

The eternal antagonism between humans and technology

A study on machine translation

Enya Fernández Muñoz
University of Córdoba

Received: 13/03/2020

Accepted: 01/07/2020

Abstract

This paper is about a controversial issue in recent times: machine translation. First of all, we will present the definitions of basic concepts such as machine translation and machine translation programme. Afterwards, we will expose the process that is carried out to perform a human translation and a machine translation, all followed by the limits of machine translation, the utility of the text corpus, the post-editing process and the objectives, purpose, functions and tools of machine translation programmes. In addition, we will also comment on the different translations that can be obtained with these translation systems depending on the type of the text that we intend to translate and we will present a brief reflection about the positive and negative points that machine translation has in order to reach a conclusion about whether it is suitable or not to get high quality translations. Finally, we will carry out a practical analysis in which several sentences will be translated through different machine translation programmes and prepare a comparative table in which we can clearly see the differences, similarities and errors provided by the different programmes.

Key Words

Machine translation, human translation, text corpus, post-editing, quality.



1. Introduction

How many times have we heard that Machine Translation engines will end up replacing human translators? It is increasingly common that many people – and some companies – decide to use free online Machine Translation engines

instead of hiring the services of a human translator. Sometimes, this is because they do not know our work, but most of the time this happens because it is faster and cheaper.

This paper aims to raise awareness of some Machine Translation engines, as well as the problems, advantages and disadvantages of this type of translation. In addition, the practical analysis that we will develop will allow us to discover the utilities of these translators and will help us to answer questions as to whether the work of the human translator will pass away or if Machine Translation engines will replace us.

We believe that the best way to know how a program or application works is to test it. So, to answer our questions, we will study all our premises by analysing the translations obtained through different Machine Translation engines and we will prepare a comparative table in which we can clearly observe the differences, similarities and errors provided by the different programs. To do this, we will translate from English to Spanish four sentences that contain different verbatim records, some expressions, etc. with different Machine Translation engines and we will analyse and compare the results obtained.

2. Theoretical framework

2.1. Machine translation

Machine translation is a form of language processing in which its ultimate goal is the achievement of a translation by telematic systems; that is, without the active participation of a human. We can define automatic translation as a fully automatic processing, performed by a computer that uses appropriate software, of a text written in a source language that produces another text in a target language, which can sometimes be used as raw translation. In other words, machine translation is achieved through computers or machines suitable for this purpose.

Consequently, a Machine Translation engine is a system that uses the information of the source language to break it down into syntactic and morphological units and analyse them to compose the text in the target language. The machine translation process does not require human

intervention, since the automatic translator replaces the terms in the source language with their equivalent in the target language.

2.1.1. *Differences between human translation and machine translation*

With regard to the translation process performed by a human, at first, translating a word from one language to another may seem simple. However, translation does not only consist of this, but also requires many other processes that go beyond a simple word-by-word translation. Thus, the human translator must not only know the working languages in depth, but also must have basic knowledge of translation that will take into account the different aspects that arise during the translation process.

Although it was clear that there are different types of texts and, therefore, different types of translations, we can establish a series of general steps that a human translator must perform to develop a quality translation. However, as we have just indicated, these steps may undergo changes depending on the text you are working with.

There are several authors who try to establish these steps. García Figuerola et al. (1996: 60-61), in his work *Consideraciones sobre el desarrollo de un sistema experto para la traducción automatizada*, establishes two major steps. Firstly, an understanding of the text through reading and searching for the meaning; and, secondly, a reconstruction of the meaning in the target language. However, other authors such as Gouadec (2007: 57-81), establish these steps from the first moment of the assignment, thus including other steps outside the translation process itself.

After knowing some of the classifications of the different authors, we can establish a series of steps to carry out a translation. The first step is the understanding of the text. To do this, the human translator must make as many readings as necessary to achieve a full understanding of the text. The second step would be to make a new reading to identify possible translation problems, whatever type they are (terminological, syntactic, cultural, meaning, etc.). In the third step, we would proceed to solve these possible problems with the help of monolingual dictionaries, bilingual dictionaries, encyclopaedias, glossaries or even consulting an expert to find out the meaning of some terms or concepts. Then, a first translation would be done respecting the meaning of the original text and using the necessary translation techniques to create a clear and understandable text in the target language. Finally, the text will be reviewed to

ensure that there are no mistakes and, if necessary, to make changes that help to improve the understanding of the text in the target language.

We can conclude that these steps can be used for any type of text. However, we must always take into account certain aspects depending on the type of text which one works with and the previous knowledge of the human translator on the area.

On the other hand, machine translation involves a quite different process. Firstly, the original text to be translated must be pre-edited by a Machine Translation engine in order to simplify syntax and grammatical structures, eliminate typing and spelling mistakes, etc. Once we have the pre-edited text, we will introduce it in the software that will carry out the automatic translation. Finally, when we have already obtained the automatic translation, we must post-edit the text if we want to obtain a quality translation, since Machine Translation engines stand out for translating word by word and, sometimes, the meaning of the text varies considerably. In addition, the quality of the translations provided by Machine Translation engines is often less than satisfactory, as they include terminological mistakes, false senses, untranslated elements, repetitions, changes in the word order, incorrect verb tenses, poor punctuation, etc. Due to these mistakes, the human translator must review and post-edit the text in depth, a task that can sometimes take longer than it would have taken to translate the text manually.

2.1.2. Limits of machine translation

Once both the concept and the machine translation process are clear, a series of language and processing features should be taken into account in order to face the difficulties and limitations that may appear in this process. The language is not as simple as it might seem. Although knowing the lexicon, grammar and syntax of it is the first step to be able to understand it in depth and to be able to transfer it through lists of rules in its different aspects or equivalences with the desired target language, the language is also characterized by containing other features and tools such as double meanings, idioms, puns, etc. which can severely distort the meaning of a segment and lead to serious misunderstandings.

In this way, machine translation implies that the software not only handles the different interpretations and uses of the different lexical, grammatical and syntactic components of the pair of languages, but must explicitly understand

the way in which it uses each one of them just as a native person would use it, who uses the language with implicit information that he has acquired naturally during his learning.

In conclusion, the language has complex implicit and explicit tools, so that the processing of all this information adapted to the corresponding cultural and idiomatic context is limited to the point that neither experts in the field can affirm that “one day can be reached high quality fully machine translation” (Alcina, 2010-2011: 1).

2.1.3. Versions of translated texts produced by machine translation systems

To approach this issue, it is necessary to know that machine translation is based on different translation systems that will vary the results depending on the used one. Thus, we can distinguish:

1. Machine Translation System that produces word-by-word translations, looking at the syntactic rules of the text and using bilingual dictionaries. That is, these are systems that attempt to transfer the syntactic structures from one language to another. This is the case of machine Translation based on rules and transfers.
2. Machine Translation System that is based on Translation memories or corpora. These systems look for similar contexts that allow them to get closer to a successful translation. This is the case of machine translation based on corpora, statistics, examples, context and neural networks.

In short, the different versions of the translated texts will depend on the type of translation system that the Machine Translation engine use. In this way, you can get literal translations word-by-word or more contextual translations where you don't need so much post-editing.

2.1.4. Utility, advantages and disadvantages of machine translation

At this point, we must take into account that there are two positions from which the advantages and disadvantages of machine translation can be assessed: the point of view of professional translators and the view of people outside of translation.

As a person outside of translation (as a company, for example), the main advantages that machine translation programs can have are the cost savings involved in translations and the speed of reception of the translated text, as they would not have to hire a professional translator or pay their fees and they would get a translation immediately. However, the exclusive use of machine translation programs would provide them with dubious quality results due to the mistakes presented by those programs.

On the other hand, if we talk as professional translators, as it has been described previously, machine translation entails a translation process almost more laborious than human translation, since the original text has to be pre-edited before entering it into the Machine Translation engine, we must post-edit the text if we want to obtain a quality translation, review it in depth, etc. For this reason, what at first may seem to favour us in terms of saving time, it may turn against us and cause us more problems.

Conversely, with regard to their utility, Machine Translation engines can be useful for translating texts that do not require creative translations, have a very simple syntax and do not contain a language crowded with cultural elements. However, if we intend to provide a quality translation, machine translation programs are not suitable for translation. For example, we would not use them for a literary translation, but they could be useful for very technical texts with simple syntactic structures, although the translation should always be accompanied by a reviewing and a post-editing process. We also consider them useful for those translations that do not have to be published later but only intended to show us the meaning of the text or provide us with the key ideas about its content, and therefore they can contain mistakes. For example, to have a general idea on the subject of an email received in a company.

2.2. Post-editing process

After learning about the operation and limitations of machine translation, we observe that human intervention is often necessary after machine translation. This activity is called post-editing and to carry it out, a language professional—translator or linguist—review the text prepared by the Machine Translation engine in order to obtain a final text coherent, without mistakes and easy to understand for the target reader.

The language professional must know in depth the working languages and their main characteristics, but also those aspects that the software cannot reach, such as double senses, puns, etc. In this reviewing process, the professional will

have to modify everything that produces serious mistakes, such as false senses or nonsenses, but also must make changes in the syntactic structure or in another level of the language to create a natural and coherent text in the target language and that can be easily understood.

In some cases, it happens that the post-editing process is more expensive and arduous than the translation process itself because the language professional may take longer to correct the mistakes than to manually translate the text. However, there are other cases in which machine translation is not entirely incorrect and the post-editing process is not so laborious, so that a translation can be obtained quickly.

Originally, machine translation emerged with the idea of allowing communication between speakers of different languages as quickly as possible. In addition, it was opted for creating fast and effective translations. However, as said before, human language is not so simple, and the translation process is not just about translating word by word.

So far, machine translation has failed to interpret those aspects of human language that go beyond significant-meaning, therefore, on several occasions, the results are translations of low or very low quality and post-editing is necessary. For this reason, the process of machine translation accompanied by the post-editing one facilitates the creation of translations quickly. Machine translation provides, in no time, a translation of a source text, a process faster than human translation. However, in many cases this translation is not adequate or presents too many mistakes, hence the need for post-editing.

Finally, we can conclude that this process is carried out to achieve a greater number of translations in less time. Machine translation and post-editing are the perfect combination for this, since by itself the machine translation cannot guarantee, in most cases, a quality translation, therefore, post-editing is our best ally. Thus, we observe that despite the advances that machine translation has brought about, human translators are still essential in this work.

2.3. Textual corpora

Machine translation has been used since its creation from different strategies based on linguistic rules to achieve the most appropriate translation.

From the nineties, one of the tools that were developed were those based on textual corpora. Textual corpora are documents that collect information of

translations already made, so it allows to machine translation system to have very valuable information from which it can generate the new translation. Textual corpora are used to compare the text that has to be translated with the stored texts and, from that comparison, the new translation is developed.

The strategies used by machine translation system based on corpus are the statistical strategy and the example strategy. On the one hand, the statistical strategy is based on the old texts stored to create a new translation. In this way, different translations of the same segment can be developed, but only one of them will be the most statistically correct. On the other hand, the example strategy is based on the search and extraction of coincidences between the text to be translated and the bilingual texts available in the textual corpora, and the combination of these coincidences for the creation of a new translation.

In conclusion, textual corpora provide great information and bilingual texts to the machine translation tools with which they can rely to develop new translations to the new source text.

3. Practical case

3.1. Analysis

Once the theoretical part about machine translation is exposed, then we will try to show the use and utility of the Machine Translation engines. For that purpose, first of all, we will present the sentences that we are going to use and highlight in bold the words or expressions that we believe could be mistakes in the translation:

1. El amigo de María no ha salido de la cama en todo el día, pues es muy perezoso y **duerme como un tronco**.
2. El **tronco** de María no ha **salido del saco** en **to el día** pues **es muy perro**.
3. Juan **está en el Ajo** pues su tío Vicentito y su amiga Marieta le han explicado el asunto y está dispuesto a **coger el toro por los cuernos**.
4. Si llego a ver al **pendejo** que me robó el **cel** le voy a **partir su madre**.

We present below a comparative table in which you can see the different Machine Translation engines used and the English translation they have provided for our four sentences.

MT engine	Sentence 1	Sentence 2	Sentence 3	Sentence 4
AltaVista – Babel Fish Translator	Maria's friend hasn't gotten out of bed all day, because he's very lazy and sleeps like a log.	Maria's trunk has not come out of the sack in the day because it is very dog.	Juan is in the Garlic because his uncle Vicentito and his friend Marieta have explained the matter to him and he is willing to take the bull by the horns.	If I ever see the asshole who stole my cel, I'm going to break his mother.
Ajax translator	Mary's friend has not been out of bed all day, for he is very lazy and sleeps like a log.	Mary's trunk has not come out of the sack in the day because it is very dogy.	John is in the Garlic because his uncle Vicentito and his friend Marieta have explained the matter to him and is willing to take the bull by the horns.	If I get to see the asshole who stole my cel, I'm going to break his mother.
Google Translate	Maria's friend has not gotten out of bed all day, because he is very lazy and sleeps like a log.	Maria's trunk has not left the bag for the day because it is very dog.	Juan is in the Garlic because his uncle Vicentito and his friend Marieta have explained the matter and are willing to take	If I get to see the asshole who stole my cell, I will leave his mother.

			the bull by the horns.	
ImTranslator	Maria's friend has not gone out of the bed in the whole day, since it is very lazy and sleeps as a trunk.	Maria's trunk has not gone out of the sack in to the day since it is much a dog.	Juan is in the Garlic since his uncle Vicentito and his friend Marieta they have explained to him the matter and it is ready to take the bull for the horns.	If I go so far as to see the pendejo that stole the cel from me I am going to destroy him.
Babylon Translation	The friend of Mary has not come out of the bed all day, because it is very lazy and sleeps like a log.	The trunk of Mary has not come out of the bag in the day because it is very dog.	John is in garlic because his uncle Vicentito and her friend Marieta have explained the matter and is willing to take the bull by the horns.	If you came to see the asshole who stole the cel i am going from his mother.
Translate.eu	Maria's friend hasn't gotten out of bed all day, because he's very lazy and sleeps like a log.	Mary's trunk has not come out of the sack in the day because it is very dogy.	John is in the Garlic because his uncle Vicentito and his friend Marieta have explained the matter to him and is willing to take the bull by the horns.	If I get to see the asshole who stole my cel, I'm going to break his mother.
WordLingo	The friend of Maria has not left the bed in	The trunk of Maria has not left the	Juan is in Garlic then his Vicentito uncle	If I get to see the asshole that

	all the day, because he is very sluggish and to duer to me like a trunk.	coat in to the day because it is very dog.	and their Marieta friend has explained him the subject and is arranged to take the bull by the horns.	robbed cel to me I am going to him to divide its mother.
DeepL	Mary's friend has not been out of bed all day, for he is very lazy and sleeps like a log.	Mary's trunk has not come out of the sack in all day because it is very doggy.	Juan is in Ajo because his uncle Vicentito and his friend Marieta have explained the matter to him and he is willing to take the bull by the horns.	If I see the asshole who stole my cel I'm going to split his mother.

3.2. Results

Finally, we will present the differences, similarities and mistakes that we find in the different translations.

With regard to the first sentence, as we had thought at first, the biggest problem would be the phrase *dormir como un tronco*. To our surprise, six of the eight Machine Translation engines used have translated it correctly (sleeps like a log), while the other two have provided a different translation: on one hand, ImTranslator has translated the phrase word-by-word (sleeps like a trunk) and, on the other hand, WordLingo has proposed a nonsense translation (to duer to me like a trunk).

In addition to the problem of this phrase, we have found other differences regarding the proposed translations. These differences are:

- To translate *no ha salido de la cama*, five different proposals appear and two of them are wrong: *be out of* and *come out of*, proposed by Ajax and Babylon respectively.

- Although proper names should not be translated, Ajax, Babylon and DeepL have translated it and included *Mary* instead of *Maria* in their proposal.
- WordLingo also makes a grammatical mistake in the phrase *en todo el día* and translates it as **all the day* instead of *all day*.
- Lastly, ImTranslator and Babylon confuse the subject of the second sentence, so they propose **it is very lazy* instead of *he is very lazy*.

Regarding the second sentence, there were four words or expressions that we thought were going to be problematic since, despite having the same meaning as the first sentence, it uses a much more vulgar language (*tronco* instead of *amigo*, *salir del saco* instead of *salir de la cama* and *to* instead of *todo*). In addition, it contains the idiom *ser un perro*. This time, we found the following mistakes:

- Firstly, no Machine Translation engine interprets a different meaning for the term *tronco* and they unanimously translate it literally as *trunk*. Likewise, they do not interpret the colloquial expression *salir del saco* as *salir de la cama*, so they propose translations like *sack*, *bag* and *coat*.
- Regarding the translation of proper names, Ajax, Babylon and DeepL have translated it in the first sentence, while Translate.eu has also translated it in the second sentence.
- In the first phrase, all Machine Translation engines use a different verb as a proposal for the verb *salir*, except ImTranslator, Babylon and WordLingo, which maintain the same verb that they used in the first sentence (*go out*, *come out* and *left* respectively).
- On the other hand, regarding the colloquial expression *to el día*, it has gone unnoticed and has been correctly understood by six Machine Translation engines, but ImTranslator and WordLingo have not understood it and have left the term *to* in Spanish. So, they propose the following wrong translation: **in to the day*.
- Finally, in relation to the expression *ser muy perro*, no Machine Translation engine has interpreted it as a synonym for *ser vago* and all of them have proposed a literal translation word-by-word. However, in this sense, we are not only faced with a misunderstanding, but there are also two repeated mistakes:
 - o All Machine Translation engines have changed the subject, so all their proposals are: **it is...*

- Ajax and Translate.eu have used the wrong term **dogy* instead of *doggy* or *dog*.

In the case of the third sentence, we started from the premise that there would only be two expressions that would give problems: *estar en el Ajo* and *coger el toro por los cuernos*. In the case of the first one, not only due to the fact that it is an idiom, but also because we feared that the term *ajo* would be taken as a proper name as it was capitalized.

Regarding the idiom *estar en el ajo*, it can be translated into English as *to be in on* or *to be in the loop*. As we can see in the table, no Machine Translation engine has translated this idiom correctly, but all of them except DeepL have translated the idiom *estar en el ajo* as **be in (the) garlic*. In addition, of these seven Machine Translation engines, all of them except Babylon have interpreted it as proper name, so they capitalized it. By contrast, DeepL has understood the term *ajo* as proper name and has kept it untranslated and capitalized.

On the other hand, with respect to the idiom *coger el toro por los cuernos*, it has been translated into English as *take the bull by the horns*. As we can see, this idiom has been translated word-by-word, so all the Machine Translation engines used have provided a correct translation except ImTranslator, which fails in the preposition and proposes **take the bull for the horns*.

Concerning the translation of proper names, Ajax, Babylon and DeepL had translated it in the first sentence; Translate.eu translated it in the second sentence; and in the third one, DeepL maintains the name in Spanish, so it is only translated as *John* by Ajax, Babylon and Translate.eu. Instead, the names *Vicentito* and *Marieta* remain at all times in Spanish.

Finally, we find some mistakes in the order of the words or their concordance, such as:

- WordLingo alters the logical order of the noun and adjective and proposes **his Vicentito uncle and their Marieta friend*.
- Babylon makes a concordance mistake with Juan's friend and translates **her friend* instead of *his friend*.
- Google Translate makes another concordance mistake in the phrase *quiere coger el toro por los cuernos*, because it conjugates it in plural, so we find **and are willing to take...*

Lastly, the fourth sentence includes specific terms of Latin American Spanish and, as we said above, machine programs do not usually differentiate between

European Spanish and Latin American Spanish, so, in the case of this specific sentence, we believe that the highlighted terms will not be recognized correctly. First of all, all the Machine Translation engines used have translated the term *pendejo* as *asshole*, except ImTranslator, which keeps it in Spanish. In this sense, we were confused thinking that it was going to be a problematic term and that they were not going to translate it correctly. Regarding the term *cel*, all Machine Translation engines maintain it as *cel*, except Google Translator, which translates it as *cell*, so this time, none of them have understood that this term *cel* referred to the *phone*, so none of them have proposed a correct translation. Finally, the expression *partir la madre* could be considered as a synonym of *partir la cara*. Although they use different expressions, they all made a literal word-by-word translation, so this expression has not been understood by any of the Machine Translation engines used and they have all provided a wrong translation proposal.

4. Conclusions

Firstly, if we look at the results obtained from the analysis carried out, we can affirm that these Machine Translation engines are not suitable for quality translations since they do not detect idioms or colloquial language, they use to translate word-by-word and they make grammatical and concordance mistakes. Regarding the concordance mistakes, we were particularly struck by the proper names' translation, since not only some Machine Translation engines translate them and others keep them in the original language, but the same Machine Translation engines sometimes translate proper names and at other times keep them in the original language.

Secondly, regarding the utility of Machine Translation engines, as has been mentioned, Machine Translation engines can be useful to understand texts in other languages or to help professional translators to translate faster, but they are not useful to perform a complete translation without post-editing or reviewing the text later.

On the other hand, regarding the quality of the translations obtained, this could be improved by making changes to the original text using some techniques that simplify syntax and language, such as simplify syntactic structures by using compound sentences instead of simple sentences, avoid ambiguous terms, or do not use the gerund or the passive.

In addition, if you do a good post-editing and reviewing to the translation obtained by Machine Translation engines, you could get a quality translation. For this reason, machine translation and post-editing are the perfect combination to achieve a greater number of translations in less time, since machine translation cannot guarantee, in most cases, a quality translation, therefore, post-editing can be a good ally.

All this leads us to deduce that, despite the many advances that machine translation has brought to the world of translation, as long as the translation system used by Machine Translation engines does not change, human translators will continue to be essential in this work and Machine Translation engines will not be able to replace human translators or take away their job.

References

- Alcina Caudelt, M. A. (2010-2011) *Los traductores automáticos en la red. El español en el mundo*. Madrid. Ed. Instituto Cervantes. 2011. Científic.
- García-Figuerola Paniagua, C.; Domínguez Ollero, A.; Alonso Berrocal, J. L. et al. (1996) Consideraciones sobre el desarrollo de un sistema experto para la traducción automatizada. *Trans.* N.º 1, pp. 64. Available on: http://www.trans.uma.es/pdf/Trans_1/t1_59-73_CFigueroa.pdf
- Gouadec, D. (2007) *Translation as a Profession*. Philadelphia: John Benjamins Publishing Company.
- Parish, J. (n.d.) Machine Translation engine Ajax Translator. Available on: <http://www.ajaxtrans.com/>
- Smart Link Corporation (2019) Machine Translation engine ImTranslator. Available on: <http://www.imtranslator.com/>
- Various Authors (BabelFish) (2019) Machine Translation engine AltaVista – Babel Fish Traducción. Available on: <https://www.babelfish.com/>
- Various Authors (Babylon) (2017) Machine Translation engine Babylon Translation. Available on: <https://traductor.babylon-software.com/>
- Various Authors (Bing) (n.d.) Machine Translation engine Translate.eu. Available on: <http://translate.eu/>
- Various Authors (DeepL) (2019) Machine Translation engine DeepL. Available on: <https://www.deepl.com/translator>
- Various Authors (Google) (2019) Machine Translation engine Google Translate. Available on: <http://translate.google.com/>

Various Authors (Real Academia Española) (2019) *Diccionario de la lengua española*.
Madrid, España. Available on: <http://dle.rae.es/>

Various Authors (WordLingo) (2014) Machine Translation engine WordLingo.
Available on: <http://www.worldlingo.com/>