

Isabel DE LA CRUZ-CABANILLAS

Cristina TEJEDOR-MARTÍNEZ

University of Alcalá, Spain

The Error Analysis Approach for the Assessment of Automatic Translation¹

Abstract:

The Error Analysis Approach has been widely used to improve the learning of foreign languages and to evaluate human translation, but Error Analysis can also be applied to automatic translation, as shown by different scholars (see M. Koponen 2010). This article reports on a European project, called Organic.Lingua, which tries to demonstrate the potential of a multilingual web portal for Sustainable Agricultural & Environmental Education by using a machine translator in order to make available all materials in different languages. The selected texts are taken from the corpus compiled for this project. The source texts are translated into various target languages using computer translation tools. We have chosen some of these texts in English and their translations into Spanish, and have applied a classification of human errors to check whether the errors generated by machines could be avoided if we can ‘teach’ the machines or to what extent they are specific to computer tools or human beings. Some metric systems, such as BLEU (Bilingual Evaluation Understudy), are being employed to evaluate the quality of the target translation. However, our role as philologists is to concentrate on more specific linguistic features. By defining the different types of errors and by trying to establish a scale of gravity, we intend to determine the quality of the automatic translations. After analysing the data, we will propose linguistic measures for improvement that would be implemented by computer experts working on the project.

Introduction

There are some issues that must be addressed concerning the process of translation. An important aspect is quality, which is often difficult to define in translation terms. Different scholars point out what should be understood by quality. Koponen claims it covers accuracy, fluency and fitness for the purpose (see M. Koponen 2010: 1). Other authors (see R. Rabadán/ B. Labrador/ N. Ramón 2009: 303) warn about the elusive nature of the concept because of the lack of conceptual clarity and the inadequacy of the tools used to measure it. In addition, Colina (see S. Colina 2009: 237) remarks that the absence of theoretical foundations on the part of the experiential approaches contributes to the fact that the existing tools do not allow the results to be transferred to other

¹ This article includes research results which were enabled by the EC-funded project CIP-ICT-PSP.2010.6.2, reference number: 270999 under the ICT Policy Support *Programme*.

environments. Finally, Popovic & Ney (see M. Popovic/ H. Ney 2011: 658) mention some of the constraints on the assessment of machine translation: “A limitation of most measures of translation quality, including human judgments, is that they are not diagnostic, and there have been few attempts to analyze MT errors so that researchers can identify the types of errors that are made. Analyses of the errors produced by machine translation (MT) systems have the potential to focus research aimed at improving translation performance, but translation error annotation is problematic because there are many ways to translate a single expression from one language to another”.

However elusive the concept of quality may be, our study focuses on the comparison of parallel corpora to check the equivalence of the original and translated texts by means of error analysis. Parallel or translation corpora are understood as “original texts in one language and their translations into one or several other languages” (see B. Altenberg/ S. Granger 2002: 8). In order to assess the quality of the target corpora, English original texts on agricultural topics have been selected from the Organic Edunet web portal. The Organic.Edunet portal aims to facilitate access, usage and exploitation of digital educational content related to Organic Agriculture and Agroecology. This is achieved by deploying a multilingual online federation of learning repositories, populated with quality content from various content producers (see J.C. Luengo/ M.A. Sicilia/ S. Sánchez, forthcoming). The total amount of words of the English corpus is 13,506. The original corpus has been translated into Spanish using two different translators: Google Translator (henceforth, GT) and Yahoo Bablefish (henceforth, YB). In order to carry out this study, eight texts from the corpus were selected, a total amount of 1,088 words, and the translations in the target language, Spanish, have been analysed.

Our research focuses on human translation assessment of machine translation texts. Automated metrics, such as BLEU metric, have not been applied in this pilot study as automated quality metrics have been criticised when dealing with accuracy of content. In fact, studies have demonstrated that higher score by the metric does not guarantee better translation quality (see C. Callison-Burch/ M. Osborne/ P. Koehn 2006). Nevertheless, the application of this evaluation method should be the next step in order to obtain more detailed information about the types of errors shown by automated metrics and compare the results with human assessment.

1. Error Analysis Approach Applied to the Target Corpora

The taxonomy of errors is wide and varied when applied to the learning of foreign languages and also to evaluate human translation. It can also be easily adapted to machine translation, as Popovic and Ney (see M. Popovic/ H. Ney 2011: 686) remark that “the results obtained by the proposed framework [automatic error analysis] correlate very well with the results of human error analysis”.

After reviewing several error classifications designed to assess translations and language learning (I. Borrego Ledesma 2001, M. Koponen 2010, M. Popovic/ H. Ney 2011, I. Santos Gargallo 1993, G. Vázquez 1999), a taxonomy of error categories has been established in order to analyse the different translations outputs using both machine

translators. These error categories deal mainly with the use of words, but also with syntactical structures.

Broadly speaking, the categorization that has been set is that which involves errors of:

- 1) omission, a concept in the source text that is not conveyed in the target text or a word that should appear in the target text because it is not redundant in this language;
- 2) addition, a concept that is not present in the source text but appears in the target text or a word that appears in the target text, but is redundant;
- 3) mistranslation, the incorrect selection of terms in a specific context, the wrong formation of terms or the literal translation of a term in the target text;
- 4) untranslated concepts, a source language word that appears in the target text or the use of recent loan words; and
- 5) syntactical errors, lack of gender or number agreement, wrong or not totally correct order of elements in a sentence, wrong use of verbal forms and literal translation of syntactical structures. In some cases, the mismatches will be further classified into smaller groups within its category, depending on the type of error detected.

(1) Omission

A very common mistake has to do with the omission of the article. When generalizing, the English language can make use of different devices. One of them is the use of nouns in their plural form without the article. However, the article is compulsory in Spanish. Thus, multiple instances of this kind of omission are observed, as illustrated by:

- (1) Topographic and climatic conditions, neighbouring vegetation, soils and erosion risks were evaluated.
Condiciones topográficas y climáticas, la vegetación vecina, los suelos y los riesgos de erosión fueron evaluados. (gt)

In this case the Yahoo Bablefish translator includes the compulsory article in Spanish.

- Las condiciones topográficas y climáticas, la vegetación vecina, los suelos y los riesgos de la erosión fueron evaluados. (yb)

Likewise, the use of the article is also compulsory with nouns in singular on most occasions. This type of omission is not so frequently found, but it also occurs in:

- (2) This study analyses the use of an agroforestry system to reclaim landscape spoiled.
Este estudio analiza el uso de un sistema agroforestal para recuperar paisaje dañado. (gt)

As in the previous example Yahoo Bablefish has included the article in its translation:

- (3) Este estudio analiza el uso de un sistema de la agrosilvicultura de reclamar el paisaje estropeado. (yb)

There are also examples of omission of prepositions in the translated texts. These omissions can be present in both target texts or just in one, as in example (4) where the article is absent just in the case of Yahoo Bablefish:

- (4) discussed the challenges
discutieron los desafíos (gt)
discutieron los desafíos (yb)

- (5) influence the conversion process
 influyen en el proceso de conversión (gt)
 influncian el proceso de conversión (yb)

As the use of articles and prepositions is in some results correct in one of the machine translations, it means that it would be possible to improve the internal computer algorithm to improve the final output.

(2) Addition

The additions are found in different word categories. In the first case, the determiner *one hundred*, when accompanying a noun, does not need a preposition in front of it; it does not need either to be turned into a noun *cientos*, which makes the noun phrase absolutely wrong employing a noun as modifier of another noun. This word structure is sometimes found when calquing English noun phrases in Spanish, but it is not the usual pattern.

- (6) one hundred high-level Ministers
 de cien ministros de alto nivel (gt)
 cientos ministros de alto nivel (yb)

Additions can also affect verbs. In example (6) the verb *funcionar* is not reflexive in Spanish. That is why the first translation provided by Google Translator is correct because the verb takes the particle *se*, but completely incorrect with a verb like *funcionar*. Likewise, the use of two prepositions one after another is not grammatical in this same example.

- (7) Scenarios are run on a 2013 baseline.
 Los escenarios se ejecuta en una línea de base 2013. (gt)
 Los panoramas se funcionan con en una línea de fondo 2013. (yb)

In the following example, we are dealing with the addition of a preposition that is not correct in Spanish although necessary in English to indicate that it is an infinitive.

- (8) To review
 Para supervisar (yb)
 Para revisar (gt)

(3) Mistranslation

Wrong formation. Sometimes the translator divides the lexical unit into different components creating a new word in the target language that is completely wrong. Thus, from *agroforestry* a new coinage *agrosilvicultura* is provided. Both the *Diccionario de la Lengua Española* (henceforth, *DRAE*) and *Diccionario de uso del español* record *agro* and *silvicultura*, but there is no such a term as *agrosilvicultura*, while the adjective *agroforestal* is properly documented.

- (9) an agroforestry system
 un sistema agroforestal (gt)
 un sistema de la agrosilvicultura (yb)

In addition, there are several examples of wrong morphological word-formation. The right verb in number (9) is *influir*, although the backformation *influnciar* is widely spread among some native speakers in spite of the fact of not being correct:

- (10) influence
 influncian (yb) → influyen

In the next example, the machine translator creates a verb which follows the regular pattern of applying the first conjugation ending to the noun from which it derives; however, the prevalent word in Spanish is formed from the base noun plus the suffix – *izar*:

- (11) monitor
 monitorear (gt) → monitorizar

Use of the possessive. While the possessive determiner is frequently found in English, in Spanish it is not so extensively used, as in the following example where the definite article would be the common determiner in such a context:

- (12) its Twenty-Fifth Session in Ottawa
 su vigésima quinta sesión en Ottawa (gt)
 su vigésimo quinto período de sesiones (yb)

The concept is not a direct lexical equivalent. The word could be right in other contexts but not in this one, which implies the use of a substituted concept. This holds true in the case of *atar* which is used to mean ‘to link, to bind’, but not ‘to attach a document or an appendix to a report’:

- (13) The complete list of participants is attached as Appendix I to this report.

La lista completa de participantes se adjunta como Apéndice I del presente informe. (gt)

La lista de participantes completa se ata como el apéndice I a este informe. (yb)

False friends. These are paronyms in both languages, because they share a similarity in form but not in meaning. Thus, in English the lexeme *vegetables* is used to designate ‘greens, leafy vegetables and others eaten by human beings and animals’; although theoretically the Spanish term *vegetal* has the same meaning of ‘living organism belonging to the vegetable kingdom’, when we talk about edible vegetables the words *hortaliza* and *verdura* are preferred.

- (14) Field vegetables
 Vehículos de campo (yb)
 Vegetales de campo (gt) → hortalizas

Another false friend is *paper*, which can have different meanings in an academic context, where *trabajo* or *comunicación* are used, but not *papel*.

- (15) The paper presents an alternative approach.
 El papel presenta un acercamiento alternativo. (yb)
 El documento presenta un enfoque alternativo. (gt)

Another example that implies the literal translation of the term is number (15) where *leche orgánica* is provided by both translators instead of *leche ecológica*. The term *orgánica* exists in Spanish, but has a different meaning:

- (16) organic milk
 leche orgánica (gt) (yb)

In (16) *ganar* does not collocate with understanding. For instance, you can *ganar* money, a prize, but not *comprensión*.

- (17) gain a better understanding
 ganar una mejor comprensión (yb)
 obtener una mejor comprensión (gt)

(4) Untranslated concept

Acronyms that are left in English or whose constituents are translated into Spanish but the initials follow the English order are considered as untranslated items. Just on one occasion the NGO has been correctly rendered into Spanish ONG:

(18) The Intergovernmental Group on Bananas and Tropical Fruits (IGG) is exploring what steps can be taken with non-governmental organizations (NGOs).

Del Grupo Intergubernamental sobre el Banano y las Frutas Tropicales (IGG) está estudiando qué medidas se pueden tomar con las organizaciones no gubernamentales (ONG). (gt)

El grupo intergubernamental en los plátanos y las frutas tropicales (IGG) está explorando qué pasos se pueden tomar con las organizaciones no gubernamentales (ONGs). (yb)

Some other examples of untranslated terms have been found when the Yahoo Bablefish was used:

(19) animal health care, housing and biosecurity

cuidado médico animal, cubierta y biosecurity (yb)

atención de la salud animal, la vivienda y la bioseguridad (gt)

(20) The web site

El web site (yb)

El sitio web (gt)

(5) Syntactical errors

Within the syntactical errors we have distinguished several kinds of errors: a) wrong use of verbal forms; b) reordering errors; c) lack of gender and number agreement; and d) lack of subject and verb agreement.

Wrong use of verbal forms. Among the wrong use of verbal forms, the passive voice is pervasive in the translated texts. Passive can be formed in Spanish by means of the auxiliary *ser* + past participle. Nevertheless, this kind of passive is very rarely used. Unlike English, another type of passive voice can be formed with the verb in the indicative mood preceded by the pronoun *se*. The target texts show the periphrastic passive with *ser* + past participle, as it is found in English. On the contrary, very few instances of reflexive passive with *se* are found all through the target texts. Thus, the instances of passive are numerous.

(21) Visual impacts of the vegetation recover of the area were observed.

El impacto visual de la vegetación recuperarse de la zona se han observado. (gt)

Los impactos visuales de la vegetación se recuperan del área fueron observados. (yb)

A special case of passive occurs with the verb *asistir* ('to attend') which is transitive in English and can be used in the passive voice, while in Spanish is intransitive and does not have this possibility.

(22) The meeting was attended by 213 delegates.

La reunión contó con la participación de 213 delegados. (gt)

La reunión fue asistida por 213 delegados. (yb)

Reordering errors. As head nouns in noun phrases in English tend to be modified by adjectives as well as nouns, when these nouns are translated into Spanish the order will be the opposite: The head takes the first position followed by the adjective and finally the noun is turned into a prepositional phrase.

- (23) longitudinal case studies
 los estudios de caso longitudinales (yb)
 los estudios longitudinales de caso (gt)

Likewise, the following example shows clearly the problems of reordering the elements of a long noun phrase in English with several premodifiers when it is translated into Spanish: “*en la Conferencia sobre Investigación Ecológica celebrada en el Reino Unido en 2002*”. It implies not only a reordering of the elements but the need of prepositions and articles. In this case, both translators produce a totally incorrect sentence in the target language.

- (24) at the UK Organic Research 2002 Conference
 en la conferencia orgánica BRITÁNICA de la investigación 2002 (yb)
 en el Reino Unido Orgánica Investigación 2002 de la Conferencia (gt)

Lack of gender and number agreement. As gender is an inherent grammatical feature of all nouns in Spanish, it is necessary to agree the noun with every adjective, pronoun, determiner or noun in apposition that refer to a given noun. Thus, in the case of the person mentioned in the example (24), the feminine reference must be present in the article (*la*), the title (*Doctora*) and her post (*Directora*):

- (25) Dr. Anne MacKenzie, Director-Generalla Dra. Anne MacKenzie, Director General (gt)
 el Dr. Anne MacKenzie, director general (yb)

Lack of subject and verb agreement. Likewise, although sometimes the collective nouns can be used in plural in English, the tendency in Spanish is to follow the grammatical agreement rather than what is known as agreement ad sensum, that is, according to the meaning. It follows from here that, in the case of *grupo*, the correct agreement should be in singular, rather than in plural:

- (26) The IGG discussed the topics of fair trade and organic bananas.
 El Grupo Intergubernamental discutieron los temas de comercio justo y el banano orgánico. (gt)
 El IGG discutió los asuntos del comercio justo y de los plátanos orgánicos. (yb)

2. Conclusions

In the previous pages we have presented one of the few studies that have been carried out applying the Error Analysis Approach to machine translations. The error analysis categories applied to human learning errors are often reproduced when the translators are human beings; this work has made use of these categories to analyse machine translator outputs. We have based our analysis on parallel corpora which are understood as “original texts in one language and their translations into one or several other languages” (B. Altenberg/ S. Granger 2002: 8). We have illustrated our findings with examples from these corpora: the original English corpus and two Spanish corpora which

are the result of the automatic translations generated by two translators: Google and Yahoo Bablefish.

The focus of this work has been the classification of errors to determine whether they are similar to those produced by human beings and to what extent they are specific to computer tools and how we can compare two machine translators by their outputs. The taxonomy is established following five categories: omission, addition, mistranslation, untranslated concepts and syntactical errors. This has been just a sample of research that needs further exploration, but generally speaking it can be stated that the final texts offered by Google Translator are more accurate and present fewer errors than those displayed by Yahoo Bablefish. Nevertheless, both automatic translators produce quite a number of mistakes, so we should try to implement some general rules to improve the quality of the target texts generated automatically.

According to the results obtained after analysing the translations, the most common errors are those of mistranslation, followed by syntactical errors, omission, addition and, finally, untranslated concepts. The automatic translators do not often add terms or forget to translate one when working, but very often they select the wrong word in a context, or a wrong formation of terms, or produce a literal translation of a concept, or not totally correct order of elements in a sentence or lack of correct syntactical rules. Thus, in the light of the data we can conclude that the same kind of errors produced by human translators is generated by automatic translators. But, how can we 'teach' automatic translators to improve their results? Maybe we can start with the lack of gender and number agreement and the order of the elements in a sentence. Some rules can be defined so that the machine can process the information and obtain better results. Although some of the mistranslation errors are quite difficult to correct, because in most cases it seems complex to explain why a specific term is used in such a context, often we are dealing with a word selection based on the usage, rather than on grammatical rules. Nonetheless, we should try to implement some general rules to improve the quality of the target texts generated automatically.

References

- Altenberg, B./ S. Granger (2002), *Recent trends in cross-linguistic lexical studies*. In: B. Altenberg/ S. Granger (eds), *Lexis in Contrast. Corpus-Based approaches*. Amsterdam, 3–48.
- Borrego Ledesma, I. (2001), *Errores y aprendizaje*. In: A. Larrañaga Domínguez (coord.), *FORMA. Formación de Formadores. Vol. 2 Interferencias, Cruces y Errores*. Madrid, 85–100.
- Callison-Burch, C./ M. Osborne/ P. Koehn (2006), *Re-evaluating the Role of BLEU in Machine Translation Research*. In: *EACL-2006: 11th Conference of the European Chapter of the Association for Computational Linguistics*, 249–256.
- Colina, S. (2009), *Further evidence for a functionalist approach to translation quality evaluation*. In: *Target* 21/2, 235–264.
- Condon, S./ D. Parvaz/ J. Aberdeen/ C. Doran/ A. Freeman/ M. Awad (2011), *Machine Translation Errors: English and Iraqi Arabic*. In: *ACM Transactions on Asian*

- Language Information Processing 10/1. (URL doi>10.1145/1929908.1929910). [Retrieved April 2015].
- Koponen, M. (2010), *Assessing Machine Translation Quality with Error Analysis*. In: Electronic proceedings of the KäTu symposium on translation and interpreting studies 4, 1–12. (URL http://www.skil.fi/@Bin/40701/Koponen_MikaEL2010.pdf). [Retrieved March 2015].
- Kutuzov, A. (2008), *Using descriptive mark-up to formalize translation quality assessment*. In: Translation industry and information supply in international business activities: materials of international conference. Perm, 90–101.
- Luengo, J.C./ M.Á. Sicilia/ S. Sánchez (forthcoming), *Evaluating MT Systems for Metadata Records in Organic Edunet*.
- Popovich, M./ H. Newy (2011), *Towards Automatic Error Analysis of Machine Translation Output*. In: Computational Linguistics 37/4, 657–988. (URL <http://aclweb.org/anthology-new/J/J11/J11-4002.pdf>). [Retrieved April 2015].
- Moliner, M. (2008), *Diccionario de uso del español*. Madrid. 3rd ed. on DVD.
- Rabadán, R./ B. Labrador/ N. Ramón (2009), *Corpus-based contrastive analysis and translation universals*. In: Babel 55/4, 300–328.
- Real Academia Española. *Diccionario de la Lengua Española (DRAE)*. Available in URL www.rae.es
- Santos Gargallo, I. (1993), *Análisis Contrastivo, Análisis de Errores e Interlengua en el marco de la Lingüística Contrastiva*. Madrid.
- Vázquez, G. (1999), *¿Errores? ¡Sin falta!* Madrid.