Universidad de Valladolid

FACULTAD de FILOSOFÍA Y LETRAS
DEPARTAMENTO de FILOLOGÍA INGLESA
Grado en Estudios Ingleses

TRABAJO FIN DE GRADO

An analysis on the effectiveness of Machine Translation Tools for the translation of medical texts from English to Spanish

Miguel Díaz García de las Bayonas

Tutor: María Belén López Arroyo

2021-2022

ABSTRACT

Nowadays, translation has an important role in society, since it allows people to progress in knowledge in any field that may arise. For this reason, translation has become very usual and necessary. The objective of this project is to analyze MT systems in order to discover how reliable they are and to detect the main reasons why these systems may cause problems. Moreover, the result of this project solves possible doubts among translators and raises awareness about the use of these MT systems. Finally, readers will be able to observe the importance of following some procedures when dealing with a translation project and the relevance of facilitating a post-edition version to have a better understanding of the suggestions provided to improve the quality of the translation. It can also be useful for those people who do not really control the language used and want to seek information about this disease.

Keywords: machine translation, *Google Translator*, *DeepL*, *medicine*, *scleroderma*, scientific language.

RESUMEN

Hoy en día la traducción es un aspecto muy importante en la sociedad, ya que permite progresar en conocimientos en cualquier ámbito que se presente. Por este motivo, la traducción es una tarea de gran importancia. El principal objetivo de este trabajo es analizar los sistemas de TA con el fin de descubrir cuan fiables son y detectar los principales motivos por los que estos sistemas pueden provocar problemas. Además, el resultado de este proyecto resuelve posibles dudas entre los traductores y da conciencia sobre el uso de estos sistemas de traducción automática. Finalmente, los lectores podrán observar la vitalidad de seguir unos procedimientos a la hora de afrontar una traducción y la importancia de facilitar una posedición para que se comprendan con mayor facilidad las sugerencias facilitadas para perfeccionar la traducción. Además, puede también servir a aquellas personas que desconozcan el idioma y deseen buscar información sobre esta enfermedad.

Palabras clave: traducción automática, *Traductor de Google*, *DeepL*, *medicina*, *esclerosis sistémica*, lenguaje científico.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Translation can be found everywhere from your favorite movie to a web page. Translation is essential in professional and daily lives. Without translation there wouldbe a lack of knowledge about new studies and technological advances as well as a language barrier among people. Therefore, the main role of translation in such a changing world is to help to globalize knowledge among different cultures, and facilitate the access to literature and educational resources in different languages.

In the last few decades, translation has become a cornerstone in the fields of healthcare and medical studies. The translation of medical texts has helped not only to disseminate knowledge about different diseases all over the world, but also to develop new save-life treatments and improve those that already exist. The recent interest in medical translation along with the advances in language technologies has made possible the evolution of Machine Translation (MT) systems in this research field. MT systems are capable of providing automatic and precise translations faster than human beings can do, thus accelerating the translation process. However, though useful and fast, MT systems are not entirely reliable, so they often require human post-editing.

This project is necessary since it will make us aware of the need of post-editing the translations provided by MT systems since they do not produce high-quality translations. This project also gives importance to translators since thanks to these kinds of projects their work will be valued in a very positive way. This implies the thought that MT systems will replace the job of translators will disappear in the readers of this project. Finally, there is another clear need for this project since a good Spanish quality translation of a particularity (*red flags*) of a disease that is not very well-known today will be created.

The project is organized into three different sections: The first part consists of the reinforcement of my knowledge about translation with fundamental theoretical aspects of MT. The second part deals with the analysis and comparison of MT systems (*Google Translator* and *DeepL*) with the objective of translating a medical text about *systemic sclerosis*. To simplify the analysis of this project, the aspects to improve belonging to lack of adequacy found in the translations will be classified into two different categories: *accuracy* (*omission*, *addition*, *untranslated elements*, *mistranslation* and *terminology*), and *fluency* (*spelling*, *typography*, *grammar*, and *unintelligible*). This classification is based on the use of the Multidimensional Quality Metrics (MQM) from the German Research Center for Artificial Intelligence (DFKI, 2015) which is currently the most important center of Artificial Intelligence (Burchardt, 2017). Finally, the third section of this project will deal with a post-edition version of the text where all the changes made can be seen, thus achieving a high-quality translation. The post-edition version will also help the reader understand the text and appreciate the importance of the post-edition, since without it the text would lose a lot of quality.

1.1 Purpose

The general aim of this paper is to give an account of the efficiency and accuracy of two of the most important MT systems nowadays (*Google Translator* and *DeepL*), by translating a medical text about *systemic sclerosis*, (also called *scleroderma*), from English into Spanish. These two MT systems has been chosen because according to Blasco (2018), *Google Translator* is the most popular MT system in the world; and it has been compared with *DeepL* since according to some professional translators such as it is said in Mego (2019, p.27), *DeepL* is a MT system with a superior quality than *Google Translator*. This may lead some readers to want to check if this statement is true. So in this project, we will find out if this statement can be considered true with certainty by using a specific *terminology* text from a specific field, *medicine*. The main reason why we have chosen this text is because it combines medical and technological aspects which may pose important problems to MT systems due to the specificity of these fields. The present study includes several steps that may be useful for future translation projects in general, and for those oriented to MT in particular. Furthermore, there are other objectives such as discovering which MT system is more precise between *Google Translator* and *DeepL* in a specific text by classifying the aspects to improve made into subcategories belonging to two different categories: *accuracy* and *fluency*.

This general aim is specified in the following objectives:

1. Detect the main aspects to improve in terms of *accuracy* and *fluency* that both systems commit while translating.
2. Find out if *DeepL* is a MT system with a superior quality than *Google Translator*.
3. Conclude final aspects comparing the results obtained from both systems.

2. THEORETICAL BACKGROUND

2.1 Definition of machine translation

According to Kumar and Kumar (2013, p.318), “machine translation is a process to generate words automatically from one language to another language”. In other words, MT consists of translating a text with no human intervention. Currently, there are many different MT systems to which any user can easily get free access. The most popular ones are *Google Translator*, *DeepL*, *Reverso*, *Systran*, *Bing Translator*, *Babylon*, etc. In this project we will use *Google Translator* and *DeepL* since as we have said before, *Google Translator* is the most popular MT system and *DeepL* is considered a MT system with a better-quality translation according to Mego (2019, p.27).

It should be noted that the MT tools have been very controversial in recent years since these tools can be considered as a substitute for human translators by some users. Due to the fast progress in technology, human translators are afraid of being replaced by translation tools. However, another reason that has created concern is the poor translation quality that these tools offer us.

At the beginning, MT programs “were based on linguistic rules that were used to parse the source sentence and create the intermediate representation, from which the target language sentence was created” (Sepesy and Donaj, 2019, p.2).

2.2 Machine translation approaches

MT approaches are the criteria by which MT systems are designed. MT approaches are usually organized according to the level of attention that should be given to aspects of syntax, morphology and semantics of the source and target text (Hettige and Karunananda, 2011).

According to Sepesy and Donaj (2019, pp 5-6), there are three different MT approaches: *rule-based approach*, *corpus-based approach* and *Neural MT approach*.

1. *Rule-based approach.* This is a system that consists of a huge collection of semantic, syntactic and morphological rules, which is developed manually every so often by expert translators, representing structures from the source language to develop their translation to the target language. This system has an expensive cost and it requires a lot of time to maintain and run (Sepesy & Donaj, 2019, p.2).

According to Sepesy and Donaj (2019, p.2), *rule-based approach* includes three different types of MT: *dictionary-based MT*, *transfer-based MT* and *interlingual MT*.

1.1 Dictionary-based MT. Regarding to Shalini and Hettige (2017, p.24), this system "translates source language to target language by using word-to-word or phrase-to-phrase mapping". This means, as it is explained in Sepesy and Donaj (2019, p.3) that this type of *rule-based MT* uses entries in a bilingual dictionary of any language with the aim of finding equivalent words in the target language. However, this system creates significant problems since this system cannot solve ambiguity problems. Also, this system does not perform a linguistic analysis of the source language text before translating it, which creates low-quality translations. This system had a very slow processing time. Furthermore, it was designed to translate between two related languages. Some of the early MT systems are *Meteo*, the old *Systran* and *Weidner* (Jurafsky & Martin, 2000, p.813).

1.2 Transfer-based MT. This system analyzes the text of the source language in a syntactic, semantic and morphological way to establish its grammatical structure. In this way the system leaves a structure known as "intermediate". Then, the structure is processed into an appropriate structure in order to translate the text into the target language (Sepesy and Donaj 2019, p.3).

1.3 Interlingual MT. This system uses an intermediate language to translate the source language text, that is, it uses a language as a "bridge" that has the syntactic, morphological and semantic characteristics of the target language. This system is very economical. However, it has important drawbacks since it is difficult to find a suitable interlingual language. In addition, by using two different translations instead of making a direct one, the system loses information that makes the quality of the translation lower (Sepesy and Donaj, 2019, p.3).

2. *Corpus-based approach.* This method uses linguistic information from different corpora with the objective of creating new translations. This approach is divided into two different types*: example-based MT* and *statistical MT* (Quah, 2006, p.76).

2.1 Example-based MT. This type of MT belongs to corpus-based approaches because in this type of approach, the main source to carry out the translations is the use of examples of corpora from large collections of bilingual corpora (Sepesy and Donaj, 2019, p.3).

2.2 Statistical MT. Unlike *example-based MT*, this type of MT uses translation models whose parameters depend on the analysis of bilingual corpora composed of original text and their respective translations through the application of algorithms. This approach can provide many translations for the same source language sentence. Nevertheless, this system is expensive and problematic when it comes to storage. However, if sufficient bilingual data are available, it can easily be adapted to a specific domain. It does not require linguistic knowledge (Quah, 2006, p.77).

3. *Neural MT approach.* This system is an approach to MT in which a computer uses deep learning to create an artificial neural network and thus it teaches itself to translate between different languages. This MT system arose from the statistical MT, but the main difference is that this type of approach has neural networks that work by relating words from the same

semantic field in order to provide translations. For example, the words *tiger* and *lion* are more related between them than the words *tiger* and *chair* (Molina, 2019, p.11). It should be noted that it has improved greatly in recent years. In addition, it is actively entering the translation industry. However, there are some problems in expanding the semantic field. The training speed of the models should also be improved (Sepesy and Donaj, 2019, p.5).

2.3 Machine translation and its evaluation

Nowadays, MT systems are tools that are becoming quite popular in the workplace. In fact, Arevalillo (2012, p.181) assures that many companies are increasingly introducing this type of tools to expedite and facilitate their work as much as possible. However, the translations provided by these programs, though comprehensible, are far from being perfect. As a result, they tend to require human reviews. Thus, companies also need to be awareof the importance of a good reviewer and/or translator who performs post-editing tasks to achieve high quality translations.

Post-editing means having the need for a human being to review the translated text obtained from the automatic translation program and thus improve or adjust it to the wished result. According to Sepesy and Donaj, (2019, p.15) 70% of translation machines use post-editing and also ensure that it is faster than assisted translation through the use of translation memories. In addition, both these authors and Mendoza (2017, p.185) state that there are 3 post-editing effort methods that greatly influence the quality of translation.

- *Temporal effort*. This effort refers to the time the translator spends post-editing a text. This effort is easy to measure quantitatively since we can know how long the post-editor takes to complete the task.
- *Cognitive effort*. This effort refers to the mental efforts made by the translator. This effort occurs when cognitive processes are shown during the post-editing process. However, this effort cannot be measured directly since a person's mental effort cannot be calculated. However, Krings (1986,

p.137) used a method called *Think Aloud Protocol* (TAP) in which post-editors were recorded and expressed what they thought while performing the post-editing process. Nervertheless, it should be clarified that speaking out loud can slow down the cognitive process and there is no guarantee to ensure that what is expressed by the post-editor is an accurate representation of the cognitive process of the post-editor (Jakobsen, 1998, p.82).

- *Technical effort*. This type of process is by which corrections are made. In this effort the post-editor can make insertions and changes in the punctuation. Moreover, the post-editor will reorder and delete words in order to get a better translation.

Hence, it seems that post-edition is not the simple correction of a text, but a whole process that requires great efforts and hard work on the part of the translator/editor. In simple words, this means that translators and editors need to spend a lot of time evaluating the quality of the translation obtained by the MT system, to finally provide the best version of the original text in the target language. As a way of helping professionals with this exhausting and tedious work, recently some tools, such as *BLEU* (Bilingual Evaluation Understudy), have been created to analyze and evaluate the quality of MT translations automatically. Automatic evaluation tools judge the output of the MT systems by contrasting it to human reference translations. For example, according to Papineni et. al (2002., p.5), *BLEU* scores the quality of a translation on a scale from 1 to 0, being 1 a perfect match with the referent translation and 0 a complete mismatch with the referent translation. This type of automatic evaluation tools is generally free and quite fast, thus drawing the attention of many companies whose main concerns are that of saving time and reducing costs. However, it is important to consider that even though automatic evaluation tools are quite practical, they are still more ineffective than the work performed by professional translators because these tools can produce more structures that can be improved in terms of *accuracy* and *fluency* more easily than professional translators since they do not have all the means that a professional translator can obtain to evaluate the quality of a translation. In addition, these tools cannot be objective as a human professional translator depending on the translation to be carried out, since these tools follow specific rules that are used in the same way in all translations. Nevertheless, translating a medical text is not the same as translating a literary text for example, since different steps must be used.

Anyway, the conclusion reached at this point is that translations produced by MT systems are not perfect and therefore, they always need to be reviewed before being published, either by a professional editor/translator or by an automatic evaluation tool. This fact makes us wonder whether MT programs are sufficiently capable of producing quality translations andmeet the needs of their users or not. For this reason, and as previously said, the present paper aims at analyzing the efficiency of MT systems by translating a medical text on *systemic sclerosis* in *Google Translator* and *DeepL,* i.e., two of the most influential and worldwide known MT systems nowadays. In the following sections we will explain the procedure that has been followed and the aspects we have taken into account to evaluate the quality of each MT system.

2.4 Post-editing

In this section we try to explain what post-edition is and what it consists of, since its use is fundamentally important in MT due to the fact that these systems are not perfect.

The post-editing process consists of editing the translated text that a MT system provides us with the goal of getting a higher quality translation. This edition can include the correction of semantic and linguistic aspects such as those of *grammar*, *spelling*, *terminology*, *typography*, etc. Post-editing can be applied in different measures. On the one hand we find the light post-editing, which consists of performing a partial edition of the text, making it understandable enough. Among the changes that are made in this type of post-edition we find changes in *terminology* or *grammatical* aspects, but these aspects can be allowed while the message of the original text can be understood by the reader. On the other hand, we have full post-editing, whose purpose is to achieve a high-level translation, thus eliminating all the failures that MT offer us. In other words, full post- edition seeks that the text translated gets a native level, in which no terminological or grammatical aspects to improve can be found. However, nowadays it is difficult to categorize the quality level of a translation, since each client has their own criteria, although it is possible to differentiate between a good translation and a bad one. Today, there are companies such as *Systran* or *SDL* that offer courses to help translators make better post-editions (Aranberri, 2014, p.473). Depending on the quality of the MT system translators, it will be necessary to post-edit more or less; if the MT system is usually very good and gives us a high-quality translation, it will only be necessary to adapt the format but there will not be much more work.

The possibility that MT systems end up extinguish translators' work in the future is something that is questioned nowadays by many people. Serrano, (2015, p.12) says that MT has been growing for 50 years, but even so, that moment will never come, as professional translators are needed to adapt the translations produced by these MT systems. However, this aspect can also be solved in this project, since two well-known MT systems will be analyzed, *Google Translator* and *DeepL*. The translated texts will be analyzed and thus it will be checked whether they provide an acceptable translation in a specific field.

2.5 Machine translation systems

There are many MT programs in the world. In this project we will focus on comparing *Google Translator* and *DeepL* due to its popularity. However, there are also other MT programs that are very interesting to mention such as *Reverso*, *Systran, Babylon*, *Bing Translator*, since they have the same functions and they are also free. These programs are capable of automatically translating texts into different languages. In addition, these programs allow users to know the pronunciation of the transcribed text.

2.5.1 Translation systems to be compared: *Google Translator* and *DeepL*

Google Translator and *DeepL* are two of the most well-known translators. The function of both is simple: the MT systems automatically translate the text that has been introduced into practically any language. *Google Translator* is able to translate texts into 103 languages, while the *DeepL* translator just into 9 languages. It should be clarified that the number of languages does not mean that it has a better translation quality. This can be proved in Mego, (2019, p.27), since using the test bench called *BLEU*, which as mentioned before measures the quality of translation, professional translators seem to prefer *DeepL* results over *Google Translator* ones. In this project, the translation made by these two MT systems will be also analyzed to prove which system is more accurate in a particular field (*medicine*). On the other hand, both *Google* and *DeepL* translators can include a maximum of 5,000 characters. It is also important to mention that both services are free and let the user hear the pronunciation of the translation. Moreover, *Google* claims that its translator contains more than a billion words and that for European languages *Google Translator* uses a corpus made up of debate texts from the European Parliament (Azou, 2016, p.16). For these reasons it is interesting to check the reliability of these MT systems by comparing their possible types of aspects to improve and thus see which one creates the best translation of a scientific article.

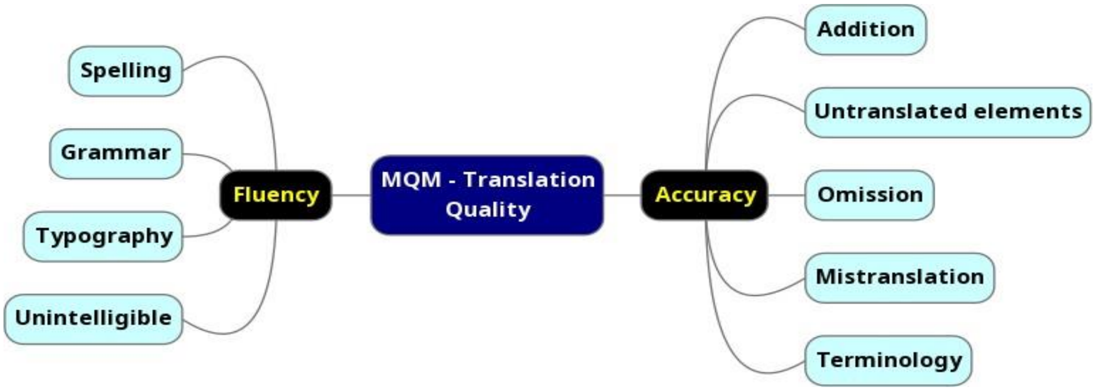
3. METHODOLOGY

In order to carry out this project, several fundamental steps have been followed. Firstly, documentation and a previous organization with a supervisor were needed on how to do this kind of project. Then, we proceeded to search for information from MT and *systemic sclerosis* to improve and increase the knowledge in these fields. We found several helpful parallel texts about *scleroderma*, such as the one from Oliveró (2013): *Esclerosis sistémica: supervivencia y factores pronósticos*. In addition, help has been required from a medicine graduated doctor from the University of Valladolid and from a veterinary student who suffers from this disease. Finally, the use of some online monolingual and bilingual dictionaries such as the *Real Academia Española* (*RAE*) (2001), the *Diccionario Panhispánico de Dudas* (2005), the *Merriam-Webster* dictionary (2011) or the *Diccionario crítico de dudas ingles-español de medicina* by Navarro (2005), medicine dictionaries such as *MedlinePlus* (2019) and Spanish corpora databases like *CREA* and *CORPES* from the *RAE* (2020) have been very useful to post-edit some specific terms.

Moreover, it will be made a contrast among the number of aspects to improve made by both MT systems, both in a general and specific way to appreciate if a MT system is better than the other in general terms, paying attention to the different categories and subcategories. This is interesting because in general terms a MT system could be "better" for the fact of presenting a lower number of aspects to improve, but these aspects might be more problematic.

As we have mentioned before we will follow the MQM rules from the DFKI (2015). In the following figure (1), we can appreciate an adapted scheme based on these rules. This figure deals with the quality criteria that the MT must carry out. These patterns provide a clear definition of what quality control is. Moreover, this project offers a detailed categorization list, showing the definition and some examples of each type. The quality of MT is divided into these two different categories:

Figure 1. MQM Translation Quality scheme adapted from Lommel, A., Burchardt, A. & Uszkoreit, H (2013, p.4).



As we can see in this figure (1), the classification will be divided into two different categories: *fluency* and *accuracy*. Within the part of *fluency,* we will find those of *spelling*, *grammar*, *typography* and *unintelligible* aspects, while within the *accuracy* category we will find those of *addition*, *untranslated elements*, *omission*, *mistranslation* and *terminology*.

The translation process to follow is to analyze each type of aspect to improve as it is explained in the MQM from the DFKI (2015) and to classify them in different tables showing the original text, the result that *Google Translator* and *DeepL* versions produced of every aspect in question and the corresponding part of the post-edition version. In addition, the different examples shown in the tables will be explained.

Finally, once the text is translated, we will facilitate a post-edited text to improve the translation and thus understand the adequacy aspects to improve, committed by these programs, if any. Then, the results of the MT systems used will be compared to evaluate which program is the most reliable from an objective point of view. Finally, in the last section a brief conclusion on the translation procedure and the problems encountered mainly with *terminology* will be developed. In addition, we have also included in the CD a zip file with the original text and a personal post-edited version.

4. ANALYSIS AND COMPARISON BETWEEN *GOOGLE TRANSLATOR* AND *DEEPL*

4.1 Comparison between *Google Translator* and *DeepL*

In this section the aspects to improve made by *Google Translator* and *DeepL* are going to be classified into two categories: *accuracy* and *fluency*. The aspects that can be improved from the translations provided by these two MT systems will be collected from a scientific article by Li, Sahhar and Littlejohn. These authors are part of the Rheumatology Department from Melbourne. The article deals with *red flags* in *scleroderma*. The reason why this article has been chosen is because we wanted to use a specific text from such an interesting branch as *medicine*. This article is really interesting as it defines what is *scleroderma*, how to recognize it, possible complications and treatment, as well as the provision of images and examples of some of the signs mentioned in the article. This section is divided into two different parts: in the first one, the aspects to improve which belong to the *accuracy* category are going to be analyzed and compared, while the second part of the section will deal with a classification based on the aspects to improve from the *fluency* category. In order to make this classification, several tables have been made to show examples of each type of aspect within each category.

4.2 Classification of the aspects to improve

In this part of the project, we are going to give account of some aspects that have drawn my attention from the translation made by *Google Translator* and *DeepL*. In order to obtain the expected result, those aspects that can be improved will be classified into two different categories: *accuracy* and *fluency*. These aspects will be chosen from the translation provided by *Google Translator* and then compared with the same translation from *DeepL*. The aspects to improve will be classified into different tables according to their category. These aspects can be interpreted as recommendations for these two MT or as precautions tothose translators who use MT systems.

4.2.1 *Accuracy* classification

The aspects to improve belonging to the *accuracy* classification will be sorted in five different categories: *omission*, *addition*, *untranslated elements*, *mistranslation* and *terminology*. Some of these aspects are going to be classified in tables comparing the original text, the translation provided by these two MT systems and the post-edition version. Then, we will provide an explanation of each example in each table.

In the following table (1) we will deal with *omission* aspects to improve. We find this type of aspects when an MT system omits necessary words in the translation of a text. Below, we can see some examples of *omission* aspects to improve made by these MT systems:

Table 1. ***Accuracy* classification: *Omission***

MQM – Translation Quality			
Accuracy - Omission			
Original text	***Google Translator***	***DeepL***	**Post-edited version**
“CREST syndrome”	“síndrome CREST”	“síndrome CREST”	“Síndrome ***de*** CREST”

“Significant interstitial lung disease occurs… in limited and diffuse SD”	“La enfermedad pulmonar intersticial significativa ocurre… en SD limitada”	“La enfermedad pulmonar intersticial significativa ocurre… en ***la*** SD limitada”	“La enfermedad pulmonar intersticial significativa ocurre … en ***la*** ES limitada”
“to improve exercise capacity”	“mejoran la capacidad de ejercicio”	“mejoran la capacidad de ejercicio”	“mejoran la capacidad de ***hacer*** ejercicio”

In table (1), aspects to improve related to prepositions and determiners that have been omitted by these two MT systems can be appreciated in the two first examples. The structure “CREST syndrome” has been translated as “síndrome CREST”, but in Spanish this syndrome is called “síndrome *de* Crest” as it can be seen in the medical dictionary of Clínica Universidad de Navarra (2020). In addition, despite the fact that in English there are no determiners when talking about diseases, such as in the example “in limited and diffuse SD”, in Spanish it is required the determiner *la* when we specify the disease in which something *occurs*. To prove this, it has been necessary the use of Spanish original parallel texts to compare the utilization of this type of determiners in Spanish similar texts, such as the article about *systemic sclerosis* from Oliveró (2013), where we can see the structure *la ES limitada* in the same context in pages 48 and 49. The last example deals with a Spanish expression used to exercise. According to the Spanish legal dictionary from the RAE (2020), *capacidad de ejercicio* is the suitability of a person to perform rights and fulfill obligations in a personal way. However, if we want to refer to practicing sport, in Spanish we must add the verb *to do*: *capacidad de hacer ejercicio*. In English we can use “exercise” as a verb but in Spanish it is required the verb *hacer* since without this verb the meaning of the sentence changes.

In the next table (2), we are going to deal with *addition* aspects to improve, it is said, those words added by these MT systems without been needed.

Table 2. ***Accuracy*** **classification:** ***Addition***

MQM – Translation Quality			
Accuracy - Addition			
Original text	***Google Translator***	***DeepL***	**Post-edited version**
“…is an uncommon connective tissue disease characterised by vascular …dysfunction of multiple organ systems”	“… es una enfermedad poco común del tejido conectivo caracterizada por disfunción vascular…”	“es una enfermedad poco común del tejido conectivo que se caracteriza por ***la*** disfunción vascular”	“es una enfermedad poco común del tejido conectivo que se caracteriza por disfunción vascular”
“emotional stress”	“estrés emocional”	“***el*** estrés emocional”	“estrés emocional”
“Characteristic features suggesting its presence include Raynaud phenomenon, skin thickening, calcinosis and telangiectasia”	“Las características que sugieren su presencia incluyen el fenómeno de Raynaud, engrosamiento de la piel, calcinosis y telangectasia”	“Los rasgos característicos que sugieren su presencia incluyen el fenómeno de Raynaud, ***el*** engrosamiento de la piel, ***la*** calcinosis y ***la*** telangectasia”	“Las características que pueden determinar su presencia incluyen fenómeno de Raynaud, engrosamiento de la piel, calcinosis y telangectasia”

As we have seen in this table (2) above, *DeepL* causes several problems with determiners while *Google Translator* has some aspect to improve when dealing with *addition*. In order to make this classification, the use of Spanish parallel texts has been very helpful. In the first example, the determiner has not native-like sound since it is not required in the sentence. As we can observe in an example from Oliveró (2013, p.40), she does not use that determiner when enumerating symptoms or signs of the disease. The second example has to do with a determiner which is not needed in that structure since that example is taken from a subsection within a table. As we can see in the tables from the original text from Li, Sahhar and Littlejohn (2008) and in the tables from the Spanish parallel text from Oliveró (2013, p.22), the use of determiners in these tables is unnecessary since the structures found are part of a list which indicates possible causes, complications, treatments, symptoms or signs of the disease. Regarding the last example, it is not common in scientific articles to add a determiner to every characteristic feature in Spanish medical texts, as we can see in an example from Oliveró (2013, p.40).

In the following table (3) the words that these MT systems have not translated are going to be analyzed and compared.

Table 3. ***Accuracy*** **classification:** ***Untranslated elements***

MQM – Translation Quality			
Accuracy – Untranslated elements			
Original text	***Google Translator***	***DeepL***	**Post-edited version**
“SD”	***“SD”***	***“SD”***	***“ES”***
“PAH”	***“PAH”***	***“PAH”***	***“HAP”***
“ILD”	***“ILD ”***	***“EPI”***	***“EPI”***
“ENA”	***“ENA”***	***“ENA”***	***“ENA”***
“DLCO”	***“DLCO”***	***“DLCO”***	***“DLCO”***

Although there are some acronyms that should not be translated such as *FVC* or *DLCO* since these acronyms are also used in Spanish as it can be proved in the *MedlinePlus* dictionary (2019), there are other such as *SD* (*Scleroderma*), *PAH* (*Pulmonary Arterial Hypertensión*) or *ILD* (*Intersticial Lung Disease*) which require an adaptation and these MT systems did not provide it. The correct acronyms in Spanish are: *ES* (*Esclerosis Sistémica*), *HAP* (*Hipertensión Arterial Pulmonar*) and *EPI* (*Enfermedad Pulmonar Intersticial*), as we can see in the *MedlinePlus* dictionary (2019). However, as we have seen in this table above (3), there is one acronym known by *DeepL* of these examples which has been rightly adapted. In addition, these MT systems have not translated correctly other acronyms such as *ENA* or *DLCO*. However, we cannot affirm that this fact is a sign of success by the MT systems since, as we have seen in the other examples, it may be because these MT systems do not have an equivalent for these acronyms.

In the following table (4) we will see some *mistranslation* aspects to improve, which are those aspects that are poorly translated, not because of the specificity of the text, but rather because they are general terms poorly translated according to the *RAE* (2001).

Table 4. ***Accuracy*** **classification:** ***Mistranslation***

MQM – Translation Quality			
Accuracy – Mistranslation			
Original text	***Google Translator***	***DeepL***	***Post-edited version***
"It is ***now*** the leading cause of..."	"***Ahora*** es la principal causa de…"	"Es ***ahora*** la principal causa de…"	"***Hoy en día*** es la principal causa de".
"Scleroderma renal crisis is a ***rapidly progressive***"	"La crisis renal de esclerodermia es una forma de insuficiencia renal ***rápidamente progresiva***"	"La crisis renal de la esclerodermia es una crisis renal de ***rápida progresión***"	"La crisis renal de la esclerodermia es una crisis renal de ***rápida progresión***"
"This can lead to flexion contractures ***which*** limit hand function"	"Esto puede conducir a contracturas de flexión ***que*** limitan la función de la mano".	"Esto puede llevar a contracturas de flexión ***que*** limitan la función de la mano"	"Esto puede conducir a contracturas de flexión, ***las cuales*** limitan la función de la mano"

As it can be seen, in the first and third cases, there is almost no difference between the translation provided by *Google* and *DeepL* as both of them have poorly translated *now* and *which* as *ahora* and *que*. The words *ahora* and *que* in these two specific cases make the sentence loose quality since in the first example, it is more appropriate in this kind of texts to translate *now* as *hoy en día*, since according to the *RAE* (2001), *ahora* is a specific word of a certain moment (in this specific moment). However, the expression *hoy en día* is a more general expression which covers more time and refers to the times in which we currently live. In the third example, the reason is the same; according to the *RAE* (2001), it is more formal to say *las cuales* in order to add more information to a defining relative clause. Hence, a possible way to increase the quality and comprehension of these two sentences in Spanish can be the use of the expressions *hoy en día* and *las cuales*, respectively. Looking at the second example, it can be seen that there is a difference between the results obtained by both translation tools as, in *Google Translator*, *rapidly progressive* has been translated as *rapidamente progresiva* whereas in *DeepL*, as *de rápida progresión*. In this case, *DeepL* has provided a more adequate and native-like version because in medical texts, according to the *CREA* corpora (2020), it is not common to see that expression in *medicine* texts to refer something which has a fast progression since there are only two results, while the frequency for the expression *rápida progression* is much higher. For this reason, the *DeepL* version has been the one selected for the post-edited text.

In the following table (5) we will deal with one of the most important aspects of health science translation, *terminology*. However, this type of text, as we mentioned before, can present more problems because of its specificity. For this reason, the quality of the translation is crucial in the translation of medical texts. We will show some examples of aspects to improve made by these MT systems due to the specificity of the text:

Table 5. *Accuracy* classification: *Terminology*

MQM – Translation Quality			
Accuracy – Terminology			
Original text	***Google Translator***	***DeepL***	***Post-edited version***

"Iron ***deficiency***"	"***deficiencia*** de hierro"	"***deficiencia*** de hierro".	"***déficit*** de hierro"
"***HRCT chest***"	"***cofre HRCT***"	"***Tórax de HRCT***"	"***TAC de alta resolución pulmonar***"
"***Renal physicians***"	"***renal médicos***"	"***riñones médicos***"	"***nefrólogos***"

As we can see in this table above (5), MT systems present problems when facing to *terminology* of a specific field. *Google Translator* and *DeepL* have translated in a literal way words such as *deficiency*, while in Spanish medical texts it is more accurate to say *déficit*, as we can prove in the Spanish corpora *CORPES* (2020), the frequency for the word *deficiencia* in *health* texts is much fewer than the frequency obtained for the word *déficit* in the same search. In the second example, *HCRT chest* should be translated as *TAC de alta resolución pulmonar*, since that acronym in Spanish needs an adaptation, as it can be proved in the *MedlinePlus* dictionary (2019). In addition, the word *chest* is bad translated in both MT systems. In the third example, we can appreciate another *terminology*aspect to improve; since these MT systems did not recognize in Spanish the structure *renal physicians*. In Spanish, according to Navarro (2005), the word *nefrólogo* is used to refer to those doctors specialized in the kidneys.

4.2.2 *Fluency* classification

In this classification we are going to deal with *spelling*, *grammar*, *typography* and *unintelligible* aspects to improve. *Spelling* aspects to improve are those capitalized/lowercased words or misspelled. The next table (6) shows some examples of this kind of aspects.

Table 6. ***Fluency*** **classification:** ***spelling***

MQM – Translation Quality

Fluency – Spelling			
Original text	***Google Translator***	***DeepL***	***Post-edited version***
“***Skin***”	“***piel***”	“***Piel***”	“***Piel***”
“***Thinning*** of the lips”	“***adelgazamiento*** de los labios”	“***Adelgazamiento*** de los labios”	“***Adelgazamiento*** de los labios”
“***severe*** Raynaud phenomenon”	“fenómeno ***severo*** de Raynaud”	“***Fenómeno*** de Raynaud grave”	“fenómeno ***acusado*** de Raynaud”

As we have seen in this table (6) above, *Google Translator* and *DeepL* have some problems when dealing with short sentences within a longer text or single words which appear out of a context. However, these MT systems do not present in this text misspelled aspects to improve. In order to explain these aspects, the context is going to be displayed: A part of the text classifies the *red flag* signs in a table, divided into sections and subsections, so sections must be capitalized while those subsections belonging to different sections must be classified in lowercase letters according to the *RAE*’s *Diccionario Panhispánico de Dudas* (2005). The two first examples show two cases where the first word must be capitalized since these words belong to sections and the last example show a case where *DeepL* has capitalized a word that must be in lowercase letters because that word belongs to a subsection.

The next table (7) deals with *grammatical* aspects to improve. In this classification table, it can be observed several aspects dealing with *agreement* (those words which do not match with their number, person or case), *tense* (wrong verbal forms), *word order* (incorrect word order in a sentence), *gender* (bad use of masculine or feminine) and *part of speech* (words with an inappropriate category).

Table 7. ***Fluency*** **classification:** ***Grammar***

MQM – Translation Quality			
Fluency – Grammar			
Original	***Google Translator***	***DeepL***	***Post-edited***

text			*version*
“Skin changes ***usually*** involve the hands”	“Los cambios en la piel ***generalmente*** involucran las manos”	“Los cambios en la piel ***suelen*** afectar a las manos”	“***Generalmente***, los cambios se presentan en las manos”
“Markedly ***reduced*** if significant lung, ***cardiac*** or renal disease”	“***Reducido*** notablemente si es importante pulmón, ***cardíaco*** o enfermedad renal”	“Marcadamente ***reducido*** si el pulmón, el corazón o enfermedad renal”	“***Reducida*** si la enfermedad pulmonar, ***cardíaca*** o renal es significativa”

As we have seen in this table (7) above, there are important *grammatical* aspects to improve in MT systems. Most of the *grammatical* aspects to improve made by these two MT systems are related to *word order*, as we can see in the first example. The sentence has been badly translated since in scientific texts it is more frequent to use the word *generalmente* at the beginning of a sentence, as we can see in Oliveró (2013, p.33). The second aspect to improve has to do with a *grammatical* aspect belonging to *grammatical gender*. The problem is that in English, unlike Spanish, there is no grammatical gender, and in Spanish, the word *enfermedad* (the referent of that sentence) is a feminine word, as we can see in the*RAE* (2001). The verb *to reduce* and the adjective *cardiac* were translated as a masculine word while these two words should be translated as feminine since they are referring to a feminine noun.

The next table (8) deals with those *typography* aspects to improve. In this category we will find *punctuation* aspects to improve.

Table 8. ***Fluency*** **classification:** ***typography***

MQM – Translation Quality
Fluency – Typography

Original text	***Google Translator***	***DeepL***	***Post-edited version***
"pallor ***and/or*** cyanosis of fingers"	"palidez ***y / o*** cianosis de dedos"	"palidez ***y/o*** cianosis de los dedos"	"palidez ***y/o*** cianosis de dedos"

In the table (8) above, we find the formula and/or. Although in Spanish this formula is considered unnecessary in many areas since the conjunction *o* has an inclusive and exclusive value, in medical texts we continually find this formula in order to avoid ambiguity problems. As we can prove in Oliveró (2013, p.5, 33, 56-58, 68, 72, 75), this formula is often used throughout the article in order to avoid ambiguities. According to the example of the table (8) above, *Google Translator* has wrongly translated the formula *y/o* since this MT adds a space before and after the symbol /, something that is wrong in both languages. However, this is the only *typography* aspect to improve found in this text. There are no unpaired quote marks or brackets made by these MT systems in the translation of this text.

The next table (9) shows *unintelligible* aspects to improve, in other words, those sentences that have been translated in such a way that are very difficult to understand, either due to several aspects to improve in the same sentence or to significant aspects to improve due to the fact that the *terminology* of a specific field is unknown. Although *unintelligible* aspects to improve have not been very common in the translation that these MT systems have showed us, since in *medicine* literal translations must be made in some cases, we can see an example below:

Table 9. ***Fluency*** **classification:** ***unintelligible***

<table>
<tr><th colspan="4">MQM – Translation Quality</th></tr>
<tr><th colspan="4"><i>Fluency – Unintelligible</i></th></tr>
<tr><th>Original text</th><th><i>Google Translator</i></th><th><i>DeepL</i></th><th><i>Post-edited version</i></th></tr>
</table>

“***fine end expiratory crackles at base***”	“***crujidos espiratorios finos en la base***”	“***finas crepitaciones espiratorias en la base***”	“***estertores espiratorios secos en la base***”

Table (9) above shows us a sentence that must be considered as *unintelligible* since there is no more information and it is difficult to understand the meaning of the sentence without reading the original text and getting familiar with the illness analyzed in this project. The aspects to improve in that sentence may have been caused by the unfamiliarity of the *terminology* of this specific field because both systems have literally translated the words*fine* and *crackles*, while these words in medical terms must be translated as *secos* and *estertores*, as we can see in the *RAE* (2001) for the word *estertor* and in the United States National Library of Medicine, *MedlinePlus* (2019) for the term *secos*. Although there are aspects to improve due to the *terminology*, as we have mentioned before and it is well-explained in the MQM rules from DFKI (2015, p.4), by putting together several aspects to improve, of whatever type, sentences categorized as *unintelligible* can be created.

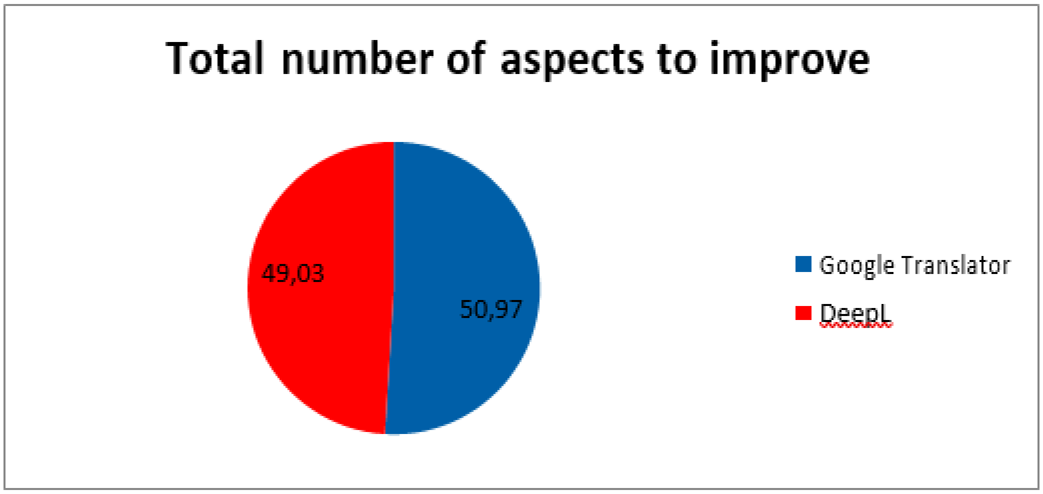
5. RESULTS

As we have proved in the section above, MT programs make several aspects to improve, which do not guarantee great reliability for the user. In this section, a recount of each type of aspects will be carried out in order to see in which aspects one translator is better than the other and appreciate the differences and similarities between the aspects to improve by these MT systems. A comparison between the total aspects to improve of *Google Translator* and *DeepL* is going to be made and finally an analysis of the two different categories of aspects to improve is going to be developed.

5.1 Google Translator vs. *DeepL*

In this subsection it is going to be appreciated which MT system has committed more aspects to improve. To verify which system has produced more structures that should be improved, all the aspects to improve of both MT systems have been counted, no matter the type of aspect produced, so the following figure (2) represents the total number of aspects to be improved between *Google Translator* and *DeepL*.

Figure 2. Total number of aspects to be improved between *Google Translator* and *DeepL*

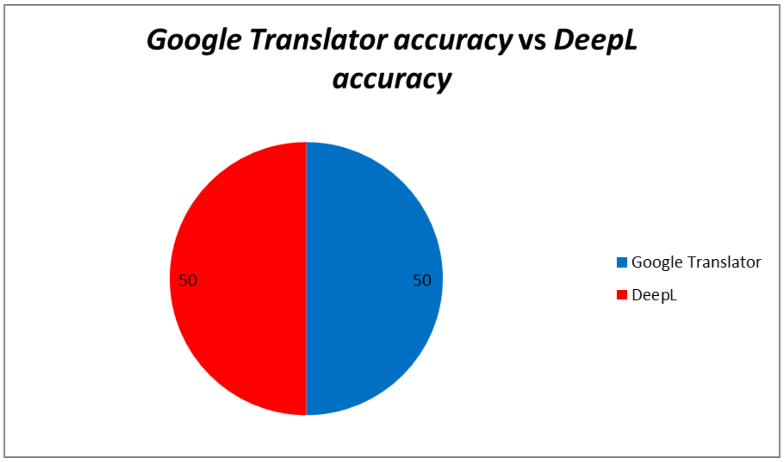


As we can see in the figure (2) above, 50.97% of the aspects to improve have been produced by *Google Translator*, while 49.03% of the aspects to improve are from *DeepL*. This figure (2) shows that as stated in Mego (2019, p.27), *DeepL* presents fewer aspects that can be improved than *Google Translator*. However, as the difference is not representative since the percentages are similar, an analysis will be carried out to see in which category and what kind of aspect to improve each MT system has made in order to see if one MT system makes more relevant aspects to improve than the other.

5.2 *Accuracy* aspects to improve in both MT systems

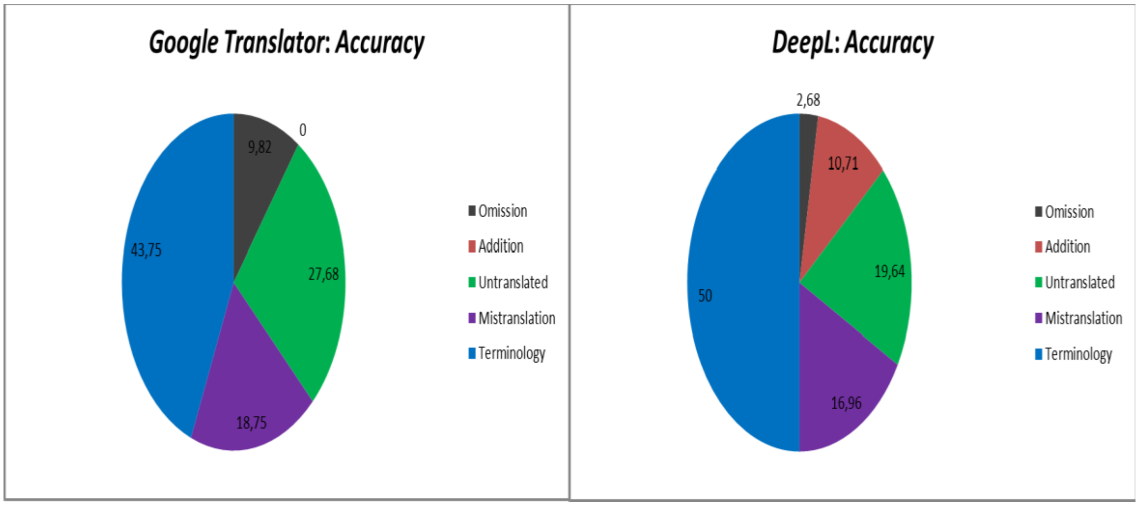
In the following figure (3) there will be a comparison of the total number of aspects to improve in the *accuracy* category by both MT systems to check which MT system has made more aspects to improve in the same category:

Figure 3. ***Google Translator accuracy*** **vs.** ***DeepL accuracy***



As it can be appreciated in figure (3), *Google Translator* and *DeepL* have presented the same number of aspects to improve, so the percentage is the same. To see if there are any differences or similarities within this category, another figure (4) will be made comparing both MT systems within the same category to see the differences and similarities between the two MT systems:

Figure 4. Comparison between ***Google Translator*** **and** ***DeepL*** **according to the** ***Accuracy*** **category**

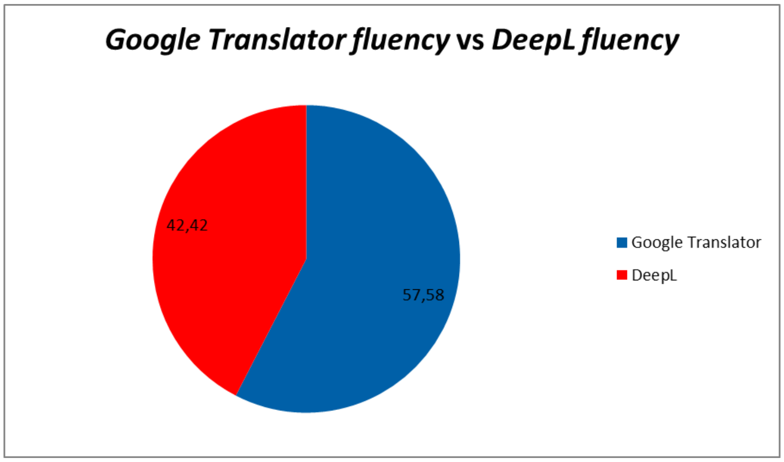


As it can be estimated in figure (4), the data obtained show that *Google Translator* and *DeepL* produce similar results in some aspects within the *accuracy* category. These results prove that 43.75% of the aspects to improve caused by *Google Translator* have to do with *terminology*. Moreover, the 50% of the aspects to improve dealing with accuracy when using *DeepL* are also related to *terminology*, so this figure projects which is the main problem of these MT systems, concluding that it is the specificity of the text. In addition, *Google Translator* displays that 18.75% of the aspects to improve committed belong to *mistranslation*, and in *DeepL*, the percentage is similar, 16.96%. *Untranslated elements* aspects to improve are also important in MT systems, since 27.68% of the total aspects to improve are produced by *Google Translator*, and 19.64% in the case of *DeepL* belong to this type of aspect. However, these MT systems show a difference in *omission* and *addition* aspects to improve. While *Google Translator* has not made any aspect to improve related to *addition*, *DeepL* show 10.71% of *addition* aspects to improve. However, *DeepL* has only 2.68% in *omission* aspects to improve while *Google Translator* has 9.82%. So, in these features, they confirm different results.

5.3 *Fluency* aspects to improve in both MT systems

In the next figure (5), we are going to follow the same steps as in figure (3); a comparison between the total aspects to improve in the *fluency* category within both MT systems is going to be carried out to prove which MT system is more precise in the *fluency* category than the other:

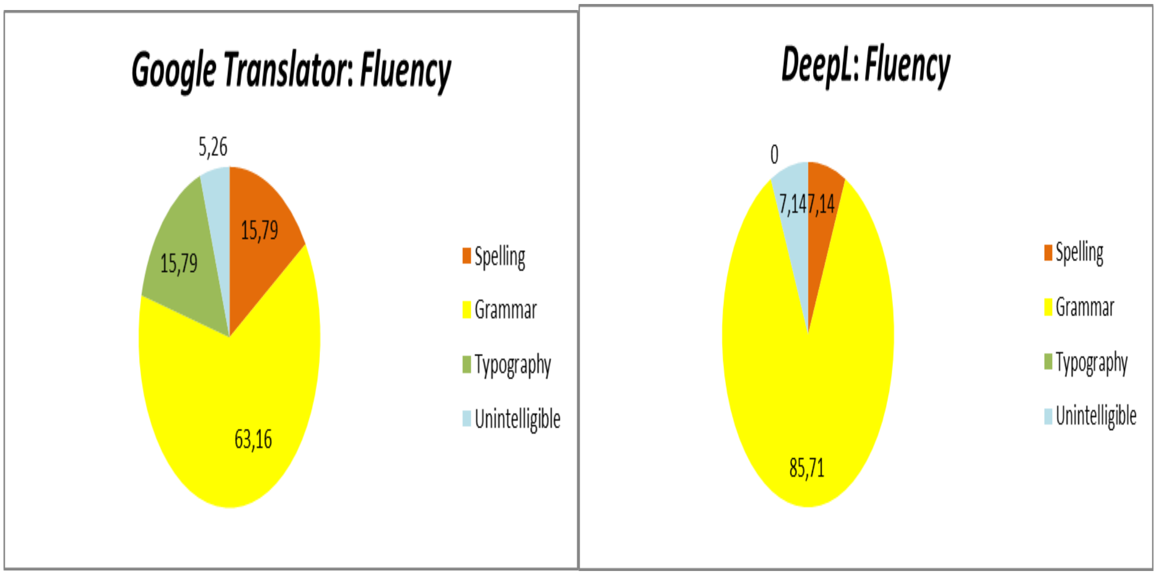
Figure 5. ***Google Translator fluency* vs. *DeepL fluency***



In figure (5) there is a significant difference between these two MT systems. The results show that 57.58% of the total aspects to improve in the *fluency* category have been produced by *Google Translator*. This indicates that in this type of text, *DeepL* works better in terms of *fluency*, since it has produced fewer aspects to improve in the same text. A figure will also be made to assess what types of aspects to improve have been produced by each MT system. We will also analyze the differences and similarities between them.

In the next figure (6) we are going to observe a comparison between *Google Translator* and *DeepL* dealing with the *fluency* category:

Figure 6. Comparison between *Google Translator* and *DeepL* according to the *fluency* category



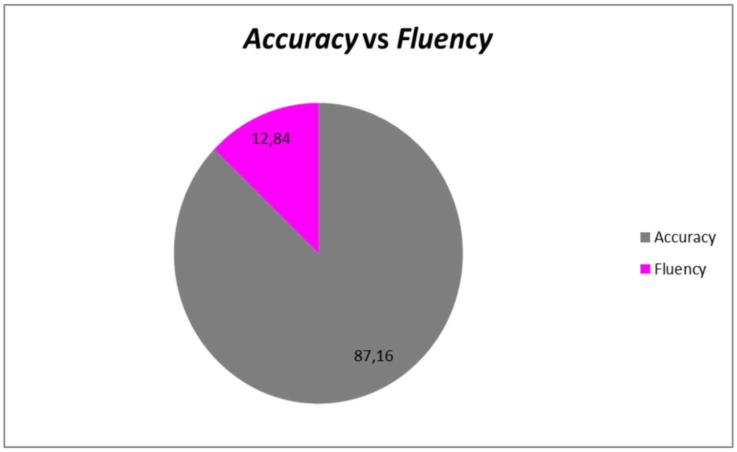
Grammar aspects to improve represent the majority of aspects that can be improved within the *fluency* category. In view of this, 63.16% of the total aspects to improve from *Google Translator* are those related to *grammar*, and in *DeepL* the percentage is even higher, 85.71%. As it is demonstrated, *grammar* is the main problem for these MT systems in the *fluency* category. *Spelling* and *typography* are also a problem in *Google Translator*, since both show 15.79%, while *DeepL* do not present any *typography* aspect to improve.

However, there is 7.14% in *spelling* aspects to improve; even so, the percentage in *Google Translator* is higher. Finally, the percentage of *unintelligible* aspects to improve in *Google Translator* is lower than the *DeepL* one (5.26% vs 7.14%), what proves that these types of aspects are the least usual of this category in *Google Translator*.

5.4 Accuracy vs. *Fluency*

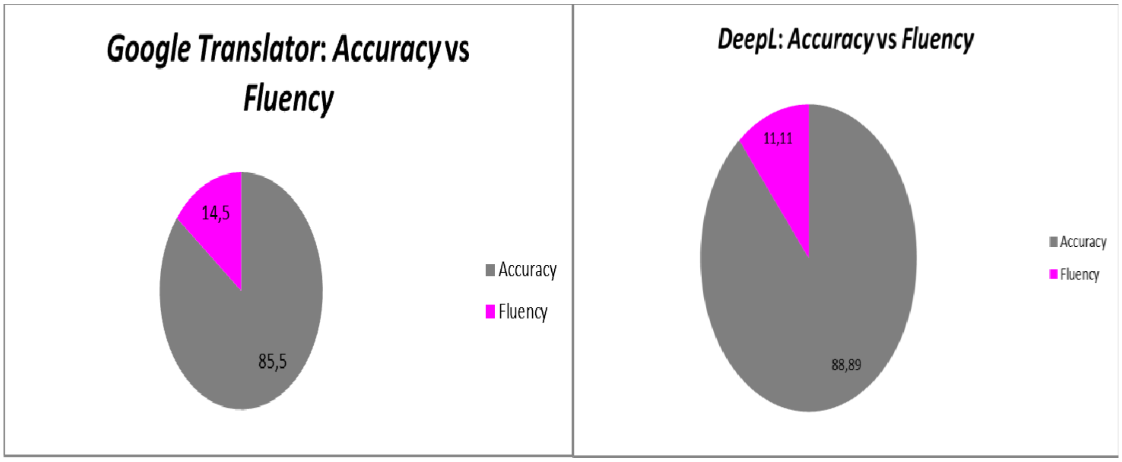
In the following figure (7), the total number of *accuracy* and *fluency* aspects to improve is going to be compared to see if one category is more problematic than the other:

Figure 7. *Accuracy* vs. *Fluency* in general terms



The result is that *accuracy* represents 87.16% of the total aspects to improve while there is only 12.84% belonging to *fluency* aspects to improve. This figure (7) shows us that these MT systems tend to make more aspects to improve when dealing with those aspects that belong to the *accuracy* category. This means that MT systems show more difficulties when dealing with *omission*, *addition* and vocabulary aspects (*terminology*, *mistranslation* and *untranslated elements*). However, in the last figure (8), we will be able to compare *accuracy* and *fluency* in both MT systems to see in which category there are more aspects to improve. But, as it is possible that this fact could only happen in one of the two MT systems, another figure (8) will clarify whether this occurs in both MT systems or just in one.

Figure 8. ***Accuracy* vs. *Fluency* in *Google Translator* and *DeepL***



As we can see in figure (8) above, the *accuracy* category represents the majority of problems in both MT systems. In the case of *Google Translator* 85.5% of the aspects to improve are from the *accuracy* category while only 14.5% is from the *fluency* category. *DeepL* shows similar percentages: 88.89% of the aspects to improve belong to the *accuracy* category and 11.11% to the *fluency* one. This figure proves that both *Google Translator* and *DeepL* are more susceptible to make more structures that can be improved, the more specific the text to be translated is.

As it has been verified in this section there are great similarities between *Google Translator* and *DeepL*. However, the results have shown that *DeepL* performs a better translation in general terms even though it produces more *grammatical* and *terminological* aspects to improve, which are two very important types of aspects to improve. For these reasons, it is clear to us that MT systems can be used as a helpful tool, but not as a reliable toolif users apply them without taking steps to improve the quality of the translation.

6. CONCLUSION

This paper has focused on the process of analyzing and comparing the aspects to improve caused by two important MT systems of a specialized medical text, facilitating a post-editing proposal in order to have a better understanding of the aspects to improve as well as the search for information about *scleroderma* both in English and Spanish. In addition, the original text and the post-edition version are available in a zip file in the CD of this project.

As we already know, MT systems are tools that can facilitate translation work on some aspects. However, it is important to clarify that they are not perfect systems and people should not be afraid of the fact that these programs are going to replace the great work translators do since they still have much to improve. This project has been made to raise awareness on the importance of making edits to those translations made by MT systems since these programs make several aspects and mistakes which cannot be considered native-like structures. In addition, this project can also be considered as a suggestion for improvement in those areas where MT systems have more problems.

As we have proved in the results section, these tools can cause important aspects to improve when translating a text from a specific field, in this case, *systemic sclerosis*. The reason for this is that in medicine, people use specific *terminology*. It is also important to mention that the elected text deals with an illness that nowadays is not very well known and developed, so we should be careful if we opt to choose a MT system to translate a medicine text since we have been able to verify that *Google Translator* and *DeepL* produce several structures that should be improved in terms of *accuracy* and *fluency*. Although we have seen in the results section that *DeepL*, in general terms, produces fewer aspects to improve than *Google Translator*, the results have been clear: both of them produce such a large number of aspects to improve that it has been shown that we should not use MT systems without making a post-editing version. We have also been able to verify that most of the aspects to improve have been produced in the *accuracy* category for both MT systems, and within this category, the biggest problem has been the *terminology*, in both MT systems. This fact has shown us that these MT systems present greater difficulty when translating a specific text. There are other important aspects to improve despite the fact that most of the aspects to improve are related to specialized *terminology*. During the classification of these different aspects and the post-editing process, it can be seen that a large number of other aspects to improve belong to word order (*grammar*).

It is worth mentioning that a large part of the presented aspects to improve are caused by the literal translation of the MT systems. Furthermore, resolving *terminology* aspects to improve was not as easy as expected, Spanish parallel texts had to be used but this was not enough. Information about the specific *terminology* (that this text contains) had to be gathered from experts in this field such as doctors and a patient who suffers the disease, which have turned out to be fundamental to get myself acquainted with general medical terms and specific terms related to the disease. Moreover, different types of dictionaries have also been used, such as English monolingual medical dictionaries like the *MedlinePlus* (2019) dictionary, bilingual English/Spanish medical dictionaries such as *Diccionario crítico de dudas ingles-español de medicina* by Navarro (2005) and other online Spanish monolingual dictionaries such as the *RAE* (2001).

In short, MT systems can be useful tools to translate general things from unknown languages to get an idea of what is wanted to transmit in the text. But to improve the translations of well-known MT systems, it is necessary to take into account certain aspects:

- To be aware that MT systems are not a reliable tool to copy and paste in specialized texts, since they make important aspects to improve, as we have been able to prove in the results section. This is one of the main reasons why professional translators affirm that MT systems will not replace their work, since today these MT systems continue making countless aspects to improve.
- Apply the concept of usefulness but taking into account that professional translators are required to improve MT systems' translations to be more specific and accurate.
- Emphasize that it will be essential to perform a post-edition when dealing with specialized texts to guarantee a good quality of the text, since it is in this area where most of the aspects to improve are produced due to the scientific language.
- To know that globality plays an important role when using MT systems due to the fact that they can be considered helpful for people who do not know the language they have to face. Even professional translators can decide to work by doing post-edition versions instead of translating from scratch. Translators will always help citizens have better communication and understanding around the world.

Finally, it is necessary to continue advancing in technology and in the training of professional translators specialized in medical texts for the future of medical translation. The importance of translation lies in transmitting information on the translated subject, in this case, *scleroderma*, in a reliable, faithful and accurate manner.

7. REFERENCES

Aranberri, N. "Posedición, productividad y calidad". Universidad del País Vasco. Revista Tradumàtica: tecnologies de la traducciò. pp. 471-477. 2014.

Arevalillo, J. "La traducción automática en las empresas de traducción". Hermes Traducciones y Servicios Lingüísticos, SL. Revista Tradumàtica: tecnologies de la traducció, pp. 179-184. 2012.

Azou, K. "La post-edición. Un análisis crítico de unas post-ediciones del crowd, de un hablante nativo y de una estudiante de traducción". Faculteit Letteren & Wijsbegeerte. Universiteit Gent. 2016.

Babych, B. "Automated MT evaluation metrics and their limitations". Revista Tradumàtica: tecnologías de la traducció, pp. 464-70. 2014.

Blasco, L. "¿Cuáles son los traductores que compiten con Google Translate y cómo funcionan?". BBC News. Web. Available at: https://www.bbc.com/mundo/noticias-42819225. 2018.

Burchardt, A. "Engines of invention: How DFKI became the world´s biggest nonprofit AI Research Center". Reykjavík Al Festival (Iceland). 2017.

Clínica Universidad de Navarra (CUN). Diccionario médico. Available at: https://www.cun.es/diccionario-medico. 2020.

German Research Center for Artificial Intelligence (DFKI); QTLaunchPad. "Multidimensional Quality Metrics (MQM) Definition", QTLaunchPad. Available at: http://www.qt21.eu/mqm-definition/definition-2015-12-30.html. 2015.

Hettige, B. & Karunananda, A. "Existing systems and approaches for Machine Translation: a review". ResearchGate. Department of statistics and computer science. Faculty of applied science. University of Sri Jayewardenepura, Sri Lanka & University of Moratuwa, Sri Lanka. 2011.

Jakobsen, A. "Logging time delay in translation" Copenhagen working papers in LSP. pp. 1:73-101. 1998.

Jurafsky, D. & Martin, J. "Speech and language processing. An introduction to natural language processing, Computational linguistics and speech recognition". Upper Saddle River, NJ: Prentince Hall. 2000.

Kumar, P. & Kumar, E. “Statistical machine translation based Punjabi to English transliteration system for proper nouns”. *International Journal of Application or Innovation in Engineering & Management (IJAIEM).* Department of Computer Engineering, Guru Kashi University, Talwadi Sabo, Punjab. Vol. 2. 2013.

Krings, H. “Was in den Köpfen von Übersetzern vorgeht”. Tübinger Beiträge zur Linguistik. Tübingen: Narr verlag. 1986.

Li, Q., Sahhar, J. & Littlejohn, G. “Red flags in scleroderma”. Clinical Practice. Reprinted from Australian Family Physician. Vol 37, No. 10. AFP. Rheumatology Department, Monash Medical Centre, Melbourne, Victoria. PDF. Available at: https://www.racgp.org.au/afpbackissues/2008/200810/200810Li.pdf. 2008.

Lommel, A., Burchardt, A. & Uszkoreit, H. “Multidimensional Quality Metrics: A flexible system for assessing translation quality”. ASLIB. DFKI Berlin. PDF. Available at: https://www.semanticscholar.org/paper/Multidimensional-Quality-Metrics-%3A-A-Flexible-for-Lommel/f07842338ba08748f0a9eb3dcda277dd85ab78df. 2013.

MedlinePlus. MedlinePlus en español. “Bethesda (MD): Biblioteca Nacional de Medicina (EEUU)”. Available at: https://medlineplus.gov/spanish/. 2019.

Mego, Y. “Automatización de la subtitulación del primer capítulo de la temporada 14 del documental “Forensic Files” del inglés al español con traducción automática neuronal y posterior posedición”. Universidad Autónoma de Barcelona. Máster en Tradumática. TFM. Facultad de Traducción e interpretación. 2019.

Mendoza, M. “La posedición de traducciones de textos técnicos del alemán al castellano”. Universidad Autónoma de Barcelona. 2017.

Merriam-Webster. Merriam-Webster dictionary. Available at: https://www.merriam-webster.com/. 2011.

Molina, A. “Creación de un motor de traducción automática estadístico (EN>ES) para textos del ámbito farmacéutico. Comparación con otros motores de traducción automática neuronal existentes”. Universidad Autónoma de Barcelona. Master Tradumática: Tecnologías de la Traducción. 2019.

Navarro, F. “Diccionario crítico de dudas inglés-español de medicina”. Madrid: MacGraw-Hill Interamericana. 2005.

Oliveró, S. “Esclerosis sistémica: supervivencia y factores pronósticos”. Departamento de Medicina Interna. Universidad Autónoma de Barcelona. 2013.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. "BLEU: a method for automatic evaluation of Machine Translation". Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL). Yorktown Heights, New York, 2002.

Quah, C. "Translation and Technology". Palgrave studies in translating and interpreting. Series editors. Anderman, G & Rogers, M. The centre for translation studies, University of Surrey, UK. 2006.

Real Academia Española: Banco de datos (CORPES) [en línea]. Corpus de referencia del español actual. Available at: http://web.frl.es/CORPES/org/publico/pages/colocacion/coaparicion.view. 2020.

Real Academia Española: Banco de datos (CREA) [en línea]. Corpus de referencia del español actual. Available at: http://corpus.rae.es/creanet.html. 2020.

Real Academia Española. "Diccionario del español jurídico". Available at: https://dej.rae.es/.

Real Academia Española. "Diccionario de la lengua española (22.a ed.)". Available at: https://www.rae.es/drae2001/. 2001.

Real Academia Española. "Diccionario Panhispánico de Dudas (DPD)". Available at: https://www.rae.es/obras-academicas/diccionarios/diccionario-panhispanico-de-dudas. 2005.

Seoane, A. "Lenguaje controlado aplicado a la traducción automática de prospectos farmacéuticos". Universidad de Alicante. Departamento de Traducción e Interpretación. Facultad de Filosofía y Letras. 2015.

Sepesy, M., & Donaj, G. "Machine translation and the evaluation of its quality". IntechOpen. pp. 1-20. 2019.

Serrano, C. "Postedición: situación actual de la traducción automática y estudio de un caso práctico". Universitat Oberta de Catalunya. Area de Posgrado Artes y Humanidades. Máster de Traducción Especializada. 2015.

Shalini, R., and Hettige, B. "Dictionary based Machine translation system for Pali to Sinhala". Faculty of humanities and social sciences. University of Sri Jayewardenepura, Nugegoda. Sri Lanka. 2017.