UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Scienze della Mediazione Linguistica e Culturale

Corso di Laurea Triennale in Mediazione Linguistica e Culturale

UNIVERSIDAD DE VALLADOLID

Facultad de Traducción e Interpretación

Grado en Traducción e Interpretación

ELABORATO FINALE/TRABAJO FIN DE GRADO

The Translation of Spanish Agri-food Texts into English and Italian Using Machine Translation Engines: A Contrastive Study

Presentato da/Presentado por

Alberto Morán Vallejo

Relatrici/Tutoras

Paola Catenaccio – Mª Teresa Ortego Antón

June 2019

A mis padres, Jesús y Ana,

por darme la oportunidad de mi vida.

INDEX

INDEX OF TABLES

INDEX OF FIGURES

INDEX OF GRAPHICS

INDEX OF ABBREVIATIONS

EBMT: Example-based machine translation
CAT: Computer-aided translation
MT: Machine translation
NMT: Neural machine translation
PE: Post-editing
SL: Source language
SMT: Statistical machine translation
TEnT: Translation environment tool
TL: Target language
TMS: Terminology management system

ABSTRACT

The agri-food industry is one of the most important economic sectors in Spain. The work of translators in this sector enables to open new business lines in other countries and, therefore, it contributes to the economic growth. Among the most significant industries, the meat industry is in the lead on invoicing and direct employment. Nevertheless, most of the companies are small and medium enterprises and they need to adapt data about their products to attract new customers, such as dried meats companies, which have not been deeply studied. The aim of this study is to analyse the output provided from a selection of machine translation systems for a text from the agri-food industry from Spanish into English and Italian, using different machine translation engines, in order to identify the main errors, to assess the MT engine which provides better results as well as to emphasize the importance of training translators in this field.

Keywords: machine translation engines, English, Spanish, Italian, agri-food sector.

RESUMEN

La industria agroalimentaria es uno de los sectores económicos más importantes en España. El trabajo de los traductores en este sector permite la creación de nuevos negocios en otros países y, por lo tanto, contribuye al crecimiento económico. Entre los más importantes, el sector cárnico es líder en facturación y empleo directo. Por otra parte, la mayoría son pequeñas y medianas empresas y tienen la necesidad de adaptar información sobre sus productos para atraer a nuevos consumidores. Sin embargo, hay algunos ámbitos en los que aún no se ha investigado con profundidad, como los embutidos. El objetivo de este estudio es analizar el resultado de la traducción automática de un texto agroalimentario, del español al inglés y al italiano y utilizando diferentes motores de traducción automática, con la finalidad de identificar los errores principales, valorar qué motor de TA produce mejores resultados y enfatizar la importancia de formar a los traductores en este ámbito.

Palabras clave: *traducción automática, inglés, español, italiano, sector agroalimentario.*

RIASSUNTO

L'industria agroalimentare è uno dei settori economici più importanti della Spagna. Il lavoro dei traduttori in questo settore consente la creazione di nuove imprese in altri paesi e quindi contribuisce alla crescita economica. Tra i più importanti, il settore della carne è leader nel fatturato e nell'occupazione diretta. D'altro canto, la maggior parte sono piccole e medie imprese e devono adattare le informazioni sui loro prodotti per attirare nuovi consumatori. Tuttavia, ci sono alcuni settori che non sono ancora stati studiati a fondo, come le salsicce. Lo scopo di questo studio è quello di analizzare i risultati della traduzione automatica di un testo relativo all'industria agroalimentare, dallo spagnolo all'inglese e all'italiano e utilizzando diversi motori di traduzione automatica, con l'obiettivo di identificare i principali errori, valutare quale motore di traduzione automatica produce i migliori risultati e sottolineare l'importanza della formazione dei traduttori.

Keywords*: traduzione automatica, inglese, spagnolo, italiano, settore agroalimentare.*

1. INTRODUCTION

This final degree project, entitled *The Interlinguistic Transfer of Spanish Agri-food Texts into English and Italian Using Machine Translation Engines: A Contrastive Study*, focuses on two very important areas of translation. On the one hand, the use of two new techniques related to computer-assisted translation (CAT tools), in particular, machine translation and post-editing. We have decided to carry out this project with these innovative tools as we intend to demonstrate that they can be useful to increase the productivity of the professional translator. Nevertheless, these tools are not exempt from difficulties and deliver results that are generally not as good as expected. Furthermore, we would like to highlight the mistakes provided by these tools and solve them in a practical and professional way. Likewise, another field of interest of our project are the descriptive and promotional texts about the elaboration of cured meats.

1.1. JUSTIFICATION

One of the reasons why we have selected this field is related to the importance of translation in the agri-food sector, since there are few studies in this domain despite the fact that it is one of the most important economic sectors in Spain. According to Rivas Carmona & Veroz González (2018: 17), it is essential to raise awareness of agri-food translation as a thematic variety within specialised translation. Furthermore, according to the Spanish Ministry of Agriculture, Fisheries and Food (MAPA, 2019), the agri-food industry is the main manufacturing industry in the EU with a turnover of 1,109,000 M€. In addition, the Spanish agri-food industry ranks 5th in terms of turnover at the EU level (8.7 %).

Besides, we have decided to raise awareness of machine translation and post-editing because it has become a key tool for translation service providers and freelance translators. As it is shown in the ProjecTA report (Torres Hostench *et al*., 2016: 24), the 47.3 % of the translation services providers surveyed stated that they use machine translation as a working tool. However, they state that post-editing represents less than 10 % of their workload, so it seems that companies do not see post-editing as a service equivalent to normal human revision.

Throughout the Degree in Translation and Interpreting we have been able to develop new knowledge thanks to subjects such as Information Technology Applied to Translating, Computer-Assisted Translation (CAT) and ICT for Translation and Localisation. In addition, we have received complementary training thanks to the course organized by CITTAC, entitled *Advanced Seminar for Trainers and Junior Researchers on Machine Translation and Post-editing*, with the collaboration of Prof. Miriam Seghiri and Prof. Pilar Sánchez Gijón. Finally, due to the fact that we have been awarded a grant from the Ministry of Education, we are running a project parallel to our final degree work called "An approach to machine translation (MT) and post-editing (PE) tools from the translator's perspective", tutored by Dr. Mª Teresa Ortego Antón.

In this way, our final degree project is developed within this framework with the aim of providing more in-depth research on advances and new technologies in this field.

1.2. COMPETENCES

This work puts into practice the competences acquired during the Degree in Translation and Interpreting, as well as a series of competences described in the course guide of the subject entitled *Trabajo de Fin de Grado,* that is, the final degree project.

The general competencies developed throughout this final project correspond to G1, G2, G3, G4, G5, G6.

As for the specific competencies acquired during the Degree, the following are implemented in this project: E1, E5, E8, E17, E18, E19, E34, E41, E47, E49, E50, E51, E52.

1.3. AIMS

With this project we intend to verify whether machine translation engines offer an acceptable translation from Spanish to English and Italian as an output when they transfer promotional and descriptive texts from a certain field of knowledge, the agri-food industry. Meanwhile, our aim is to detect and analyse the main translation errors that these engines produce.

In addition to this main objective, we aim to achieve the following specific aims:

- To offer an overview of the main MT and PE techniques.
- To be aware of the importance of machine translation for a professional translator today.
- To compare different machine translation engines according to their software.
- To be able to detect the errors of different machine translation engines.
- To classify errors according to their category, for example, terminology, grammar, etc.
- To contrast information on which machine translation engines commit the majority of the errors.
- To observe the translation of specific terminology of the field of the agri-food industry, more specifically, the elaboration of cured meats.

2. BACKGROUND

In this chapter, we are going to deal with the main concepts of our final degree project: machine translation, its challenges and limitations, its applications and its fundamental approaches. Furthermore, it is also important to address the concept of computer-assisted translation in order to differentiate it from machine translation. Moreover, the last section of this chapter provides a summary of the main characteristics of post-editing.

2.1. MACHINE TRANSLATION

Firstly, the standard definition of machine translation is provided by ISO 18587:2017 and it is identified as "the automatic translation of a text from one language to another using a computer application". Some years before, Forcada (2010: 215) provided a similar definition for machine translation (MT): "the translation, by means of a computer using suitable software, of a text written in the *source language* (SL) which produces another text in the *target language* (TL) which may be called its *raw translation"*. In other words, it is an automatic process carried out by a software which provides a translation of the text in the target language.

2.1.1. CHALLENGES AND LIMITATIONS OF MACHINE TRANSLATION

Machine translation has a great significance today but there are still many challenges and limitations that we have to grapple with. *Raw translation* is a term used to designate the output of an MT system, which is usually very different to the output of translation professionals. However, this does not mean that MT is not useful; the key is to be aware of its specific applications. Furthermore, identifying the contexts in which MT can be used effectively is important to know what can be expected of it.

According to Arnold (2003), the obstacles faced in machine translation can be classified in four groups:

a) The first group includes the cases in which form does not completely determine the content due to the *ambiguity* of language. Sentences can be ambiguous because their words have several meanings (*lexical ambiguity*) or more than one possible syntactic structure (*syntactic ambiguity*). In addition, there are some cases in which there are both syntactic and structural ambiguity at the same time. In short, ambiguity is a problem because the MT system must choose the correct meaning of a sentence in order to produce a suitable translation.
b) Secondly, there are some cases in which content does not completely determine form because there exist different ways to express an idea in the same language.
c) In the third place, we must be aware that different languages use different structures to provide the same information. This means that the structures used by different languages can be so different that a word-for-word translation would be inacceptable or even wrong.
d) Finally, these three problems are the manifestation of intrinsic features of translation. In this way, the fourth group includes what Arnold calls the *description* problem, which states that translation theories cannot formally express all the mechanisms underlying

natural language translation. These problems are tackled using some methods such as radical simplifications or complete reformulations.

Nevertheless, despite these problems, the reality is that machine translation is increasingly being used. In 2016, almost half of the total amount of translation companies (47,3 %) used machine translation, as opposed to the 52,7 % that state that they do not use MT as a working method, as demonstrated by ProjecTA report about the use of machine translation and post-editing in Spanish by language service providers (Torres Hostench *et al*., 2016: 24).

2.1.2. APPLICATIONS OF MACHINE TRANSLATION

Machine translation has two main purposes which are formally called *assimilation* and *dissemination* (Forcada, 2010: 216).

Assimilation is a very common application used when one does not understand the source language. In assimilation, texts are machine-translated in order to have an approximate idea of the content of the text. Thus, errors are not too important if the engine provides a result in which the general sense of the text is achieved. The accuracy of the text depends on the MT system and the languages involved, but also on the user, who must know how to take advantage of this type of text.

On the other hand, *dissemination* is an application in which texts are machine-translated as an intermediate step in the production of a document in the target language that will be published (disseminated). Therefore, raw MT results have to be *post-edited*, which means that the text is revised and corrected by a professional translator. Indeed, the use of MT systems as a key component in the translation process increases the productivity compared to human translation, and it can be done in different stages:

1) *Machine translation followed by post-editing*: the raw MT output is edited by professional post-editors (preferably trained translators) to achieve an adequate text. This process seems to be advantageous when the cost of MT and post-editing is lower than the cost of human translation, but there may be some additional costs related to training and changes in the translation workflow.
2) *Pre-editing* consists on the edition of the source text before translation. The pre-editor must be trained to anticipate MT problems, and this involves extra costs. It may be helpful when the text has to be translated to several target languages since edits in the target text may be avoided. However, this does not mean that post-editing can be avoided.
3) *Controlled languages* are variants of the source languages with some lexical and syntactic restrictions designed to avoid problems in MT. It is helpful to avoid repetitive post-editing. However, designing a controlled language is costly and it is only profitable in the event of heavy repetitive pre-editing.

To conclude this section, it should be noted that applications of machine translation engines are undergoing changes and steadily increasing, and their stages are modified or altered depending on the operation of each type of engine. Thus, the section above is essential for the understanding of this changing phenomenon.

2.1.3. APPROACHES TO MACHINE TRANSLATION ENGINES

Until the 90s the dominant approach was *rule-based* or *knowledge-based* machine translation, a system based on linguistic rules that allow the words to be put in different places and to have different meaning depending on context. It is made up of both computer and translation experts that programme the MT engine and compile dictionaries and grammatical rules to transform sentence structures in formats that could be processed by that MT engine.

However, there is another approach called *corpus-based* machine translation which has experimented a growth since the beginning of the 90s. This system consists of huge corpora of bilingual texts where sentences in one language have been aligned with their equivalents in the other language. Furthermore, hybrid approaches are nowadays an active field of current research. The most prominent example are statistical MT systems, which incorporate a mixture of linguistic knowledge, such as morphological dictionaries (Koehn 2009: 314) and statistical MT techniques.

In order to establish the major differences between these two main approaches, we have to be aware that rule-based systems are more complex because it takes a lot of time to encode explicitly the linguistic information that the system will use. Meanwhile, corpus-based systems can be structured more quickly but it is necessary to compile a large volume of sentence-aligned bilingual texts.

2.1.3.1. RULE-BASED MACHINE TRANSLATION

The most common rule-based MT systems are *transfer* systems. According to Hutchins & Somers (1992: 75), an ideal transfer system has three well-defined stages:

1) *Analysis*: in this stage, the MT system produces, from the sentence in SL, an abstract intermediate representation, in which linguistic classifications and groupings are established to allow for the application of general rules of translation (Forcada, 2010: 218).
2) *Transfer* is the stage in which the intermediate representation delivered by analysis is converted into a new intermediate representation for the target language, searching for words in the bilingual dictionary (*lexical* transfer) and applying rules (*structural* transfer).
3) Finally, *generation* is the production of a concrete sentence in the target language from this abstract intermediate representation.

These intermediate representations can be more or less complex depending on the depth of the analysis, which leads to a complete syntactic parsing of the sentence, or even a semantic representation of it.

Furthermore, there exist some systems in which only analysis and generation are necessary, since the analysis is so deep that a language-neutral intermediate representation is obtained. These systems are called *interlingua* systems and they do not need bilingual knowledge to add a new language to an existing system (just analysis and generation modules are required). Nevertheless, designing a general-purpose interlingua is equivalent to design a

complete model of the real world. That is why interlingua systems are restricted to limited-domain translation skills.

2.1.3.2. CORPUS-BASED MACHINE TRANSLATION

Corpus-based machine translation is also called *data-driven* machine translation and it can be divided into two main models: *example-based machine translation* and *statistical machine translation*. Both of them have something in common since a corpus of sentence-aligned bilingual parallel texts is required for both paradigms.

On the one hand, example-based machine translation (EBMT) (Carl & Way, 2003) generally consists of three distinct phases:

1) *Matching* is the process of segmentation of the new sentence to be translated. Then, the segments are matched against identical or similar segments in the SL side of the bilingual examples in the corpus.
2) *Alignment* is the stage in which a correspondence between fragments in the TL side of the matched bilingual examples is determined in order to build "translation units".
3) *Recombination* is the phase in which the TL "translation units" are combined into a translation for the new sentence.

Each one of these phases is still an open subject of research: the segmentation process is essential for successful matching, alignment is important especially when languages are not closely related, and recombination must be successful by avoiding repeated words or lack of agreement.

Furthermore, an important characteristic of EBMT is that, when the new sentence is identical to a sentence in the corpus, its translation is recovered and used, as if it were a translation memory system (this translation tool is explained in the following section).

On the other hand, statistical machine translation (SMT) (Koehn, 2009: 110) is currently the dominant approach in MT research and it has experienced a growth in MT market. SMT consists on finding the sentence on TL that is the highest possible to be the translation of a SL sentence. That is, a SL and a TL sentence are a translation of each other *with certain probability* (Forcada, 2010: 220). This system operates through a *probability model* inferred from the bilingual corpus.

This probability model is made of several components with lexical probabilities and alignment probabilities, hence linguistic information is not used: the two models of the SL and TL sentences are inferred by using complex statistical estimation techniques on sentence-aligned bilingual corpora.

2.1.3.3. NEURAL MACHINE TRANSLATION

Recently, a new type of machine translation engine based on the human neural system has emerged. This method provides the best results and is therefore the system used by some of the most widely used translation engines, such as Google Translator and DeepL. The following definition of Bahdanau *et al.* (2014: 1) is the one that we considerate the most appropriate to define neural machine translation (NMT):

Neural machine translation is a recently proposed approach to machine translation. Unlike the traditional statistical machine translation, the neural machine translation aims at building a single neural network that can be jointly turned to maximize the translation performance. The models proposed recently for neural machine translation often belong to a family of encoder-decoders and consists of an encoder that encodes a source sentence into a fixed-length vector from which a decoder generates a translation.

This type of automatic translation emulates the functioning of a human brain. NMT provides better results since words and phrases are represented numerically by vectors, so that the system takes into account the phrase as a whole, the subtleties of the language and any variations that may exist.

In short, machine translation systems are a complex reality which is constantly developing, and research will improve systems that will be available as a component to be integrated in new translation workflows.

2.2. COMPUTER-ASSISTED TRANSLATION

It is important to know the differences between machine translation and other translation technologies as computer-aided translation (CAT). Bowker & Fisher (2010: 60-65) make a clear distinction between them:

Computer-aided translation (CAT) is the use of computer software to assist a human translator in the translation process. The term applies to translation that remains primarily the responsible of a person, but involves software that can facilitate certain aspects of it. This contrasts with machine translation (MT), which refers to translation that is carried out principally by a computer but may involve some human intervention, such as pre- or post-editing.

In this way, machine translation is performed by the computer without human intervention; in computer-aided translation, translation is performed by a professional with the help of translation tools. Hence, computer-aided translation is conceived as a wide range of tools that assist a human translator in the translator process, while machine translation is carried out principally by computer.

We often associate the term *computer-aided translation* with the functions that it offers, such as online dictionaries, browsers and toolbars. However, this term is reserved for software designed specifically for the development of functions of the translation process, rather than tools for general application.

2.2.1. TRANSLATION ENVIRONMENT TOOLS

The most popular CAT tool is the Translation Environment Tool (TEnT), a suite of tools which work as a translator's workstation. There are some individual components that differ depending on the product; however, the main structure of a TEnT is composed of a translation memory (TM) and a terminology management system (TMS).

A TM is a tool that stores previously translated texts and makes easier to consult them for reuse. The source and target texts are divided into segments and each segment from the source text is linked to its equivalent in the target text. These segments are stored in a TM database so that when a translator has a new text to translate, the TM system compares the segments of the new text with the content of the TM database. Using pattern-matching, the systems finds matches for a given segment. The translator does not have to accept the matches compulsory, he is simply presented with them and he decides to accept or modify them according to his professional criteria.

Furthermore, TM systems are typically integrated with a TMS, which stores terminological information in a termbase. Translators can customize each database entry with various fields (e.g., term, equivalent, definition, context, source) and they can fill them and consult them. However, termbases can also work in an automated way as TM systems do. For instance, using a feature known as active terminology recognition, a TMS can scan a new text, compare its contents against a specified termbase, and automatically display records for any matches that are found (Bowker & Fisher, 2010: 63).

To conclude this section, it is worth to highlight the great impact of CAT tools on translation. Many translation agencies, governments and freelance translators have adopted CAT tools. Nevertheless, the costs and benefits of CAT tools must be taken into account before adopting it.

On the one hand, the efficiency of TM depends on its ability to return matches. Texts that are repetitive or similar to other translations normally generate useful matches, whereas literary works or marketing material are less predictable. Moreover, the quality of the output of a CAT tool still depends on human translators. Despite of this, some clients attribute less value to the work of translators who use CAT tools. Therefore, CAT also affects the professional status of translators, their remuneration and their intellectual property rights (Bowker & Fisher, 2010: 63).

By way of conclusion, professional translators need to be equipped with specialized skills, knowledge and tools, and CAT tools therefore will continue their crucial work into the next decades.

2.3. DEFINITION AND FEATURES OF POST-EDITING

Despite the advances in the field of machine translation of the previous sections, the current development of machine translation systems requires the participation of human beings to post-edit and produce high quality translations. Moreover, machine translation and post-editing are two intrinsically linked processes, as shown by the ISO that regulates both processes (ISO 18587:2017 *Translation services – Post-editing of machine translation output*.) Thus, although there are multiple definitions of post-editing, in short it is the process of correcting and editing fragments of a text that takes place after the automatic translation process in order to improve the result and thus achieve the same translation quality that it would have had if it had only been translated by a professional.

Traditionally, machine translation engines have been known as imprecise tools that make many mistakes of all kinds. However, it is now beginning to be integrated as a working method of professionals in the translation industry because it means a greater degree of productivity and allows machine translation to be exploited as well as improved. As demonstrated in 2016 ProjecTA report about the use of machine translation and post-editing in Spanish language service companies (Torres Hostench *et al*., 2016: 32), the 80 % of Spanish translation companies use some type of pre-editing before MT. Furthermore, almost a 33 % of the companies surveyed use post-editing on more than 10 % of their translation projects.

The post-editing process is different from the traditional correction process of a conventional translated text. First of all, because errors are different in nature. Secondly, because an error in machine translation is not caused by a failure in the interpretation of meaning, but by the misapplication of statistical rules. That is to say, if the correct meaning is not reached it is because the rules were not properly applied.

Moreover, in the technological field, post-editing is increasingly present in CAT programmes. In the field of computer-assisted translation, many softwares currently include an option to incorporate the automatic translation of segments so that the translator only has to post-edit them. For example, memoQ can be linked to machine translation engines such as Google Translate or Bing.

2.3.1. TYPES OF POST-EDITING

There are many types of post-editing depending on the classification criteria. According to Tatsumi (2010: 11-14), we can talk about *human post-edition* and *systematized post-edition*, depending on whether the task is performed only by a human or also with the help of post-editing tools. For example, Koehn (2009) introduced three types of tools for pre-editing texts (sentence completion), machine translation and post-editing. According to this study, all these tools served to increase the speed and quality of the translation.

On the other hand, as mentioned above, the goal of post-editing is to increase productivity, as well as making the text more responsive to customer needs. Taking this aspect into account, there are two types of post-editing depending on the level of acceptability to be achieved (Sánchez-Gijón, 2016: 160):

a) *Human quality or full post-editing:* the quality of the text resulting from the post-editing process must be at the level of a human translation.
b) *Good enough quality or light post-editing:* the main message of the source text must be respected and understood in the target text despite loss of naturalness or fluency. A perfect translation is not obtained, but it is a less expensive service.

As a conclusion of this chapter, despite the fact that machine translation and post-editing have entered the world of translation satisfactorily, it is important that decision-making in the translation project remains in the hands of the translator, who best controls the selection of those resources that allow the objectives set in the client's assignment to be achieved (Sánchez-Gijón, 2016: 161). Once the key concepts of this research work have been defined, we will proceed by explaining the methodology used in this final degree project.

3. METHODOLOGY

Taking into account the aim of this work, in this chapter we will describe the methodology used to carry out the analysis of the output of a sample of MT engines. Thus, firstly, we will characterise the features of the input. Next, we will justify the MT engines we have selected and, finally, we will explain the reasons by which we have selected the TEnT.

3.1. THE INPUT

In the first place, it is essential to describe the characteristics of the text that is going to be translated in this project. We have selected a promotional text about the curing of ham, taken from a website[1] of a small shop of dried meats and typical artisan products from Teruel, Spain. The selected text has 220 words in Spanish.

> *La fase de curación del jamón serrano, junto con la de maduración, son el último proceso que se realiza en la elaboración de un jamón. Lo más importante de las dos fases es controlar los niveles de temperatura, humedad (que debe ser entre 68% y 76%) y ventilación. Ambas fases están separadas por el "pannage" del jamón serrano, la acción de poner una capa de grasa de cerdo en el músculo del jamón.*
>
> *En la fase de curación del jamón se acaba el proceso donde se forman los aromas, pero el jamón sigue perdiendo agua. La temperatura óptima para la curación del jamón serrano debe estar sobre los 14ºC, así el jamón adquiere la estabilidad perfecta. La humedad debe ser entre 68% y 76%. La curación del jamón serrano se puede realizar en cualquier tipo de secadero: artificial o natural.*
>
> - *El secadero artificial está completamente cerrado y controlado por un sistema que ingresa y saca aire. El sistema consiste en alternar periodos con aire y sin aire.*
> - *El secadero natural, se puede realizar solo si el clima de la región lo permite, con temperaturas poco elevadas y regulares y poca humedad, con alguna variación a lo largo del tiempo. Debe estar equipado con mosquiteros, contraventanas, ventiladores y un calentador. Estos secaderos deben ser revisados varias veces al día.*

Regarding the main features of the text, we are going to describe them using Hymes model (1974: 10). This author takes into account the cultural content of the text, which is one of the aspects that may pose greater difficulty for machine translation

[1] https://www.sabor-artesano.com/curacion-jamon.htm (Consulted on: 23rd March 2019)

engines. In this way, Hymes (1974: 10) proposed a pattern to analyse a communicative event considering not only the form and contents of speech but also the cultural environment that influence on it. He includes the following elements:

a) The participants: the sender is the author of the text. In this case, the author of the text is unknown, but it is clear that the text is written by someone who works at the company and is responsible for uploading information to the website. Moreover, the receiver of the message is anyone who is interested in buying a product in this company or wants to learn more about the production of ham. Therefore, the text is intended for a general public with an average level of culture.
b) The channel, in this case, it is a written text published on the Internet.
c) The code, which corresponds to the linguistic elements of the Spanish language.
d) The settings, that is, the moment in which communication takes place and the circumstances surrounding the communication. In this genre, communication takes place in an extended way, from the time in which the author writes the text and publishes it on the web until the reader surfs on the Internet and reads the text.
e) The form of the message and the genre. We have selected an informative text about cured ham with a clear structure. At the beginning, the fact that is going to be narrated is presented. Then, the body of the text contains the fact itself. In addition, as typical of this type of texts, it has a clear, concise and natural language.
f) The attitude conveyed by the message and the content. This text is characterized by including typical patterns of descriptive and informative texts, since it offers information about the product. At the same time, it has some persuasive nuances to urge the reader to consume the product.
g) The event itself. This may vary depending on the communicative situation. It can be a reader who wants to broaden his knowledge about dried meats or a reader who is interested in the product and buys through the website.

Once we have described the characteristics of the text with which we are going to work, we proceed to provide a justification of the selected MT engines.

3.2. THE SELECTION OF MT ENGINES

To begin with, it is important to establish the criteria used to select different machine translation engines. The first criterion is related to the type of approach of machine translation, which we have previously described. According to Casacuberta Nolla and Peris Abril (2017: 68), it has been shown that neural machine translation often provides better results than statistical translation or rule-based translation. In addition, neural models are the state of the art in machine translation. This is why we have selected two neural machine translation engines for our project.

Secondly, we are going to use engines that have an API that can be integrated into a TEnT, in order to see the differences that occur in different translation environments when applying machine translation. Moreover, it is necessary to clarify certain concepts in order to understand how this engine works as an API in a CAT program. In computer

programming, an application programming interface (API) is a set of definitions and tools for building software. In general terms, it is a set of clearly defined methods of communication among various components. In this way, this criterion has played a role key when choosing the two translation engines described below.

3.2.1. DEEPL

The first engine we have selected to carry out our contrastive study is DeepL, an online machine translation service launched in 2017. This service supports translations of 9 languages in 72 language combinations and uses convolutional neural networks based on Linguee's database. When it was published in 2017, several tests indicated that it would outperform its competitors, such as Google Translator or Bing Translator, since it provides an accurate and fast result. DeepL defines itself as an automatic translator which "trains artificial intelligence to understand and translate texts"[2], as neural networks expand human possibility, overcome language barriers, and bring cultures closer together.

Unlike the free version of DeepL, *DeepL Pro* has many features such as data confidentiality, translation of fully editable documents, unlimited web translator use, among others. In addition, *DeepL Pro Advanced* allows the integration of the engine as an API within a CAT tool without limits. Thus, the version that I have purchased can be incorporated to a CAT tool.

In this case, the API (*DeepL Pro Advanced*) is integrated into the CAT tool (*memoQ*) in order to add a new function to the software: a MT engine that facilitates the work of the translator of posteditor and increases his productivity.

3.2.2. MICROSOFT TRANSLATOR

Microsoft Translator is a multilingual neural machine translation cloud service provided by Microsoft. This service supports 65 language systems as of April 2019 and is characterized by offering text and voice translation as a service for businesses. Therefore, we decided to choose this translator because of its functionality and its easy application in several consumer, development and business products. Since the text we are going to analyse belongs to a website of a food company, it may be interesting to study it as a tool for a company that wishes to export a product and, therefore, needs good quality to boost its production and export in foreign countries.

Although Microsoft Translator was originally a statistical translation engine, today its API is a neural translation service that can be easily integrated into applications, websites and other tools. Thus, we have selected this API to integrate it into Memsource, a cloud-based translation environment.

[2] DeepL Pro. (2019). Retrieved from https://www.deepl.com/en/pro.html

3.3. THE SELECTION OF THE TEnT

Once we have set out the criteria used to select the different machine translation engines to be applied, the next step is to describe the software we are going to use to run the engines.

3.3.1. MEMOQ

MemoQ is a proprietary computer-assisted translation software suite with runs on Microsoft Windows operating systems. It is developed by one of the fastest growing companies in the translation technology sector (formerly *Kilgray*) and it provides translation memory, terminology, machine translation integration and reference information management in desktop, client/server and web application environments, among other features. It is designed for freelance translators, but it is also used by many universities for academic purposes in the training of translators.

One of the main criteria we have taken into account when choosing memoQ is its availability, as the Faculty of Translation and Interpreting at the University of Valladolid provides free academic licenses for students. Hence, we can take advantage of all the software functionalities for free. In addition, the second criterion that has been considered is the possibility of integrating APIs, as we have mentioned before.

To set up DeepL in memoQ as an API, there are some steps that need to be followed.

1. At the top of the memoQ window, in the Quick Access toolbar, we click the Options icon and a new window will open.
2. On the Default resources pane, we click into the MT icon. We select the Mt profile we are using, and under the list, we click into Edit. Next, on the Services tab, we find the DeepL MT plugin.

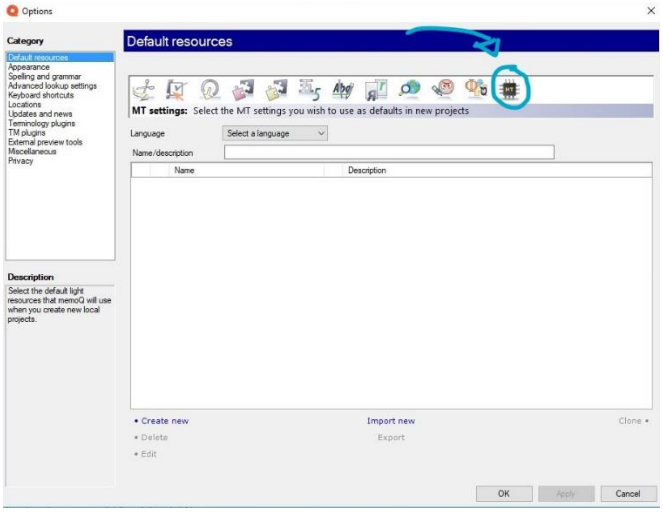


Figure 1. Integration of an API in MemoQ

3. As we have subscribed to the DeepL translation service, we have an authentication key. We introduce it into the text box and we accept.

In this way, when we open a project with a text to translate, we can use machine translation to improve our productivity without leaving the work screen.

3.3.2. MEMSOURCE

Memsource provides a cloud environment and also includes translation memory, terminology management and integrated machine translation. It was founded in 2011 and it offers the Personal, Academic and Developer editions for free, while it charges a monthly subscription fee for its most enriching features. Furthermore, Memsource has developed a unique approach to reduce translation costs by combining traditional translation with new patented technologies. Before a translation is assigned to a human translator, Memsource identifies content that can be translated automatically.

One of the main reasons why we have selected Memsource for this project is the fact that it is cloud-based, so we can see the differences with memoQ, which is a program that needs to be installed. In this way, Memsource has an easy and free access (criterion of availability), and these two factors must be considered when choosing a comfortable and productive work environment.

To integrate an API in Memsource, there are some differences in the procedure compared to memoQ:

1. First of all, we had to create a new project (in this case, it is called *TFG_AMV*).
2. Secondly, in the setting of the project, there is an option called *Machine Translation Engine.* There, in the drop-down list, we choose *Microsoft with Feedback.*

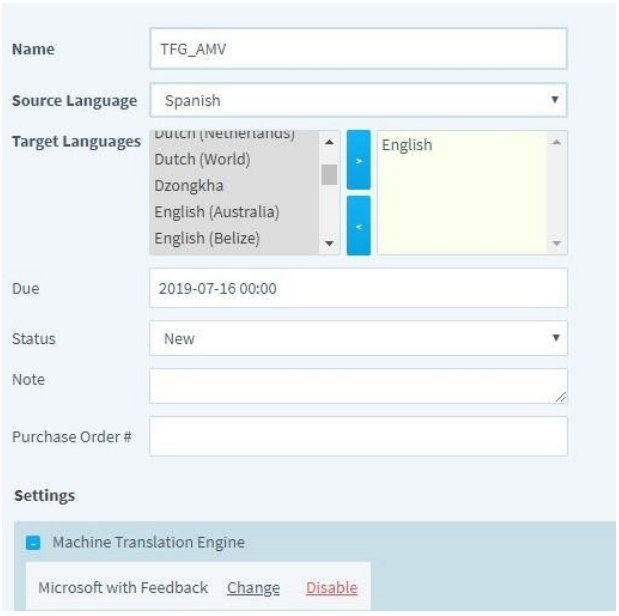


Figure 2. Integration of an API in Memsource

3. Once we have selected the MT engine, there is an option called *Pre-translate* in which we tick the box *Pre-translate from machine translation* in order to activate the option.

4. Then, we click *Create* and we add the source text to the TEnT to start the translation process

Once the MT engines and their main functionalities have been explained, we proceed to specify the criteria to analyse the output of each MT engine.

3.4. PATTERNS TO ANALYSE MT OUTPUT

In order to analyse the results of the different automatic translation engines in different languages, we have selected specific parameters with the aim of following a scheme and analysing errors in an orderly manner.

Thus, we have used a classification adapted by Ortiz (2016: 63-64) called *Multidimensional Quality Metric Error Typology (MQM)*, which has been previously used in other studies (Viver Sorolla, 2018). This classification is one of the most complex and contains specific parameters to evaluate the output of machine translation, which is the key of our project. The following table lists the error categories we are going to use in this study:

ACCURACY	**Terminology**	A term is translated with a term other than the one expected for the domain or otherwise specified.	
	Mistranslation	The target content does not accurately represent the source content.	
		Overly Literal	The translation is overly literal
		False Friend	The translation has incorrectly used a word that is superficially similar to the source word.
		Should not have been translated	Text was translated that should have been left untranslated.
		Date/time	Dates or times do not match between source and target.
		Unit conversion	The target text has not converted numeric values as needed to adjust for different units.
		Number	Numbers are inconsistent between source and target.
		Entity	Names, places or other “named entities” do not match
	Omission	Content is missing from the translation that is present in the source.	
	Addition	The target text includes text not present in the source.	
	Untranslated	Content that should have been translated has been left untranslated.	
FLUENCY	**Spelling**	Issues related to spelling of words.	
		Capitalization	Issues related to capitalization.
		Diacritics	Issues related to the use of diacritics.
	Typography	Issues related to the mechanical presentation of text. The category should be used for any typographical errors other than spelling.	
		Punctuation	Punctuation is used incorrectly for the locale or style
		Unpaired quote marks or brackets	One of a pair of quotes or brackets is missing from the text.
	Grammar	Issues related to the grammar or syntax of the text, other than spelling and orthography.	
		Morphology	There is a problem in the internal construction of a word.
		Part of speech	A word is the wrong part of speech.

		Agreement	Two or more words do not agree with respect to case, number, person or other grammatical features.
		Word order	The word order is incorrect.
		Function words	A function word is used incorrectly.
	Unintelligible	The exact nature of the error cannot be determined. Indicates a major break down in fluency.	

Table 1. MQM classification adapted from Ortiz (2016: 63-64).

In fact, we are going to concentrate on two of the categories proposed by the model, as we consider them to be the most relevant:

- Accuracy, which refers to the coherence of the text. An accurate translation conveys exactly the same meaning as the original, or at least tries to get as close as possible to the intended meaning. Therefore, the translation will meet this criterion according to its coherence with the original text. In this category we will analyse terminology errors, mistranslations, omissions or additions and untranslated content.
- On the other hand, fluency refers to spelling errors, typographical errors, grammar errors or unintelligible translations.

To carry out the analysis, we have segmented the text into 12 segments. Once the input has been divided, we have prepared a table which contains the output of both translation engines sorted by languages (Italian and English). In this way, the outputs can be easily compared. Moreover, following the model explained in the table above, the errors have been marked with different colours. A colour has been assigned to each category, as it can be seen in the table above. We have decided to classify errors into the most general types of both categories, so that the results of the analysis are more representative. Thus, there are 9 types of errors, each with its corresponding colour.

Besides, to analyse and post-edit the output in English of both MT engines, we have used C-GEFEM, an EN/ES comparable corpus of the ACTRES project (Contrastive Analysis and Translation English-Spanish in its Spanish acronym)[3]. The long-term goal of this project is the active collaboration with agents in the food processing industry, particularly quality wine producers and gastronomic businesses. Thus, C-GEFEM is an EN/ES comparable corpus of dried meats descriptive texts. It contains 245 texts in each language, comprising 70.994 words in English and 34.681 words in Spanish. It has been compiled in order to analyse and to contrast the rhetorical structure as well as the lexicon phraseology of the genre in the mentioned languages. We are going to exploit the English subcorpus of C-GEFEM.

In order to be able to exploit this corpus, we have used AntConc 3.5.7 (Anthony, 2018), a freeware corpus analysis toolkit for concordancing and text analysis. In this way, we can compare the output of the MT engines with terminology and grammatical and syntactic structures of the corpus, which was originally written in English, so it provides

[3] Actres. (2019). Retrieved from https://actres.unileon.es/wordpress/?lang=en

valid and reliable information. An example of a search for concordance is presented in the image below.

Finally, in order to post-edit the output in Italian of both MT engines, we have compiled a corpus of informative texts, DMiT (Dried Meats in Italian), composed by texts from some websites of cured meats from Italy. These texts have been compiled using the criteria and steps stated by Seghiri (2017) and Ortego Antón (2019): searching, downloading, formatting and storing. DMIT is a virtual corpora composed of 10 texts originally written in Italian about dried meats which are very useful to assess lexical parameters of Italian language in this field provided that we are not native Italian speakers, as well as we can take them as a reference to post-edit the output of MT engines.

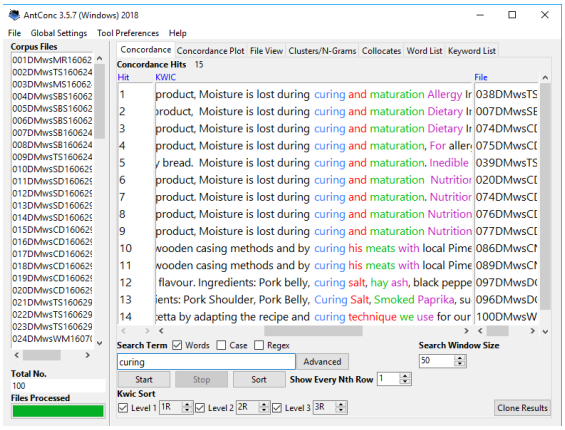


Figure 3. C-GEFEM corpus in AntConc 3.5.7

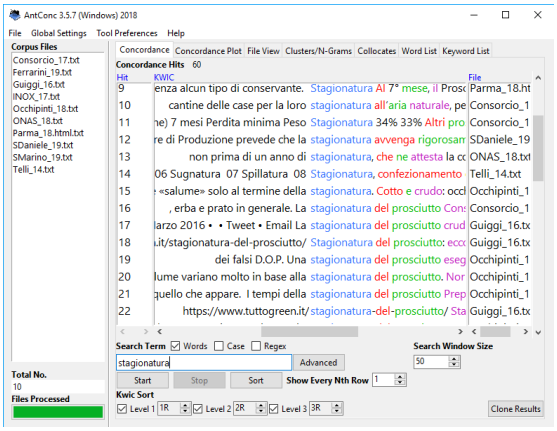


Figure 4. DMiT corpus in AntConc 3.5.7

4. ANALYSIS

Once we have described the methodology and selected the post-edition patterns, we can proceed to analyse and present the results provided by the selection of MT engines using the model described above.

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
La fase de curación del jamón serrano, junto con la de maduración, son el último proceso que se realiza en la elaboración de un jamón.	**EN-1**	**EN-2**
	The curing phase of the Serrano ham, together with the maturing phase, are the last process(es) that is carried out in the elaboration of a ham.	The healing phase of Serrano ham, together with ripening, is the last process(es) that is done in the elaboration of a ham.
	IT-1	**IT-2**
	La fase di stagionatura del prosciutto Serrano, insieme alla fase di stagionatura, sono l'ultimo processo di elaborazione di un prosciutto.	La fase di guarigione del prosciutto Serrano, insieme alla maturazione, è l'ultimo processo che viene fatto nell'elaborazione di un prosciutto.

Table 2. Segment 1

The most significant error in this segment in both languages and MT engines is the translation of "*jamón serrano*", which is not terminologically accurate. In terms of terminology, Microsoft Translator makes an error in the translation of "*curación*", both in English and in Italian. Moreover, there are some grammatical errors in English (errors in the formation of plural due to the syntactic structure). Finally, there are mistranslations in English ("*maturing*" and "*done*") and in Italian ("*l'ultimo processo*" and "*di un*".)

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
Lo más importante de las dos fases es controlar los niveles de temperatura, humedad (que debe ser entre 68% y 76%) y ventilación.	**EN-1**	**EN-2**
	The most important of the two phases is to control the levels of temperature, humidity (which should be between 68% and 76%) and ventilation.	The most important of the two phases is to control the temperature levels, humidity (which must be between 68% and 76%) and ventilation.
	IT-1	**IT-2**
	La più importante delle due fasi è il controllo dei livelli di temperatura, umidità (che dovrebbe essere compresa tra il 68% e il 76%) e ventilazione.	La più importante delle due fasi è quella di controllare i livelli di temperatura, umidità (che deve essere tra (il) 68% e (il) 76%) e ventilazione.

Table 3. Segment 2

In this second segment, there are not significant errors in the English translation. The only error that can be found in Microsoft's option is an inappropriate use of the modal verb "must". Furthermore, DeepL also commits an error in the translation of the modal verb into Italian ("*dovrebbe*".) Microsoft's translation in Italian contains more errors, since there is a grammatical error (absence of the article before the percentage) and an inaccurate translation ("*è quella di controllare*").

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
Ambas fases están separadas por el "pannage" del jamón serrano, la acción de poner una capa de grasa de cerdo en el músculo del jamón.	**EN-1**	**EN-2**
	Both phases are separated by the "pannage" of the Serrano ham, the action of putting a layer of pork fat in the muscle of the ham.	Both phases are separated by the "pannage" of the Serrano ham, the action of putting a layer of pork fat in the muscle of the ham.
	IT-1	**IT-2**
	Entrambe le fasi sono separate dal "pannage" del prosciutto Serrano, l'azione di mettere uno strato di grasso suino nel muscolo del prosciutto.	Entrambe le fasi sono separate dal "pannage" del prosciutto Serrano, l'azione di mettere uno strato di grasso di maiale nel muscolo del prosciutto.

Table 4. Segment 3

In this segment, the only error in both English versions that can be found is the terminological error mentioned in the first segment, the inaccurate translation of "*jamón serrano*." Moreover, in the Italian version, we have found the same inaccurate translation in the output of both engines ("*l'azione di mettere*".)

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
En la fase de curación del jamón se acaba el proceso donde se forman los aromas, pero el jamón sigue perdiendo agua.	**EN-1**	**EN-2**
	In the curing phase of the ham, the process where the aromas are formed is finished, but the ham continues to lose water.	In the curing phase of the ham (,) the process is finished where the aromas are formed, but the ham still loses water.
	IT-1	**IT-2**
	Nella fase di stagionatura del prosciutto, il processo di formazione degli aromi è terminato, ma il prosciutto continua a perdere acqua.	Nella fase di indurimento del prosciutto il processo è finito dove si formano gli aromi, ma il prosciutto perde ancora acqua.

Table 5. Segment 4

There is a common error in the output of both MT engines in both languages: the mistranslation of "*se acaba el proceso*" is inaccurate in the four target texts. In addition, Microsoft's translation into English is the only segment in which we can found a typography error (absence of a comma), together with a mistranslation ("*still loses*".) Last, there are significant grammatical errors in the two outputs in Italian related to word order and syntactic structures.

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
La temperatura óptima para la curación del jamón serrano debe estar sobre los 14ºC, así el jamón adquiere la estabilidad perfecta.	**EN-1**	**EN-2**
	The optimum temperature for curing Serrano ham should be above 14ºC, so that the ham acquires perfect stability.	The optimum temperature for curing Serrano ham should be about 14 º C, so the ham acquires the perfect stability.
	IT-1	**IT-2**

	La temperatura ottimale per la stagionatura del prosciutto Serrano dovrebbe essere superiore a 14ºC, in modo che il prosciutto acquisisca una perfetta stabilità..	La temperatura ottimale per la stagionatura del prosciutto Serrano dovrebbe essere di circa 14 º C, quindi il prosciutto acquisisce la stabilità perfetta.

Table 6. Segment 5

First of all, the terminological error related to the translation of "*jamón serrano*" can also be found in this segment. Moreover, DeepL provides a significant mistranslation due to a misinterpretation of the preposition "*sobre*", both in English and Italian.

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
La humedad debe ser entre 68% y 76%.	**EN-1**	**EN-2**
	The humidity should be between 68% and 76%.	The humidity must be between 68% and 76%.
	IT-1	**IT-2**
	L'umidità dovrebbe essere compresa tra il 68% e il 76%.	L'umidità deve essere compresa tra (il) 68% (il) e 76%.

Table 7. Segment 6

The only error that can be found in Microsoft's English option is an inappropriate use of the modal verb "must." Furthermore, DeepL also commits an error in the translation of the modal verb into Italian ("*dovrebbe*".) Microsoft's translation in Italian contains one significant error, since there is a grammatical error (absence of the article before the percentage.)

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
La curación del jamón serrano se puede realizar en cualquier tipo de secadero: artificial o natural.	**EN-1**	**EN-2**
	Serrano ham can be cured in any type of dryer: artificial or natural.	The cured ham can be done in any type of dryer: artificial or natural.
	IT-1	**IT-2**
	Il prosciutto Serrano può essere stagionato in qualsiasi tipo di essiccatoio: artificiale o naturale.	Il prosciutto crudo può essere fatto in qualsiasi tipo di essiccatore: artificiale o naturale.

Table 8. Segment 7

This segment is full of terminological errors, since we can find again the mistranslation of the term "*jamón serrano*", together with the inaccurate translation of both MT engines of "*se puede realizar*" into Italian. The same mistranslation can also be found in Microsoft's output in English.

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
El secadero artificial está completamente cerrado y controlado por un sistema que ingresa y saca aire.	**EN-1**	**EN-2**
	The artificial dryer is completely enclosed and controlled by an air inlet and outlet system.	The artificial dryer is completely closed and controlled by a system that enters and draws air.

	IT-1	IT-2
	L'essiccatore artificiale è completamente chiuso e controllato da un sistema di ingresso e uscita dell'aria.	L'essiccatore artificiale è completamente chiuso e controllato da un sistema che entra e disegna l'aria.

Table 9. Segment 8

In this segment, we can see a terminological error in Microsoft's version in English, since the translation of the term "*cerrado*" is more accurate in DeepL's version. Furthermore, Microsoft provides a mistranslation of "*un sistema que ingresa y saca aire*" in both languages.

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
El sistema consiste en alternar periodos con aire y sin aire.	**EN-1**	**EN-2**
	The system consists of alternating periods with air and without air.	The system consists of alternating periods with air and without air.
	IT-1	**IT-2**
	Il sistema è costituito da periodi alternati con aria e senza aria.	Il sistema è costituito da periodi alternati con aria e senza aria.

Table 10. Segment 9

In this segment, DeepL and Microsoft provide the same error in Italian: a inaccurate grammatical construction in the translation of "*alternar periodos con aire y sin aire.*"

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
El secadero natural se puede realizar solo si el clima de la región lo permite, con temperaturas poco elevadas y regulares y poca humedad, con alguna variación a lo largo del tiempo.	**EN-1**	**EN-2**
	The natural dryer can only be carried out if the climate of the region allows it, with low and regular temperatures and low humidity, with some variation over time.	The natural dryer can be carried out only if the climate of the region allows it, with low and regular temperatures and low humidity, with some variation over time.
	IT-1	**IT-2**
	L'essiccazione naturale può essere effettuata solo se il clima della regione lo consente, con temperature basse e regolari e bassa umidità, con qualche variazione nel tempo.	L'essiccatore naturale può essere effettuato solo se il clima della regione lo consente, con temperature basse e regolari e bassa umidità, con alcune variazioni nel tempo.

Table 11. Segment 10

In the tenth segment, there is a terminological error in the two English translations ("*secadero*.") This error can be also found in Microsoft's translation into Italian. There is also a misplacement of the adverb "only" in Microsoft's output. Furthermore, a significant inaccurate translation of "*se puede realizar*" can be found in both translations into Italian.

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
Debe estar equipado con mosquiteros, contraventanas, ventiladores y un calentador.	**EN-1**	**EN-2**
	It must be equipped with mosquito nets, shutters, fans and a heater.	It must be equipped with mosquito nets, shutters, ventilators and a heater.
	IT-1	**IT-2**
	Deve essere dotato di zanzariere, tapparelle, ventilatori e riscaldatore.	Deve essere dotato di zanzariere, persiane, ventilatori e un riscaldatore.

Table 12. Segment 11

The most important errors of this segment are of a terminological nature. In English, we can see a bad translation of the term "*ventilador·*" in DeepL's version. Nevertheless, in Italian, the term "*calentador*" is inappropriate in both outputs.

ES (SOURCE TEXT)	DEEPL PRO	MICROSOFT WITH FEEDBACK
Estos secaderos deben ser revisados varias veces al día.	**EN-1**	**EN-2**
	These dryers should be checked several times a day.	These dryers must be checked several times a day.
	IT-1	**IT-2**
	Questi essiccatoi devono essere controllati più volte al giorno.	Questi essiccatori devono essere controllati più volte al giorno.

Table 13. Segment 12

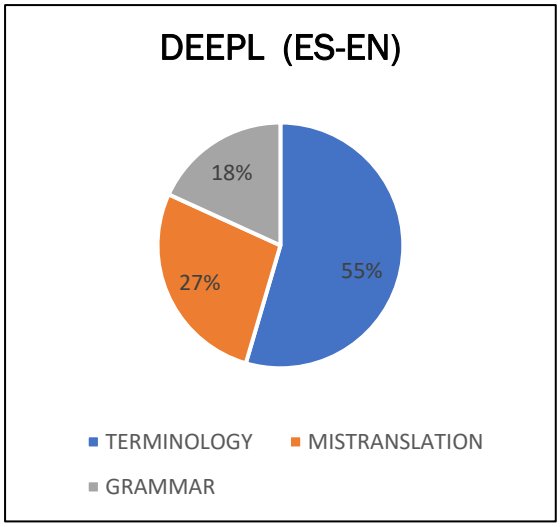
Although the output could be improved with the help of a human translator, as it happens in the majority of the segments, there are no significant errors to be mentioned in this last segment.

4.1. DISCUSSION OF RESULTS

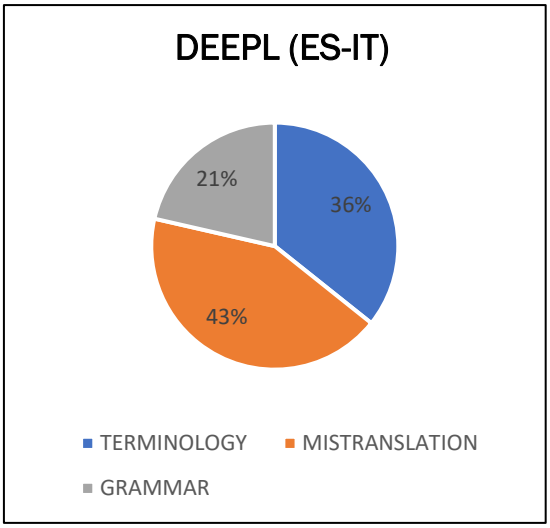
After organizing the data obtained from the different MT engines in the previous tables, we proceed to describe which are the most common errors in each of the engines and depending on the target language.

4.1.1. DEEPL

In the following figures, we present the errors of the neural machine translation engine DeepL. Figure 1 shows the errors in the Spanish-English translation. Moreover, the errors in the Spanish-Italian translation are shown in Figure 2.



Graphic 1. DeepL errors (ES-EN)



Graphic 2. DeepL errors (ES-IT)

As we can see in the figures and following the error classification mentioned above, DeepL only produces three types of errors: two of them are related to accuracy (55 % correspond to terminology and 27 % correspond to mistranslation) and 18 % are related to fluency, that is, grammar errors (18 %). We have not detected other types of errors in the translation of these twelve segments, although it is likely that there would be more errors of different types in a longer text. Thus, the 55% of DeepL errors when translating from Spanish into English are related to terminology. Provided that the text may contain specialised terminology in the field of cured meats, one of the most frequent errors is the translation of "*jamón serrano"* as "Serrano ham." Based on the corpus cited in the methodology, although we found numerous texts with this translation, it is likely that consumers in the Anglo-Saxon context do not understand this term. Therefore, a translation such as "cured ham," which is more accurate in English, would be better.

Another significant error in terminology is the translation of "*secadero natural"* as "natural dryer." In this case (segment 10), the author in Spanish refers to the drying process and not to a machine, so it would be necessary to translate it for "drying."

Furthermore, mistranslation represents 27 % of the errors. Some examples are: the translation of "*maduración"* as "maturing phase" instead of "maturation phase", some errors in the selection of the verb ("is finished," segment 4) and the translation in segment 5 of "*sobre*" as "above" instead of "about."

Finally, the remaining 18% refers to grammatical errors, such as confusions between singular and plural ("is"/ "are" or "process"/ "processes") in the first segment.)

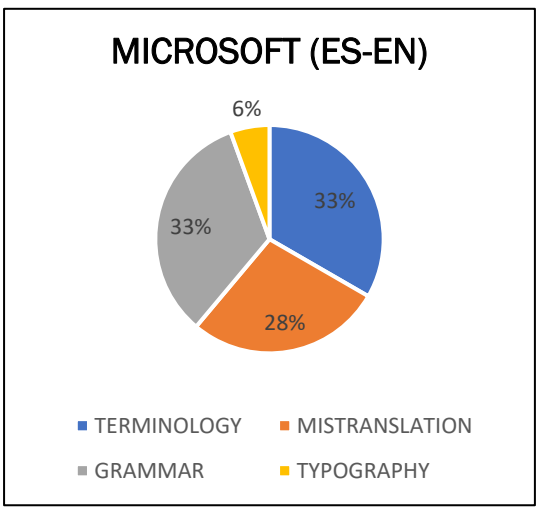
Regarding the translation from Spanish into Italian, terminology errors represent the 36% of the total and some of them are similar to the English errors mentioned above. For example, it is preferred to translate "*prosciutto crudo*" instead of "*prosciutto Serrano"*. Another example is the translation of "*calentador"* for "*riscaldatore*", since it is more accurate to translate "*dotato di riscaldamento.*"

Mistranslation represents also the 43 % of the errors in Italian and an example of this is the translation of "*proceso"* as "*processo,*" which is a false friend that should be translated as "*fase.*" In addition, there are some errors related to the accuracy of the selected verb (in the segment 4, "*è terminato il processo*" would be better translated as "*si conclude il processo*".) There is also in this category an error which can be compared to English: the translation of "*sobre"* as "*superiore*" in the segment 5, which should be translated as "*di circa.*"

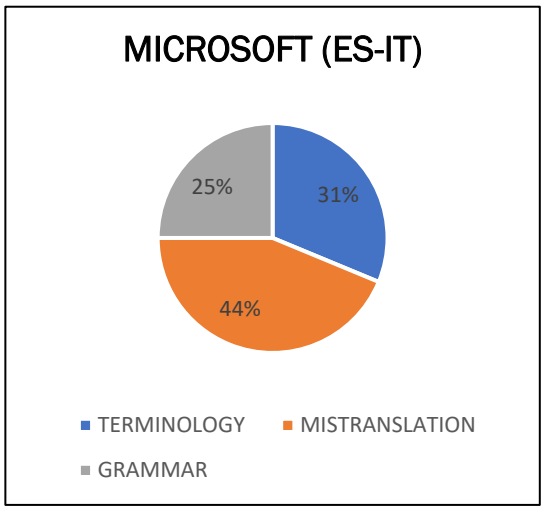
Lastly, the remaining 21 % of the translation from Spanish into Italian refers to grammar errors, as we can see in the segment 9: "*consiste en alternar periodos con aire y sin aire"* is translated as "*costituito da periodi alternati con aria e senza aria,*" and taking into account the Italian grammatical rules, it would be better translated as "*alterna periodi con aria e periodi senza aria.*"

4.1.2. MICROSOFT TRANSLATOR

In the following figures, we present the errors of the neural machine translation engine Microsoft Translator. Figure 3 shows the errors in the Spanish-English translation. Moreover, figure 4 shows the errors in the Spanish-Italian translation.



Graphic 3. Microsoft errors (ES-EN)



Graphic 4. Microsoft errors (ES-IT)

In this fragment translated by Microsoft from Spanish into English, we have detected four types of errors. Two of them are related to accuracy: terminology (33 %) and mistranslation (28 %). Moreover, the remaining two types are grammar (33 %) and

typographical errors (6 %). Thus, terminological errors are one of the largest percentages (33 %) when translating from Spanish into English. An example of this is the translation of "*jamón serrano"* as "Serrano ham". Furthermore, one of the most serious errors in this fragment is the translation of "*curación"* as "healing" in the first segment, since the English verb "to heal" is only used in the health context.

Moreover, mistranslation also represents the 33 % of the errors when translating into English. For instance, in the segment 8, the MT engine selects the participle "closed" but in this context it would be more accurate to use "enclosed". Another significative error can be found in the eighth segment, in which "*un sistema que ingresa y saca aire"* has been translated overly literal as "a system that enters and draws air". In this case, a good alternative could be "an air inlet and outlet system."

The only typographical error that can be found in our study is located in the translation into English of the segment 4, since we have detected the absence of a comma to differentiate the main clause and the subordinate clause. Besides, grammatical errors are basically confusions between singular and plural, as we can see in the first segment ("is"/"are" or "process" or "processes.") In addition, there are some errors related to the use of modal verbs ("must be" instead of "should be" in the segment 2) and one error which refers to word order (location of the adverb in the segment 10.)

Regarding to the translation from Spanish into Italian, mistranslation represents the 44 % of the errors. Furthermore, we have detected some terminological errors (31 %) and grammar errors (25 %).

As it is said above, mistranslation occupies the largest percentage in Microsoft's translation into Italian (44 %). One of the errors that are likely to be mentioned is the use of the term "*processo*" instead of "*fase*" in the first segment, or the use of inaccurate verbs to describe processes (in the segment 4, "*è terminato il processo*" would be better translated as "*si conclude il processo*.") Alongside to these errors, we would like to highlight the mistranslation of the segment 2 ("*controlar los niveles de temperature"* has been automatically translated as "*è quella di controllare i livelli*" and it should be better translated as "*è il controllo di*.") Moreover, a mistranslation can be found in the eighth segment, since "*sistema que ingresa y saca aire"* has been translated by Microsoft as "*sistema che entra e disegna l'aria*", which is overly literal. A good alternative to solve this issue is "*sistema di ingresso e uscita dell'aria*."

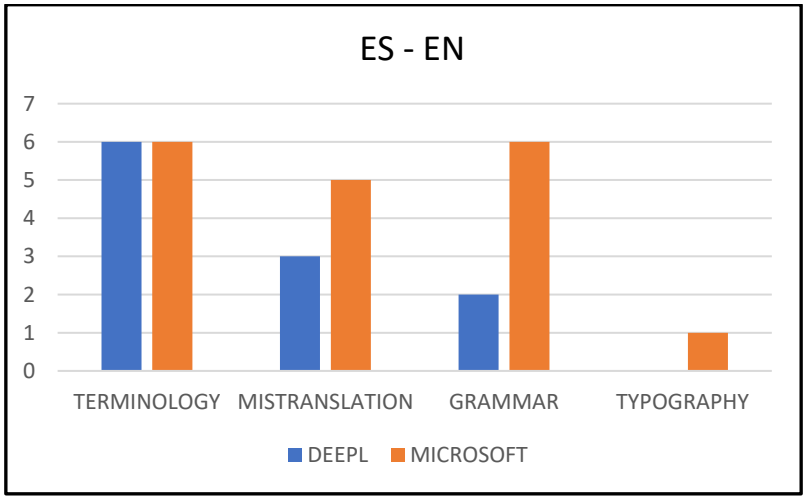
Besides, terminological errors in Italian represent the 33 % of the total. Among the most outstanding errors we can highlight the translation of "*curación"* as "*guarigione*" in the first segment, since the Italian verb "*guarire*" is only used in the health context and a good alternative could be "*stagionatura*." In addition, the translation of "*secadero natural"* as "*essicatore*" is not accurate since the autor in Spanish refers to the drying process and not to a machine, so it would be necessary to translate "*essiccazione*". There are also some errors due to the cultural content of the text: the use of "*prosciutto Serrano*" instead of "*prosciutto crudo*" throughout the text. Finally,

"*calentador"* in the segment 11 would be better translated for "*dotato di riscaldamento*" instead of "*riscaldatore*".

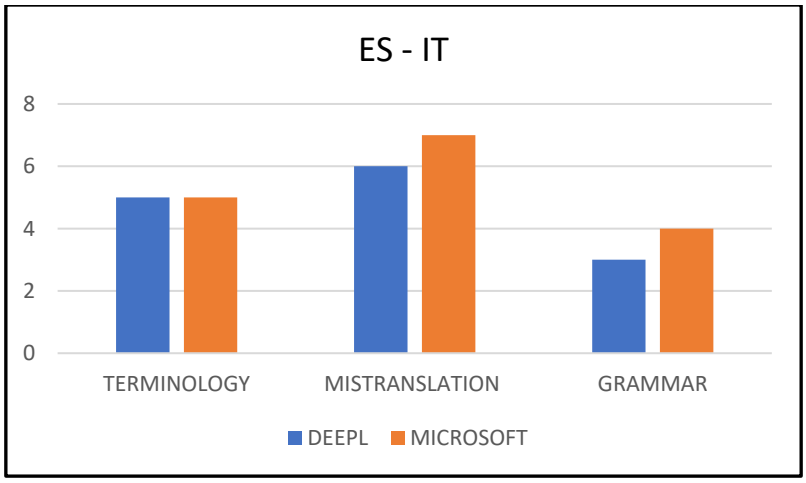
Finally, grammar mistakes (25 %) are also significative in this linguistic combination. For instance, in the last part of the ninth segment, "*consiste en alternar periodos con aire y sin aire"* is translated for "*è costituito da periodi alternati con aria e senza aria*", following the order of the Spanish structure. An accurate alternative could be "*alterna periodi con aria e periodi senza aria*." Furthermore, there is a significant error that can be appreciated in the fourth segment and it is related to word order: "*el proceso donde se forman los aromas"* is translated for "*il processo è finito dove si formano gli aromi*" because the MT engine has followed the Spanish order and syntactic structure. A proper alternative for this part of the segment could be "*si conclude il processo di formazione degli aromi*."

4.2. COMPARISON OF RESULTS

In the following two figures, all the errors of both translation engines according to its language combination are brought together. This comparison serves to evaluate which translation engine is suitable for each language combination depending on the type of error that is given priority and in which cases it would be better to use one engine or another depending on the target language.



Graphic 5. ES - EN errors



Graphic 6. ES - IT errors

In Figure 5, we can see that both translation engines commit the same number of terminological errors, and this is due to the difficulty that MT engines have in achieving precision in certain areas of specialised translation due to the lack of adequate terminology. However, the graphs show that Microsoft produces more mistranslation errors and grammar errors besides it is the only one of the two engines that produces typographical errors.

On the other hand, Figure 6 shows nearly the same results depending on the type of error, but the differences are much smaller, which shows that these translation engines provide similar results when translating into Italian, despite the significant difference between them when translating into English.

5. CONCLUSIONS

In the first place, we have obtained an overview of the main techniques of machine translation and post-editing in the background. Thus, we can state that machine translation can offer results that would serve as a starting point for subsequent post-editing, which is necessary to obtain an optimum result, since machine translation engines do not provide perfect translations.

Secondly, once the analysis has been carried out, we can draw conclusions about the functioning of the MT engines in general and the functioning of each of them depending on the target language. So, after analysing each segment, we have come to the conclusion that there are many prejudices against machine translation because, before analysing the output, I thought that the results would be worst. Although post-editing is necessary, the overall result of the engines is not so bad since neural technology has been implemented. However, a human post-editing is necessary, so this project also highlights the important value of machine translation and the importance of training translators in post-editing. In addition, we can say that the analysis of the errors produced by the engines has been of great difficulty, and this is related to the knowledge of the language and the need to be an expert in the field in the working languages of the translator.

Moreover, at a more specific level, we have obtained the conclusion that DeepL offers better results than Microsoft Translator, despite the two of them are neural translation engines recently developed. On a terminological level, both engines produce almost the same amount of errors. However, when it comes to grammar, typography and fluency, DeepL provides better results.

In addition, there are differences depending on the target language. Due to the similarity in syntactic and grammatical structure between Italian and Spanish, the result of this linguistic combination is better. However, the number of terminological errors does not vary regardless of the target language. This is due to the lack of cultural content in the target languages, since in the field of cured meats, many concepts are words and expressions for culture-specific elements.

In addition, the classification of patterns to post-edit shows that a huge part of the errors committed by these tools correspond to translation errors derived from literalism.

In short, this project gives rise to future lines of research, since it would be interesting in the future to test whether neural translation engines evolve as improvements in neuronal technology are implemented. It would also be useful to investigate other fields of specialisation in order to draw different conclusions depending on the field.

6. REFERENCES

Anthony, L. (2018). *Laurence Anthony's AntConc.* Retrieved from: https://www.laurenceanthony.net/software/antconc/ [Consulted on 1st April 2019]

Arnold, D. (2003). "Why translation is difficult for computers." In: Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies: Volume 1*. Amsterdam/Philadelphia: John Benjamins, 60-65.

Bahdanau, D., Kyunghyun, C., Bengio, Y. (2014). *Neural Machine Translation by Jointly Learning to Align and Translate.* Notas de la ponencia. Retrieved from: https://arxiv.org/abs/1409.0473 [Consulted on 22nd May 2019].

Bowker, L., & Fisher, D. (2010). "Computer-aided translation". In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies: Volume 1*. Amsterdam/Philadelphia: John Benjamins, 60-65.

Carl, M. & Way, A. (Eds.). (2003). *Recent advances in example-based machine translation*. In: Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies: Volume 1*. Amsterdam/Philadelphia: John Benjamins, 60-65.

Casacuberta Nolla, F. & Peris Abril, A. (2017). "Traducción automática neuronal". *Tradumàtica. Tecnologies de la Traducciò, 15*, 66-74. https://doi.org/10.5565/rev/tradumatica.203

Forcada, M. (2010). "Machine translation today". In Gambier & L. van Doorslaer (Eds.). *Handbook of Translation Studies: Volume 1*. Amsterdam/Philadelphia: John Benjamins, 215-220.

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

Hymes, D. H. (1974) "Models of the interaction of language and social life". In J. J. Gumperz & D. Hymes (Eds.). *Directions in Sociolinguistics: The ethnography of Communication*. New York: Holt, Rinehart & Winston, 10.

ISO (2017). *International Organization for Standardization. ISO 18587:2017 Translation services -- Post-editing of machine translation output – Requirements*. Retrieved from: https://www.iso.org/standard/62970.html [Consulted on 19th May 2019].

Koehn, P. (2009). *Statistical machine translation.* Cambridge: Cambridge University Press.

MAPA (2019). *Marco Estratégico para la Industria de Alimentación y Bebidas*. Retrieved from https://www.mapa.gob.es/es/alimentacion/temas/industria-agroalimentaria/marco-estrategico/ [Consulted on: 15th May 2019].

Ortiz, C. (2016). *Implementing Machine Translation and Post-Editing to the Translation of Wildlife Documentaries through Voice-over and Off-screen Dubbing*. Doctoral thesis. Barcelona: Universitat Autònoma de Barcelona.

Ortego Antón, M. T. (2019). *La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos.* Bern: Peter Lang.

Rivas Carmona, M. & Veroz González, M. (2018). *Agroalimentación*: *Lenguajes de especialidad y traducción.* Granada: Comares, 15-31.

Sánchez-Gijón, P. (2016). "La posedición: hacia una definición competencial del perfil y una descripción multidimensional del fenómeno". *Sendebar*, 27, 151-162.

Seguiri, M. (2017). "Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores." *Babel*, 63 (1), 43-64.

Tatsumi, M. 2010. *Post-Editing Machine Translated Text in A Commercial Setting: Observation and Statistical Analysis.* Doctoral Thesis. Dublin: DCU.

Torres-Hostench, O., Presas, M., & Cid-Leal, P. (Coords.). (2016). *El uso de traducción automática y posedición en las empresas de servicios lingüísticos españolas: Informe de investigación ProjecTA 2015*. Barcelona: Universitat Autònoma de Barcelona. Retrieved from: https://ddd.uab.cat/record/148361 [Consulted on 1st May 2019].

Viver Sorolla, P. (2018). *La evaluación de las herramientas de traducción automática (TA) desde la perspectiva del traductor: Google Translate, Bing, Babylon y Systran*. TFG. Valladolid: Universidad de Valladolid. Retrieved from: http://uvadoc.uva.es/handle/10324/33981 [Consulted on 22nd April 2019].