

THE RELIABILITY AND LIMITATIONS OF GOOGLE TRANSLATE: A BILINGUAL,
BIDIRECTIONAL AND GENRE-BASED EVALUATION

Blanca Roig Allué
Universidad de Zaragoza

ABSTRACT

The popularity of machine translation systems (or CAT, computer-assisted translation), which enable their users to obtain automatically generated translations of any text, has been increasing ever since they were created. One of the most widely used machine translation is Google Translate, a statistical system whose performance is the object of study of this paper. In order to evaluate its reliability, a small-scale study has been carried out in which translations of tourist texts and football match reports published online generated by the tool have been analysed, and the most representative mistakes found in terms of frequency have been classified at a lexicogrammatical, syntactic, pragmatic and punctuation level. Based on these findings, the main linguistic limitations of Google Translate have been established. It is worth highlighting the fact that three crucial variables have intervened in the corpus compilation, namely language (English/Spanish), direction of the translation (English into Spanish and vice versa) and genre of the texts (tourist and sports).

KEYWORDS: Machine translation, Google Translate, Reliability, Corpus, Linguistic limitations

RESUMEN

La popularidad de los sistemas de traducción automática, que permiten a los usuarios obtener al instante traducciones de cualquier texto, se ha visto aumentada desde su creación. Uno de los sistemas más utilizados es Google Translate, herramienta de traducción automática estadística en la que se centra este trabajo. Con el objetivo de evaluar su fiabilidad, se han analizado traducciones de textos turísticos y de resúmenes de partidos de fútbol en línea, proporcionadas por la herramienta, y los errores más relevantes de las mismas se han clasificado en cuatro categorías: léxico-gramaticales, sintácticos, pragmáticos y de puntuación. A través de esta clasificación, se han determinado las principales limitaciones lingüísticas de Google Translate. Es importante mencionar las tres variables fundamentales que se han combinado para la creación del corpus: idioma (inglés/español), dirección de la traducción (de inglés a español y viceversa) y género de los textos (turísticos y de prensa deportiva).

PALABRAS CLAVE: Traducción automática, Google Translate, Fiabilidad, Corpus, Limitaciones lingüísticas.

FECHA DE RECEPCIÓN: 13/09/2016

FECHA DE ACEPTACIÓN: 20/12/2016

PÁGINAS: 67-80

1. INTRODUCTION

1.1. NEW TECHNOLOGIES AND TRANSLATION

It is undeniable that the rise of new technological advances which began during the second half of the 20th century has contributed to creating an interconnected world which can be defined as a 'global village'. This term, coined by McLuhan (1962), perfectly illustrates the idea of a world in which the instantaneous communication and flow of information make physical barriers and geographical boundaries disappear. One of the many fields which are affected by these new technologies is linguistics, and more specifically, translation. It is worth noting that a crucial element in this new technological world is, undoubtedly, the search engine Google. In fact, Herrera (n.d.) defines it as an "irreplaceable tool" in the translator's documentation process and acknowledges the fact that Google is the preferred option when documentation is needed in order to produce a target text.

The impact of the application of new technologies to the field of linguistics gave birth to the so-called automatic or machine translation. As can be inferred by its name, this technology provides its users with automatically generated translations of a source text. In the entry "What Are the Main Types of Machine Translation?", retrieved from the webpage of the company Safaba Translation Solutions, four main types of automatic translation are identified, namely rule-based machine translation, statistical machine translation, hybrid machine translation and next generation approaches. The first one consists of systems which use rules developed by human experts in order to accurately translate structures from the source language into the target language. Statistical machine translation, by contrast, relies on computer algorithms of statistical models of words and phrases in order to create bilingual databases, whereas hybrids constitute a blend of the previous ones, thus representing an improvement of statistical systems but increasing their costs. Finally, the so-called new generation systems are statistical ones to which sophisticated advances in Language Transformation, Language Optimization Technologies and Terminology Management solutions have been added, therefore producing better translations.

One of the most popular systems of automatic translation is, indeed, Google Translate (hereafter GT). According to the classification which has been proposed above, the tool can be defined as a statistical system, although currently there are some hints which might indicate the fact that syntactic rules are being added, which would therefore categorise it as a hybrid tool. Nevertheless, this conclusion is entirely hypothetical, due to the secrecy which has always surrounded the way in which the tool works. In spite of its limitations, which are going to be explored throughout this paper, the usage of GT has been increasing dramatically since it was created in 2006. In fact, more than five hundred million people from different places around the world use it, according to Turovsky (2016). Turovsky argues that GT helps people to make connections, to reflect trends and events, and to enable its users to have a conversation no matter what language they speak. All these factors contribute to the rise of the tool's popularity, and therefore it is useful to establish its limitations in order for its users to know to what extent it is reliable. Nevertheless, it should be highlighted that if the hybrid mode is followed and GT becomes a hybrid tool, its limitations will presumably be reduced in a few months' time.

1.2. THE TRANSLATION FROM ENGLISH INTO SPANISH AND BACK: A FEW HINTS ON CONTRASTIVE LINGUISTICS

Out of the 103 languages with which Google Translate works, the reliability of the tool is going to be explored throughout this paper by analysing translations bidirectionally, in both Spanish into English and English into Spanish. Therefore, it is worth summarising the main differences between both languages from a contrastive point of view, considering their syntax, morphology and punctuation (López Guix & Wilkinson 1997:83-159) in order to better evaluate the target texts generated by GT.

As far as syntax is concerned, it is worth mentioning the fact that the English language tends to use short sentences separated by full stops, whereas Spanish prefers longer constructions composed by

subordinate clauses (López Guix & Wilkinson 1997:85). Likewise, juxtaposition and coordination are more frequent in English than in Spanish, which prefers the use of subordination (p. 89). As regards word order, both languages also differ to a great extent, especially in the position of the subject in the sentence and that of the adjective in the noun phrase. Thus, the Spanish language allows variation in the position of the previously mentioned elements, whereas in English the subject is almost invariably placed before the verb, and the adjective usually precedes the noun (p. 91).

Regarding morphological differences, the most relevant words which need to be taken into consideration are the article, the adjective, the adverb and the verb. The article, for instance, is often omitted in English in cases when Spanish uses the definite article (p. 99). Furthermore, the adjective in English is morphologically invariable and tends to precede the noun, whereas in Spanish it marks both gender and number and can be placed either before or after the noun, depending on its expressive value (p. 101).

With reference to the adverb, the main difference is the fact that the flexibility of the English language enhances the constant use of adverbs created by adding the derivational suffix “-ly” to a variety of words. Despite the fact that the equivalent suffix “-mente” also exists in Spanish, it is far less commonly used (p.119). Finally, there are also significant differences between the Spanish and the English verb. For instance, the former has preserved a different verbal desinence for each person, and therefore the subject can be omitted without causing misunderstandings. By contrast, the latter is characterised by its morphological simplicity, which makes the explicitation of the subject compulsory and contributes to the rigidity of the word order in the sentence (p. 123). Another interesting aspect is the verb tense, given the fact that the correspondence between both languages is not total, but partial (p. 125). According to Valero Garcés (2012), one of the most problematic differences is the extensive use of progressive forms which characterises the English language. Whereas in English they can be applied to every verb tense, in Spanish their use is far more restricted.

The last aspect worth commenting is punctuation. Nowadays, the English language prefers the use of the so-called ‘open punctuation’, as opposed to the more strictly grammatical ‘close punctuation’ used in Spanish (López Guix & Wilkinson 1997:145). Some key differences between both languages are the use of full stops, which are far more common in English, and commas, whose use greatly differs from one language to the other (pp. 146-152). Other punctuation marks which show divergences in their use are the dash, the hyphen and the inverted commas. The use of the semicolon, by contrast, is relatively similar in both languages (pp. 152-157). Apart from these, Valero Garcés (2012) mentions other interesting aspects in which both languages diverge, such as letter greetings and hour indications. Finally, as regards orthotypography, it is worth mentioning that several divergences can be found in the use of capital letters, due to the fact that Spanish tends to prefer small letters where English uses capitalisation (López Guix & Wilkinson 1997:157-159).

1.3. RESEARCH QUESTIONS

The purpose of this paper is to provide answers for several general and specific research questions. As regards general ones, the main goal is to determine the reliability of GT, by considering the following:

- i) to what extent GT meets its users’ needs and, therefore,
- ii) to what extent GT achieves its function as an automatic translation tool which facilitates and accelerates the transference of information.

To be able to provide answers for these questions it is worth bearing in mind the fact that the aim of the users of this tool is not to obtain a professional translation, but to have a general idea of what is said in a text written in a language which they cannot master or understand.

At a more specific level, this paper aims at identifying the mistakes committed by GT, in order to propose a classification of these translation mistakes. Likewise, this categorisation contributes to

establishing the limitations of the tool, that is, the aspects which could be improved in order to make it more reliable. Furthermore, it would also be interesting to analyse the aspects in which the tool succeeds in spite of being a system based on automatic translation.

2. METHODOLOGY

In order to answer the questions which have been previously posed, a pilot corpus of 3,826 words has been utilised. It is constituted by texts available online, which can be classified according to genre as tourist texts and and football match reports. As can be observed in Table 1 below, the number of words in the corpus is similar in terms of genres included and source text's language:

| | SL: English | SL: Spanish |
|------------------------|-------------|-------------|
| Tourist texts | 792 | 992 |
| Football match Reports | 992 | 1,050 |

Table 1. Number of words per genre and source language

Before focusing on the conventions of the chosen genres, it is worth clarifying the concept of 'genre' in order to better understand this classification. According to Swales (1990:42), the different approaches to the notion of genre in the field of linguistics depend on the emphasis given to their goal-directed nature, their having a schematic structure and their being "disassociated from registers or styles". Likewise, he defines genre as "a class of communicative event, the members of which share some set of communicative purposes" (Swales 1990:58). Apart from sharing a communicative purpose, Swales considers texts belonging to the same genre as having a similar structure, style, content and intended audience. If these conditions are fulfilled, the exemplar will be regarded as prototypical.

Together with Swales's considerations on genre, Bhatia (1993) adds to the former's linguistic and sociological notions the "psychological, particularly cognitive, level of genre construction" (Bhatia 1993:16), that is, the tactical aspects which characterise genre as a dynamic social process. These tactical choices or 'strategies' are "exploited by a particular writer" in order for the writing to be "more effective" (p. 20). Thus, he defines genre as "an instance of a successful achievement of a specific communicative purpose using conventionalised knowledge of linguistic and discoursal resources", and highlights the importance of both the conventional features which are common to the exemplars of a genre and the cognitive constraints attached to these features (p. 16).

In order to evaluate GT, texts belonging to two different genres have been chosen: tourist texts and football match reports. In both cases, texts available online have been analysed, in order to facilitate the process of evaluation of the target texts generated by the tool. In the case of the tourist texts, descriptions of exotic places have been selected, given the fact that finding available translations of these locations is difficult, which would justify the use of GT. As regards football match reports, they have been taken from several newspapers available online.

According to Skibitska (2013), online tourist texts are characterised by their emphatic and motivating components. Therefore, several stylistic devices are used in order to accentuate the "expressive, vivid, persuasive and laconic" nature of these texts. Some of the most frequent devices are epithets and direct questions. Furthermore, there are others which can be found on a less regular basis, such as metaphors, repetitions, inversions, hyperboles, ellipses, euphemisms, rhetoric questions, gradations, similes, periphrases, aposiopeses, antitheses, metonymies, parallel constructions and intertextuality.

Apart from the abundance of stylistic devices, it is worth taking into account other aspects which are key to define tourist texts. Sanning (2010) argues that the expressive, the informative and the vocative function can be found in these writings. The first responds to the writer's expression of his or her own feelings, whereas the second makes reference to "the facts of a topic, that is, reality outside

language, including reported ideas or theories” (p.125). Finally, the vocative function can be seen in the allusion to the addressee, that is, to the readership.

As far as football match reports are concerned, it is worth indicating that the conventions associated to the sports press in general can be applied to this particular genre. Thus, Martínez Moreno (2013) describes the style of the texts belonging to the sports press as jargon-like, with a semi-specialised vocabulary. In spite of this, however, they are easily understandable by the reader. Furthermore, he highlights the abundance of terms related to war and the epic genre, as well as of expressions characterised by their lyricism. He also defines the sports-press style as full of borrowings, new expressions and terms which often cannot be found in dictionaries. Finally, Martínez Moreno states that standardisation is needed in order to create linguistic norms which guarantee the correct and homogeneous use of language in the texts belonging to the sports press.

Despite the fact that there is no apparent relationship between tourist texts and football match reports, there are several reasons which justify their choice as the basis for the evaluation of GT. Firstly, both are informative texts which have an international scope. Secondly, they pose different problems at a linguistic level. For example, the cultural references related to the place which is being described represent a common source of problems in the translation of tourist texts. On the other hand, the metaphorical use of language which characterises football match reports also represents a translation problem.

Finally, and on a more general level, the use of texts which belong to different genres provides a third axis of comparison which needs to be taken into account, given the fact that each genre is located in a different context of situation, has a different function and is characterised by different conventions. Therefore, it is worth bearing in mind the three crucial factors which intervene in this evaluation of GT: two languages (English and Spanish), two directions of translation (English into Spanish and Spanish into English) and two genres (online tourist texts and football match reports). It can thus be said that this evaluation is bilingual, bidirectional and genre-based.

Moreover, in general terms, it may be assumed that there is a different GT user profile for each direction. Thus, those users who demand translations from Spanish into English are likely to do so in order to disseminate them, although the need for understanding may also cater for certain cases, as will be exemplified later. By contrast, the translation of English texts into Spanish is likely to derive exclusively from the need to understand texts originally written in English. Focusing on the two genres which have been selected for the present study, the use of GT is justified in both cases. In the case of tourist texts, people who are planning to travel and do not find information about a specific place (for instance an exotic place) in their mother tongue may rely on the tool in order to have a general idea of the characteristics of that location. As regards football match reports, the potential users would be people who want to know, in general terms, what happened during a game which they were not able to watch. If they find no report in their mother tongue, for instance in English, given the fact that football has a slightly higher cultural relevance in Spain than in English-speaking countries, they may choose GT in order to get to know the most important events of the match. This case is an instance of the previously mentioned Spanish-into-English translation which does not respond to diffusion purposes.

3. ANALYSIS OF FINDINGS

Throughout this section, the most representative translation mistakes which have been found in the target texts automatically generated by GT are going to be analysed. In order to facilitate this analysis, mistakes have been classified into four categories: lexicogrammar, syntax, pragmatic aspects and punctuation. It is worth highlighting the fact that the category of lexicogrammar encapsulates mistakes related to morphology, semantics (lexis) and phraseology (multi-word lexical units, such as idioms). The research findings will be presented separately, according to the genre of the texts analysed and the direction of the translation.

3.1. TOURIST TEXTS

As regards tourist texts, Table 2 provides quantitative data about the number and frequency of each of the four categories of mistakes which can be found in the translations under study, in order to illustrate the information given below. The percentage of each type of error has been calculated, and the results are the following:

| Tourist texts | LEXICOGRAMMAR | SYNTAX | PRAGMATIC ASPECTS | PUNCTUATION |
|----------------------|------------------|------------------|-------------------|------------------|
| English-Spanish | 2.8% (22) | 1.14% (9) | 0.25% (2) | 0.63% (5) |
| Spanish-English | 0.8% (8) | 1.6% (16) | 0.1% (1) | 0% (0) |

Table 2. Raw and normalised results (per hundred words) of mistakes according to typology and direction of translation in tourist texts.

3.1.1. English into Spanish

As can be observed in Table 2 above, in the English tourist texts translated into Spanish by GT a predominance of mistakes and inaccuracies related to lexicogrammar can be appreciated. Particularly, the most frequent ones are those concerning morphology and lexis. Some relevant instances at the level of morphology are those related to gender and number agreement, such as the translation of “world-famous” as “famosas” instead of “famosos” when it is used as a modifier of the noun “parks” (Text 1), and that of “is a richly rewarding place” as “es un lugar muy gratificantes” instead of “es un lugar muy gratificante” (Text 1). As adjectives are invariable in English, this mistake does not occur in the opposite direction. With regard to lexical mistakes, several instances can also be found, such as the translation of “which left millions dead and *flattened* almost the whole peninsula” as “lo que dejó millones de muertos y *aplanado* casi toda la península” instead of “ *arrasó*” (Text 2), and that of the phraseological unit “round-the-clock business” as “negocios alrededor del reloj” instead of “negocios abiertos las 24 horas” (Text 2). The last instance is a typical case of literal translation (Zaro & Truman 1999:26). These mistakes are not likely to cause, however, important misunderstandings to the reader, who will probably infer the meaning of the mistranslated terms through the co-text.

Still on the level of lexicogrammatical mistakes, it is worth commenting on the translation of some idiomatic expressions. “If you get off the beaten track”, for instance, is translated as “Si obtiene fuera de lo común” (Text 1). A more acceptable translation would be, for instance, “si usted huye de lo convencional”. The idiom “has largely gone the way of the dodo” (Text 2) is also mistranslated as “ha ido en gran parte del camino del dodo”, which constitutes another case of literal translation (Zaro & Truman 1999:26). A more accurate option would be “en gran parte ha desaparecido”. These mistakes are likely to confuse the reader and cause a misunderstanding of the intended meaning of the source text. Other idioms, by contrast, are translated correctly, such as “nothing short of miraculous” (Text 2), which is translated as “poco menos que milagrosa”.

With regard to syntax, there are several mistakes of this category which are worth commenting on. One of them is the frequent mistranslation of noun phrases. For instance, “Kenya’s world-famous national parks” (Text 1) is translated as “parques de Kenia famosas nacionales”, instead of “los mundialmente famosos parques nacionales de Kenia”; “Korea’s very existence” (Text 2) is translated as “Corea existencia”, instead of “la propia existencia de Corea”; and “Confucian-style formal ceremonies” (Text 2) is translated as “al estilo confuciano ceremonias formales”, instead of “ceremonias formales al estilo confuciano”. We may attribute these mistakes to GT’s inability to recognise the head of noun

phrases with more than one modifier (such as nouns and adjectives), which would explain the fact that the tool tends to disorganise their constituents. Nevertheless, this mistake is not likely to cause major misunderstandings of the message of the source text.

Another case of syntactic mistake which does not alter the meaning of the source text is the literal translation (Zaro & Truman 1999:26) of syntactic structures which do not exist in Spanish. For instance, “an absolute joy to dive into” (Text 2) is translated as “una alegría absoluta para sumergirse en”, which in spite of constituting a grammatical mistake, maintains the original meaning. Other syntactic mistakes, by contrast, are likely to confuse readers and prevent them from rightly interpreting the text. An obvious instance of this is the translation of “start a conversation with any local and you’ll soon find out about the country’s deep economic and social tensions” (Text 1) as “iniciar una conversación con cualquier ley local y pronto se enteró de profundas tensiones económicas y sociales del país”. A more accurate option would be, for instance, “Inicie una conversación con cualquier lugareño y pronto descubrirá las profundas tensiones sociales y económicas del país”.

As far as pragmatic mistakes are concerned, there are some cases of lack of coherence in the target texts generated by GT. For instance, “North Korea has armed itself to the teeth since 1953, stagnated in its pursuit of a local brand of Communism and become one of the least accessible countries in the world” (Text 2) is translated as “Corea del Norte *se ha armado* hasta los dientes desde 1953, *se estancó* en su búsqueda de una marca local del comunismo y *convertirse* en uno de los países menos accesibles del mundo”. As can be appreciated in the italicised words, there is an alternation of verb tenses within the same sentence, due to the fact that the tool has not identified that “stagnated” and “become” are past participles which follow the auxiliary verb “has”. The combination of present perfect, past simple and infinitive creates ambiguity in the target text and may prevent the reader from correctly interpreting it.

Finally, it is worth mentioning the fact that there is a mistake related to punctuation which frequently appears in the generated translations. In the English language, the dash is frequently used in order to introduce explanations and recapitulations, and also in order to give emphasis, whereas Spanish uses commas, colons, semicolons and suspension points for the same purposes (López Guix & Wilkinson 1997:153). Nevertheless, GT maintains in the target texts the dashes of the source texts, in spite of the fact that it would be better to use colons, commas or brackets. This inaccuracy, however, does not produce meaning shifts.

3.1.2. Spanish into English

The target texts provided by the tool when translating from Spanish into English have a slightly higher quality than those in the opposite one, the genre being the same. This could be due to the fact that the bank of data from which GT draws choices is bigger in English than in Spanish. As regards lexicogrammar, for instance, there are cases in which the choice of vocabulary is accurate, as can be seen in the translation of “la ley jordana *establecida*” as “Jordanian law *stipulated*” (Text 3) and “Los recuerdos más *solicitados*” as “The most *popular* souvenirs” (Text 5). The tool also correctly translates “las gambas” as “shrimp” and “pescados” as “fish” (Text 5), these invariable plural forms being exceptions. Some lexical inaccuracies, however, can be found, such as the translation of “estallar” as “explode” instead of “break out” when referring to a “war” (Text 3), “albóndigas” as “dumplings” instead of “meatballs” (Text 3), and “todo español” as “all Spanish” instead of “every Spaniard” (Text 4). Nevertheless, these inaccuracies do not alter the message conveyed by the source text.

Still on the lexicogrammatical level, it should be noted the fact that some phraseological units are successfully translated. For instance, “se mantuvo al margen” (Text 3) is translated as “stood on the sidelines”, and thus not only the message is transmitted, but the metaphorical way of expressing it is also maintained. Other expressions, such as “al alcance de todos los bolsillos” (Text 4), are also translated in a clear and unambiguous way (“within reach of every budget”). Although “affordable for everybody”

would sound more natural, the choice of GT perfectly encapsulates the idea conveyed by the source text. Furthermore, in some cases the tool omits the definite articles of the source text, due to the fact that their use is more frequent in Spanish than in English (López Guix & Wilkinson 1997:99-100). An example of this is the translation of “la carne de foca, la ballena, las gambas, el reno (caribú), el buey almizclero y pescados” (Text 5) as “seal meat, whale, shrimp, reindeer (caribou), musk ox and fish”.

In comparison to the translations from English into Spanish analysed above, in which errors related to lexicogrammar abound, in the Spanish-into-English translations there is a predominance of syntactic mistakes in the texts generated by GT (See Table 2), the genre being the same. These mistakes may lead the reader to misunderstandings, as can be appreciated in the translation of “solo podrían ser herederos al trono los nacidos de padres árabes y musulmanes por nacimiento” (Text 3) as “could only be born heirs to the throne of Arab and Muslim by birth parents”, instead of “only those born to parents who were Arab and Muslim by birth could be heirs to the throne”, and that of “unas figuritas talladas en colmillos de morsa o en cuernos de reno a los que llaman "tulipaks"” (Text 5) as “a carved walrus tusk or reindeer antlers callers "tulipaks" figurines”, instead of “figurines, carved in walrus tusk or reindeer antlers, called “tulipaks”.

Another frequent source of syntactic mistakes is the translation of noun phrases. Sometimes, noun phrases composed by a noun as head and an adjective as modifier are mistranslated. For instance, “la corona noruega” (Text 5) is translated as “the crown Norway”, instead of “the Norwegian crown”, and “la alimentación inuit” (Text 5) is translated as “food Inuit”, instead of “Inuit food”. Other instances of mistranslation of noun phrases are “albóndigas de garbanzos fritas” (Text 3) translated as “dumplings fried chickpeas” instead of “fried chickpea dumplings”, and “especialidad de los beduinos” (Text 3) translated as “specialty Bedouins” instead of “specialty of the Bedouins”. This problem is common to both directions and genres, and it is often caused by the lack of a preposition which would clarify the meaning of the target text. However, these mistakes are not likely to mislead the reader.

Furthermore, it is worth mentioning the problems derived from the translation of some discourse markers, which relate to both semantics and syntax. For instance, “*y es que* dada su cercanía” (Text 4) is translated as “*and that* given its proximity”, which is likewise a case of literal translation (Zaro & Truman 1999:26), instead of “*In fact*, given its proximity”. Another example is the translation of “*por lo que* se considera” (Text 5) as “*which* is considered” instead of “*that is why* it is considered”. Other discourse markers, by contrast, are correctly translated, such as “*En cuanto a* su gastronomía” (Text 5), which is translated as “*As for* its gastronomy”. Finally, it is worth noting the fact that no relevant mistakes related to pragmatics and punctuation have been found. In fact, no punctuation mistakes are identified in the Spanish-into-English direction, whichever the genre is. Thus, it can be concluded that it does not depend on the genre, but on the direction.

3.2. FOOTBALL MATCH REPORTS

Table 3 below provides quantitative data about the frequency of each of the four categories of mistakes which can be found in the translations of football match reports generated by GT, in order to complement the information given in 3.2.1. and 3.2.2. The percentage of each type of error has been calculated, and the results are the following:

| Football match reports | LEXICOGRAMMAR | SYNTAX | PRAGMATIC ASPECTS | PUNCTUATION |
|-------------------------------|-------------------|-------------------|-------------------|-----------------|
| English-Spanish | 2.21% (22) | 1.41% (14) | 0.3% (3) | 0.1% (1) |
| Spanish-English | 1.14% (12) | 1.8% (19) | 0.28% (3) | 0% (0) |

Table 3. Raw and normalised results (per hundred words) of mistakes according to typology and direction of translation in football match reports.

3.2.1. English into Spanish

As can be observed in the table above, in English-into-Spanish translations both lexicogrammatical and syntactic mistakes abound. At a pragmatic level, there are also some cases of lack of coherence, which is typical in automatic translation. For instance, “aggregate” (Text 6) is translated as both “global” and “agregada” within the same text, and “effort” (Text 6) is translated as both “remate” and “esfuerzo”. In these cases, the pragmatic and the semantic level overlap. Nevertheless, translation mistakes related to lexicogrammar and syntax are far more frequent than pragmatic ones, and in several cases they are likely to mislead the reader and prevent him from understanding the text.

The first lexicogrammatical aspect to highlight is morphology, given the fact that mistakes in gender agreement are frequently found in the target texts provided by GT. These mistakes also appear in the translation of tourist texts from English into Spanish, as has been explained in section 3.1.1. However, since adjectives are invariable in English, those mistakes do not occur in the opposite direction. Some instances are the translation of “the most aesthetic of wins” (Text 7) as “el más estético de victorias”, that of “With Roma forced to push” (Text 6) as “Con Roma forzado a empujar”, and that of “a very tough Real Sociedad” (Text 7) as “una muy duro equipo de la Real Sociedad”.

As far as lexis is concerned, it is worth noting that in some cases the choice of vocabulary is accurate. Thus, “on aggregate” (Text 6) is translated as “en el global”, “injury time” (Text 8) as “tiempo de descuento”, “forward” (Text 8) as “delantero”, and “free kick” (Text 8) as “tiro libre”. A hypothetical explanation for the fact that these terms are in the database of GT might be the high impact that football has on the Spanish culture. By contrast, other terms related to the field of sports are mistranslated. For instance, in Text 6 “manager” is translated as “gerente” (instead of “director técnico” or “entrenador”), “fielding” is translated as “fildeo” (instead of “alineando”), and “stinging” is translated as “picadura” (instead of “directo”). Furthermore, some terms are not translated at all, such as “shanked” (Text 6), “ruing” (Text 6), “ratcheted” (Text 7) and “cut-back” (Text 8). If the users of GT have no knowledge of the English language, this apparently unimportant mistake may prevent them from understanding a crucial part of the source text. This lack of translation of some words can also be found in the Spanish-into-English direction, as will be seen in 3.2.2.

Another lexicogrammatical aspect which deserves attention is the translation of metaphorical expressions which are typical in this genre. These expressions are literally translated (Zaro & Truman 1999:26), and despite the fact that in the particular case of the translation of “gilt-edged chance” (Text 6) as “oportunidad con bordes dorados” the meaning is likely to be inferred by the user of GT, although “oportunidad de oro” would be more accurate, in the majority of cases the translation is not successful and the reader will probably be misled. Some instances of these mistranslations are that of “provided a lift to Los Blancos” (Text 6) as “instala un elevador a Los Blancos” (instead of “animó a los blancos”), that of “Real’s attack had a new look to it” (Text 7) as “el ataque de Madrid tuvo una nueva mirada a ella” (instead of “el planteamiento del ataque del Madrid fue diferente”), and that of “their second ‘big two’ scalp of the season” (Text 7) as “su segundo “grande dos “del cuero cabelludo de la temporada”

(instead of “su segunda gran victoria de la temporada”). These literally translated expressions are likely to cause misunderstandings to the reader, because their meaning is impossible to grasp.

With regard to syntactic mistakes, several cases can be found in the generated texts. The mistranslation of noun phrases, which is common to both translation directions and genres, is a frequent source of semantic changes, and therefore represents a major drawback for the reliability of GT. For instance, “in the tournament's final eight” (Text 6) is translated as “en la final del torneo de ocho” (instead of “entre los últimos ocho del torneo”), “manager Zinedine Zidane” (Text 6) is translated as “gerente *de* Zinedine Zidane” (instead of “el director técnico Zinedine Zidane”), “Substitute Lucas Vázquez” (Text 6) as “Sustituto *de* Lucas Vázquez” (instead of “El suplente Lucas Vázquez”), and “a scissor-kicked effort right at Szczesny” (Text 6) as “un esfuerzo correcto tijera patadas a Szczesny” (instead of “un remate de tijera directo a Szczesny”). This last case is an example of the tendency of GT to mix the constituents of phrases, thus creating chaotic syntactic structures which often make comprehension impossible.

Furthermore, it is worth noting the fact that “as” is invariably translated as “como”, even if it has a temporal or a causal value in the source text. Some examples are the translation of “And then as Barça struggled to clear a corner” (Text 8) as “Y entonces, como el Barça tuvo problemas para despejar una esquina” (instead of “Y entonces, mientras el Barça tenía problemas para despejar un córner”), that of “as Atlético tired” (Text 8) as “como Atlético cansado” (instead of “mientras el Atlético se cansaba”), and that of “as Gabi handled” (Text 8) as “como se maneja Gabi” (instead of “por una mano de Gabi”). This mistake, related to both semantics and syntax, also prevents the reader from rightly understanding the text.

Finally, it is worth commenting on some syntactic inaccuracies which are not likely to cause misunderstandings to the reader. Mistakes in subject-verb agreement, for instance, can often be found in the target texts generated by GT, as can be appreciated in the translation of “Atlético were worthy winners” (Text 8) as “Atlético eran dignos ganadores”, and that of “the Basques showed” (Text 7) as “los vascos mostró”. Besides, in the translation of “a crucial 1-0 lead which they would not surrender” (Text 7) as “una importante ventaja de 1-0 que no se rendirían”, the transitive verb “surrender” is translated intransitively, thus creating a syntactic structure which has no coherence. A better option would be, for instance, “una importante ventaja de 1-0 a la que no renunciarían”. Nevertheless, the reader will presumably infer the intended meaning through the co-text.

3.2.2. Spanish into English

In the football match reports translated from Spanish into English, lexicogrammatical and syntactic mistakes predominate, as happens in the opposite direction. Likewise, a lack of coherence in the target texts generated by Google Translate can be appreciated in the examples which are going to be given throughout this section. Furthermore, no punctuation mistakes can be found in this direction, as has been said in 3.1.2.

With reference to lexicogrammatical aspects, it is worth saying that the translations show some lexical inaccuracies, especially regarding technical vocabulary related to the field of sports, and of football in particular. In spite of the fact that some terms are correctly translated, such as “azulgranas” (Text 10) as “Catalans” and “la ida de las semifinales” (Texts 11 and 12) as “the first leg of the semifinals”, the majority of football-related lexis is mistranslated. For instance, “parada” (Texts 10 and 11) is translated as “stop” instead of “save”, and “el balón no había empezado a *rodar*” (Text 11) is translated as “the ball had not started *filming*” instead of “the ball had not started *rolling*”, therefore using a term related to the field of cinema. Furthermore, “falta” (Text 11) is sometimes accurately translated as “free kick”, whereas it is wrongly translated as “lack” within the same text. This also happens with the term “partido” (Text 11), which is alternatively translated as both “party” and “match”. However, this lack of coherence, which can be classified as an inaccuracy related to pragmatics, is not likely to cause major

misunderstandings to the potential user of GT, whose aim is to have a general idea of what happened during the match.

Apart from lexical inaccuracies, there are also some instances of morphological mistakes, such as the incorrect number of the possessive adjective in the translation of “al resentirse de *sus* problemas musculares” (Text 11) as “to resent *their* muscle problems” (instead of “*his* muscle problems”), the preceding noun being “Cristiano Ronaldo”. Still on the lexicogrammatical aspects, it can be observed that certain structures have been literally translated, word by word (Zaro & Truman 1999:26), and therefore the target text does not maintain the meaning conveyed by the source text. For instance, “*Y eso que el arranque fulgurante*” (Text 10) is translated as “*And that the meteoric start*”, “*que sí que quiso hacerse con el mando del partido*” (Text 11) is translated as “*yes that would take command of the game*”, and “*será el no va más para Madrid y City*” (Text 11) is translated as “*it will be the does not go to Madrid and City*”.

Furthermore, it is worth highlighting the fact that, as opposed to the English-into-Spanish direction, the translation of the metaphorical expressions which characterise football match reports does not cause major comprehension problems in this direction. Thus, “*blindó la moral de los suyos*” (Text 10) is translated as “*shielded the moral of their own*”, “*los grandes caen con estrépito*” (Text 10) is translated as “*the big fall noisily*”, and “*uno de esos castigos que alargan mucho la convalecencia*” (Text 13) is translated as “*one of those punishments that greatly extend convalescence*”. Despite the fact that human translators are able to provide far better options by using, for instance, the strategy of adaptation (Zaro & Truman 1999:24), GT’s translation of these metaphorical expressions succeeds in conveying the meaning intended in the source text, and therefore its user is likely to interpret the message correctly.

Regarding syntactic mistakes, there are several cases which deserve attention. There are instances of mistranslation of noun phrases, which, as has been said in the previous sections, is common to both directions. Thus, in Text 11 “*el equipo de Pellegrini*” is translated as “*team Pellegrini*” instead of “*Pellegrini’s team*”, and “*peticiones de penaltis*” is translated as “*requests penalty*” instead of “*requests for penalty*”. Although these inaccuracies are not likely to cause misunderstandings, there are other mistakes in syntactic structures which either make comprehension impossible or convey a misleading meaning. For instance, “*lesión muscular que sufrió el canario en los isquiotibiales*” (Text 12) is translated as “*muscle injury he suffered canary hamstring*” instead of “*muscle injury that the Canary suffered in the hamstring*”, “*al que esta vez no le sirvió ni jugar bien ni poner actitud*” (Text 10) is translated as “*which this time did not help him or play well and put attitude*” instead of “*to whom playing well and having attitude did not help this time*”, and “*Benzema, como se preveía, se marchó al banquillo*” (Text 11) is translated as “*Benzema, as previous, left the bench*” instead of “*Benzema, as was expected, left the pitch and sat on the bench*”. Particularly, this last example of mistranslation of syntactic structures is likely to originate misunderstandings which prevent the reader from knowing a basic event in the match.

Still on the level of syntactic mistakes, GT seems to have a tendency to mistranslate negative forms of verbs. For instance, “*no estará recuperado en una semana*” (Text 12) is translated as “*there will be recovered in a week*” (instead of “*he will not have recovered in a week’s time*”), and “*pero tampoco encontró una respuesta contundente de su adversario*” (Text 13) is translated as “*but it also found a strong response from his adversary*” (instead of “*but it did not find a strong response from his adversary*”). It also has a tendency to insert personal pronouns which do not exist in the source text, as can be seen in the translation of “*como decía Zidane*” (Text 11) as “*as I said Zidane*” (instead of “*as Zidane said*”), and that of “*pero su falta la paró Keylor Navas*” (Text 11) as “*but his lack Keylor Navas stopped her*” (instead of “*but his free kick was saved by Keylor Navas*”).

Furthermore, relative pronouns introducing relative subordinate clauses are occasionally mistranslated, which does not alter the meaning of the source text. For instance, “*La estrella fue Alves, Diego, que desquició a los azulgranas*” (Text 10) is translated as “*The star was Alves, Diego, which unhinged the Catalans*”, instead of “*who unhinged the Catalans*”. At other times, by contrast, they are properly translated, as can be seen in the translation of “*Silva que se marchó lesionado*” (Text 11) as “*Silva who left injured*”, and that of “*Casemiro, que de mayor quiere ser Mauro Silva*” (Text 13) as

“Casemiro, *who* wants to be a Mauro Silva”. Finally, it is worth noting the fact that some terms, such as “pelucón” (Text 10), “cuitas” (Text 11), “artificiero” (Text 13) and “racheadamente” (Text 13), are not translated in the target text. This mistake, which as has been indicated in 3.2.1. also appears in football match reports translated in the opposite direction, may cause comprehension problems if the user has no knowledge of the Spanish language.

4. CONCLUSIONS

The aim of this paper was to assess the reliability of GT by means of a pilot study which consisted in the analysis of target texts generated by the tool, categorising the most relevant mistakes which can be found in them. In order to draw reliable conclusions, the potential function of the translated texts has been taken into account throughout this process. In the case of both tourist texts and football match reports, the main goal of the reader is not likely to be to obtain a professional translation, but to have a general idea of the characteristics of a place and of the events which happened during a football match, respectively.

Bearing in mind what the potential users of GT expect from the tool, and therefore leaving stylistic aspects aside, it can be said that lexicogrammatical and syntactic mistakes can frequently be found in the generated translations in both genres and in both directions of translation. In some cases, the meaning intended in the source text is still inferable, but other times the reader is likely to be led to misunderstandings, and thus the tool cannot be considered to be reliable, in the sense that it is not complying with its users’ needs and is not fulfilling its aim. According to the classification which has been proposed in Section 3, the most frequent translation mistakes in the target texts generated by the tool can be divided into four categories, namely lexicogrammar (encapsulating morphology, semantics and phraseology), syntax, pragmatic aspects and punctuation.

As can be appreciated in Tables 2 and 3, the percentage of lexicogrammatical and syntactic mistakes is higher than that of those related to pragmatics and punctuation, in both genres and directions of translation. It is worth noting the fact that within the genre of the tourist text, mistakes related to lexicogrammar predominate in the English into Spanish direction. By contrast, the mistakes which prevail in the opposite direction of translation are syntactic ones, the genre being the same. The same conclusion can be reached regarding football match reports, as can be observed in Table 3.

Furthermore, it is worth mentioning the fact that one of the most frequent syntactic mistakes in the target texts generated by GT is the mistranslation of noun phrases, whichever the genre of the text and the direction of the translation are. The tool tends to disorganise the constituents of those phrases, and even though in some cases the meaning intended in the source text is still inferable by the reader, other times comprehension is made almost impossible. Moreover, it is interesting to note the fact that in the translations of football match reports generated by GT, whichever the direction of translation is, some terms remain untranslated, therefore hindering understanding if the reader has no knowledge of the source language.

It is also worth highlighting the fact that the success of the literal translations provided by GT of the metaphorical expressions which characterise the genre of the football match report depends on the direction of the translation. Thus, these expressions are mistranslated and the reader is likely to be misled in the English-into-Spanish translation, whereas in the opposite direction the tool maintains the meaning intended by the expression used in the source text, therefore fulfilling its users’ needs and achieving its aim.

In the target texts generated by GT, mistakes related to pragmatics and punctuation are far less frequent than lexicogrammatical and syntactic ones, as can be appreciated in Tables 2 and 3. Furthermore, these inaccuracies are not bound to cause semantic variations in the target texts, and therefore do not represent a major drawback for the reliability of the tool.

Further differences can be identified if texts from the same genre are looked into. Thus, as regards intra-genre differences, it can be concluded that within the genre of the tourist text, GT provides

better translations from Spanish into English than in the opposite direction. This might be due to the fact that the database of the tool contains more tourist texts originally written in English. Therefore, one way of ensuring reliability and improving the quality of English-into-Spanish translations would be the increase in the number of tourist texts in Spanish. Within the genre of the football match report, by contrast, the translations provided by GT tend to be equally inaccurate and difficult to understand whichever the direction is. This is probably due to the fact that in the bank of data from which GT draws choices, the amount of texts belonging to this genre in both English and Spanish is lower than that of tourist texts. Again, it seems to be the case that the higher number of texts contained in the repository from which GT retrieves translation choices, the better quality these translations will have and the more reliable they will be.

Apart from these intra-genre differences, it is worth commenting on the inter-genre contrasts which can be appreciated after the analysis. Thus, it can be said that the translations of tourist texts provided by GT are slightly better-quality than those of football match reports. As has been said above, this might be due to the fact that the database of the tool includes a considerable amount of tourist texts, especially of those written in English. This hypothesis would explain the difference across genres in the quality of the translations provided by GT.

Thus, it can be concluded that the reliability of the tool depends on both the direction of the translation and the genre of the source text. The tool is more reliable if its user aims at understanding the content of a tourist text, especially if the translation is from Spanish into English. By contrast, it is less reliable for those who aim at having a general idea of the events which happened during a football match, whichever the direction of the translation is. Likewise, it could also be hypothesised that the more frequent a genre is online, the better the quality of its translations will be. Nevertheless, it is worth highlighting the fact that, if syntactic rules continue to be added and GT is transformed into a hybrid system, the target texts will presumably improve to a great extent, and therefore the tool will be far more reliable than it is nowadays. In spite of these advances, however, it does not seem plausible that the day will come when human translators will be replaced by machine translation systems.

5. BIBLIOGRAPHY

- Bhatia, Vijay Kumar (1993). *Analysing genre: Language use in professional settings*. London, Longman.
- Herrera, J. [online] (n.d.). “Documentación aplicada a la traducción: Hay vida más allá de Google”. *La linterna del traductor*, 9. <http://www.lalinternadeltraductor.org/n9/documentacion-traducccion.html> [retrieved April 20, 2016].
- López Guix, J. G. y J. M. Wilkinson (1997). *Manual de traducción inglés-castellano: Teoría y práctica*. Barcelona, Gedisa.
- Martínez Moreno, R. [online] (2013). “Breve análisis del lenguaje periodístico deportivo en España”, *Papel de periódico*. <http://papeldeperiodico.com/2013/09/breve-analisis-del-lenguaje-periodistico-deportivo-en-espana/> [retrieved May 7, 2016].
- McLuhan, Marshall (1962). *The Gutenberg galaxy: The making of typographic man*. Toronto, University of Toronto Press.
- Safaba Translation Solutions [online] (n.d.). “What are the main types of machine translation?”, *A quick guide to machine translation*. <http://www.machinetranslation.net/quick-guide-to-machine-translation/machine-translation-technologies> [retrieved May 5, 2016].
- Sanning, H. (2010). “Lost and found in translating tourist texts: Domesticating, foreignising or neutralising approach”, *JoSTrans: The Journal of Specialised Translation*, 13, 124-137.
- Skibitska, O. [online] (2013). “The challenges of translation of tourist e-text”, *Translation Journal*, 17 (4). <http://translationjournal.net/journal/66tourism.htm> [retrieved March 5, 2016].
- Swales, John Malcolm (1990): *Genre analysis: English in academic and research settings*. Cambridge, Cambridge University Press.

Turovsky, B. [online] (2016). "Ten years of Google Translate", *Google Translate blog*.

<http://googletranslate.blogspot.com.es/> [retrieved April 20, 2016].

Valero Garcés, C. [online] (2012). "Inglés y español mano a mano: Dos lenguas y dos formas de ver el mundo", *Cuadernos Cervantes*. http://www.cuadernos cervantes.com/lc_ingles.html [retrieved

May 20, 2016].

Zaro, J. J. y M. Truman (1999). *Manual de traducción: Textos españoles e ingleses traducidos y comentados*. Madrid, Sgel.

6. APPENDIX: SOURCE TEXTS SELECTED

6.1. TOURIST TEXTS

6.1.1. English into Spanish

Text 1- Kenya [<http://www.roughguides.com/destinations/africa/kenya/>]

Text 2- South Korea [<http://www.roughguides.com/destinations/asia/south-korea/>]

6.1.2. Spanish into English

Text 3- Jordania [<http://www.guias-viajes.com/content/view/1325/294/>]

Text 4- Lisboa [<https://www.disfrutalisboa.com/>]

Text 5- Groenlandia [<http://www.guias-viajes.com/content/view/1308/278/>]

6.2. FOOTBALL MATCH REPORTS

6.2.1. English into Spanish

Text 6- Real Sweat To Reach Quarters (extracts) [<http://www.marca.com/en/football/real-madrid/2016/03/08/56df4744ca4741cc198b462f.html>]

Text 7- Bale Header Keeps Real In Title Race (extracts) [<http://www.marca.com/en/football/real-madrid/2016/04/30/5724d5c2468aeb77048b4611.html>]

Text 8- Atletico's Griezmann At The Double To Knock European Champions Barça Out (extracts) [<http://www.telegraph.co.uk/football/2016/04/13/atletico-madrid-vs-barcelona-live---what-time-is-tonights-champi/>]

Text 9- Atlético Madrid 4-Real Madrid 0 (extracts)

[<https://www.thesun.co.uk/archives/football/23593/atletico-madrid-4-real-madrid-0/>]

6.2.2. Spanish into English

Text 10- El Barça se desploma (extract)

[http://www.marca.com/eventos/marcador/futbol/2015_16/la-liga/jornada_33/bar_val/]

Text 11- Madrid y Manchester City lo dejan todo para el Bernabéu (extracts)

[<http://www.mundodeportivo.com/futbol/champions-league/20160426/401380478595/manchester-city-real-madrid-resultado-resumen-cronica-champions-league.html>]

Text 12- David Silva, baja para la vuelta

[http://deportes.elpais.com/deportes/2016/04/27/actualidad/1461752332_537837.html]

Text 13- El Madrid siempre vuelve (extracts)

[http://futbol.as.com/futbol/2016/04/02/primera/1459621992_594312.html]