María. 2023. ‘Life is Baseball’ vs. ‘A vida é futebol’: The Translation of Sport-Specific Metaphors from English to European and Brazilian Portuguese. *Sendebarr*, 34, 1-15.
- Lakoff, George & Johnson, Mark. 1980. *Metaphors We Live By*. University of Chicago Press.
- Lakoff, George. 1993. The Contemporary Theory of Metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (2nd ed.) 202-251. Cambridge University Press.
- Newmark, Peter. 1988. *A textbook of translation*. Prentice Hall International.
- Poibeau, Thierry. 2017. *Machine Translation*. The MIT Press.
- Shah, Priten. 2023. *AI and the Future of Education: Teaching in the Age of Artificial Intelligence*. Jossey-Bass.
- Trim, Richard. 2007. *Metaphor Networks. The Comparative Evolution of Figurative Language*. Palgrave Macmillan.