



---

# Universidad de Valladolid

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

## TRABAJO DE FIN DE GRADO

Marketing and Advertising Translation: Humans vs  
Machines in the field of cosmetics

Carmen Alonso del Caño

Tutor: Carlos Herrero Quirós

2018-2019

**HERRERO  
QUIROS CARLOS  
- DNI 09271727J**

Firmado digitalmente por HERRERO QUIROS  
CARLOS - DNI 09271727J  
Nombre de reconocimiento (DN): c=ES,  
o=UNIVERSIDAD DE VALLADOLID,  
ou=CERTIFICADO ELECTRONICO DE EMPLEADO  
PUBLICO, serialNumber=IDCES-09271727J,  
sn=HERRERO QUIROS, givenName=CARLOS,  
cn=HERRERO QUIROS CARLOS - DNI 09271727J  
Fecha: 2019.06.27 19:11:57 +02'00'





## ABSTRACT

This undergraduate thesis focuses on a very specific field of specialized translation: advertising and marketing translation. Indeed, the high degree of specialization involved in this activity provides a testing ground for a reconsideration of the importance of the human translator and a reformulation of their role. The constant development of new technologies creates ever more sophisticated translation programs, which in turn revives the long-standing machine vs human translation debate. The aim of this project is to conduct a practical exercise targeted at verifying whether specialization in translation always requires the supervision of humans equipped with the relevant linguistic knowledge and technical background, or whether, on the contrary, machine translation can at present provide valid enough results and a sufficient level of reliability.

Keywords: translation, human, machine, advertising, marketing, specialization

## RESUMEN

El presente Trabajo de Fin de Grado se centra en un campo muy concreto de la traducción especializada: la traducción para la publicidad y la mercadotecnia. De hecho, el alto grado de especialización que implica esta actividad proporciona un campo de pruebas para una reconsideración de la importancia del traductor humano y una reformulación de su papel. El desarrollo creciente e ininterrumpido de las nuevas tecnologías está produciendo programas de traducción cada vez más sofisticados, lo que a su vez reaviva el viejo debate que confronta la traducción humana y la traducción automática. El objetivo de este proyecto es llevar a cabo un ejercicio práctico destinado a verificar si la especialización en la traducción siempre requiere la supervisión de personas con la formación lingüística y los conocimientos técnicos pertinentes, o si, por el contrario, la traducción automática puede en la actualidad proporcionar por sí sola resultados suficientes y un nivel suficiente de fiabilidad.

Palabras clave: traducción, humano, automática, publicidad, marketing, especialización.





## Index

1. Introduction .....	1
2. Methodology.....	3
3. Theoretical background .....	4
3.1. Human translation (HT).....	4
3.2. Machine translation (MT).....	5
3.2.1. Pioneer years .....	6
3.2.2. First-generation systems .....	6
3.2.3. The quiet years.....	7
3.2.4. Second-generation systems .....	7
3.2.5. Third-generation systems .....	7
3.3. Main stages of the MT process .....	8
3.4. Machines and Humans.....	9
3.5. Marketing and Advertising Translation (MAT) .....	10
4. Presentation of the data: Corpus on cosmetics .....	14
4.1. Corpus methodology .....	16
4.2. Software tools .....	18
4.3. Corpus analysis results.....	25
5. Translation analysis of cosmetic labels .....	27
6. Conclusion.....	35
7. Bibliography .....	37



## 1. Introduction

Technology is ubiquitous in our everyday life and the purchasing and sales sector is by no means an exception. E-commerce and online shopping have become a staple of our modern world's consumer economy and a booming business that grows at an unprecedented rate across the globe. The Internet enables shoppers to promptly find whichever article they want to buy and provides retail companies with a unique tool to exponentially increase sales and to massively promote their products. The new web-based forms of communication are intensively exploited in advertising and marketing campaigns that are designed to influence consumer habits and therefore rely on a persuasive, appellative component that is ultimately linguistic, at least in part, and often needs to be conveyed in different languages.

In a world where we are daily surrounded by a huge amount of beauty products (skin care, hair care, colour makeup, etc) irrespective of age or gender, the flourishing sector of cosmetics is a case in point. Both technical descriptions and seductive messages need to be phrased for marketing purposes, and translation agencies like the UK-based Andiamo provide “specialist cosmetics translations for the health and beauty” including “everything from ingredients to marketing copy” (Andiamo Blog, Robertshaw). As with all advertising translation, here too the professional translator must possess a sound acquaintance with the sector's terminology, but also the language skills needed to produce accurate and persuasive messages and, last but not least, an awareness of culture-specific constraints that may require an adaptation of the original texts to the customs, traditions, conventions, etc. of the target audience.

On the other hand, the constant development of new technologies creates ever more sophisticated translation programs. Indeed, we have now reached a point where Machine Translation Systems (MTS) are able to provide target texts in practically no time. As a result of these developments, the old machine vs human translation debate is being constantly updated. The aim of this project is to conduct a practical exercise targeted at verifying whether specialization in translation always requires the supervision of humans equipped with the relevant linguistic knowledge and technical background, or whether on the contrary, machine translation can at present provide valid enough results and a sufficient level of reliability. To this end, and in order to sufficiently narrow the scope of our work, we will focus on a selected bilingual corpus of cosmetic descriptive labels that we compiled from Amazon's online catalogues. Our initial assumption is that humans



cannot be currently equated to machines in an activity that requires the kind of complex, specialized knowledge pointed above, even though human-machine collaboration can be an interesting choice to contemplate.

Before we get to the comparative exercise proper, we will provide a general background to the issue under examination. The following section details the sequence of our inquiry.



## 2. Methodology

This paper is divided into three main sections. In the first place, we address several theoretical considerations around human, machine and marketing and advertising translation. The purpose of this background section is to introduce the context of machine translation, a powerful current development whose modern inception can be traced back to the mid-twentieth century, and of the field of specialization that we put the focus on in the present study: marketing and advertising translation.

Secondly, we carry out a compilation of a bilingual corpus consisting of descriptive labels of cosmetic products in both Spanish and English with the intention of studying the phraseological units that we can find in the sample within the general framework of comparative linguistic analysis. More specifically we track the collocations involving the most frequently repeated words in the compilation so as to discern the main grammatical patterns that define each of the languages involved within the particular context of sales-targeted descriptions.

Once such patterns are established for the two languages, we test our findings by implementing a translation trial in the third and practical section of this project. More particularly, we select an amount of 10 descriptive labels (5 in English and 5 in Spanish) from the compiled bilingual corpus consisting of 139 items per language. The next step consists in translating this selection, first by using two machine translation systems (DeepL and Google Translator) on the initial assumption that their performance will not be satisfactory in the context of such a specialized language domain as cosmetics marketing; and then by providing our own human translation on the hypothesis that it will turn out to be more accurate and reliable. All the translated segments are collected in a separate document called “Annexes” that is attached in CD format (given its length) to this undergraduate thesis.

Finally, the sections devoted to analysis and conclusions summarize our findings and establish the validity or non-validity of our initial hypotheses.



### 3. Theoretical background

#### 3.1. Human translation (HT)

While modern translation studies are a relatively recent phenomenon, the activity of translation is as old as the dawn of civilization. Its history goes back over 4000 years, when it became an essential piece for the communication between different cultures and civilizations. Yet in its beginnings, human translation was very different from today's professional activity. The German hispanist Hina (31), for example, reminds us that former translators were not committed to achieving perfect equivalence with regard to the culture and style of the source text. Instead, their job was successfully performed if they were able to produce a legible text, where the meaning expressed in both languages was reasonably similar. However, progress occurred here as in all other spheres of human activity, and in the course of time translators had to adapt their craft to changing demands in a historical process that ultimately led to translation becoming a powerful industry, if only because it was called upon to serve booming economic sectors like the one that constitutes the focus of this graduation project.

What we understand nowadays as central to the translator's job involves specific skills: first, the translation worker must have a broad cultural/technical knowledge that encompasses different disciplines, from linguistics and pragmatics to multiple areas of specialization as demanded by the particular register (medium, tenor and domain) of the text that is being translated. Secondly, a good translation would be the one that best reproduces the content of the source text, both in terms of meaning and the way it is expressed, that is, the style (and in our case the marketing and technical register) of the original text. In practical terms, however, translation is confronted with a number of difficult hurdles that need to be overcome. On the one hand, the impossibility of setting-up perfect, error-proof methods that ensure the procedure to be followed in order to guarantee an optimal result (Hina 35). On the other hand, we need to bear in mind culture as the fundamental context upon which many other translation-related aspects hinge. In the world of advertising, for example, translation often veers towards cultural adaptation or localization.



The languages involved, their natural grammatical and lexical differences, are the first thing a translator must get to grips with, but also their socio-cultural contexts and the field of specialisation determine the required translation procedures. After all translation is all about communication and exchanges between individuals, communities and nations, thence its historical importance as a driver for the development of our world. As García Yebra puts it (267) “es una tarea tan importante para el progreso humano en general y para la estima y comprensión entre los pueblos que hablan distintas lenguas”. More specifically commerce and trade exchanges were historically major drivers of the development of modern translation, since every business relationship between people or organizations that operated in different languages required its services as a regular part of trade operations. Modern-day marketing and advertising are none the less dependent on the input of translation, whether explicit or implicit.

Translation, in short, is just as important as it is complex, and the translator’s job demands a sound training and a constant professional update so as to keep pace with the fast progress of the activity itself and of the many sectors it serves. In order to succeed in this, and more specifically in the context of specialised, technical professional translation, translators must be aware, among other things, of the new technological tools that are constantly emerging and being upgraded. This leads us to the topic of Machine Translation Systems (MTS), which we will address next.

### **3.2. Machine translation (MT)**

Since the early years of machine translation, the emergence of the new technologies in the translation field has always been surrounded by controversy in relation to the advantages or disadvantages they may entail, as well as the rejection or support of these new tools. Machine translation can be defined as “the production of text in one natural language from that in another by means of computer procedures” (Collins Dictionary), that is, the involvement of software in the role of translating. As in the case of human translation, we can also establish a historical framework that reflects the different stages that this new field has gone through.

Leaving aside distant historical antecedents that are rather related to proposals for the creation of universal interlinguas, “proposals for machines to perform dictionary



consultation or translation did not come until technological developments in the early twentieth century” (Hutchins), while modern MT proper did not start to be developed until the 1950s. Between then and today’s exponential growth of web- and apps-based translation systems in a digitalised world where otherwise human professionals are not enough to cover translation demands on time and at low costs, seven decades have elapsed which in terms of the development of this technology may be divided into five main phases which we summarise next.

### **3.2.1. Pioneer years**

This first stage began in 1949 with the Warren Weaver memorandum, which established the real beginning of MT research (Quah 59). The main proposal stated in this memorandum had to do with the use of cryptography in order to automate the translation process. Nevertheless, funding was critically needed in order to develop fully automated high-quality MT systems (FAHQMT), and it came from the US and the Soviet Union. While Weaver’s intention and the mission of the Rockefeller Foundation where he worked were scientific and philanthropic, research on MT was primarily fuelled by the Cold War atmosphere of mutual fear and suspicion between both countries and the pursuance of their geopolitical ambitions.

### **3.2.2. First-generation systems**

The early MT systems were known as “toy systems”, relied on word-for-word translation and could not process more than 250 words. They followed the “direct translation approach”, since they worked without the application of rules but simply transferring the word’s meaning from one language to the other. These early developments created high expectations and a positive outlook on the part of investors. It was in 1966 when the Automatic Language Processing Advisory Committee (ALPAC) decided to investigate the reliability and feasibility of MT systems. The ALPAC report concluded that this MT system was “slower, less accurate and twice as expensive as human translation” (Hutchins 439). Accordingly, there were no more financial aids by governments and the initial optimism vanished.



### **3.2.3. The quiet years**

This period encompasses the decade from 1967 to 1976 and it directly derives from the ALPAC report negative results. According to Quah (61) MT research entered a phase of decline especially in the US, but there were still some groups that continued to be active in this field of inquiry, both in the US and elsewhere.

### **3.2.4. Second-generation systems**

The 1980s saw a period of resurgence for MT systems that led to the launch of several devices like the systems named SPANAM (Spanish-English) or the ENGLSPAN (English-Spanish). Those years also witnessed the emergence of new, rule-based approaches to automated translation. These involved the use of semantic and syntactic rules for the analysis of a source language prior to dumping its content into the target language. Such rules ensure that the process of translation relies on a systematic all-level linguistic analysis that in turn underpins an equivalent rewording of meaning from STL to the TL. Following Hutchins (448), rule-based approaches include two distinct types: interlingua (the source text is transformed into an abstract, language-independent representation and then, several target language texts result from a process of conversion) and transfer (it relied on bilingual dictionaries and is at least partly dependent on the specific language pair).

### **3.2.5. Third-generation systems**

The last stage before the modern years started in the 1990s and comprises the development of corpus-based approaches which, according to Somers (114), “use a corpus or a database of already translated examples, and involve a process of matching a new input against this database to extract suitable examples”. It follows that corpus-based systems can in turn be divided into two types: example-based and statistical-based approach. In the former one, the program counts on bilingual corpora of translation pairs and the closest examples for both languages are matched together. In the statistical-based approach, algorithms are employed in order to create formulas that are applied to the software in the process of translation.



None of these systems is perfect and able to warrant a reliable result. For that reason, innovative approaches (hybrid and interactive MT systems) are the preferred ones in the latest stage of this process, the modern years. Still at an experimental stage (Quah 85), they combine the previous MT architectures and have already the great scope and huge possibilities of their implementation.

As noted some lines above, machine translation has been under the spotlight of controversy in recent years, since translation profession felt threatened by IT developments. It is and will presumably continue to be debatable whether the human translator's involvement is imperative or machines are enough to achieve results of acceptable quality. Even so, it is undeniable that the current reputation of MT is on the rise. Systems have evolved from simple word-for-word translation to sophisticated and complex analyses and transformation rules.

### **3.3. Main stages of the MT process**

In a nutshell, the source text goes through five main processes in order to be converted into a target text (Alarcón 722). First, we have the pre-editing stage or pre-processing, which involves the correction of mistakes regarding grammar, spelling or punctuation (including typographical aspects such as parentheses, apostrophes, spaces between words or hyphens). The main stage in MT is the analysis of the ST, where the procedures involved have to do with the determination of the deep structure of the sentence and identification of the components, the relationships between its syntactic and semantic elements and the cases of homophony. In the next stage occurs the transfer between the source and the target texts. Here, the different rules regarding grammar, lexis and context operate so that meaning is reassigned to the final target product. The synthesis of the target text is the next step. Here, tense, number and form must be assigned to the final product. The final stage in the machine translation process is post-editing. It consists in the revision and correction of all the errors that may remain in the target text. Post-editing is a process that involves human intervention in order to rectify software failures and close the whole process. As Wills (222) notes "...it is the person of the human translator who has the ultimate control over the MT operations".



### 3.4. Machines and Humans

In view of our brief overview of MT and the way it works, we can assume that both MT and HT are more intertwined and interrelated with one another than we may expect. This brings us back to the aforementioned debate about the rejection/acceptance of comprehensive translation software. The above considerations are compatible with the claim that while resorting to a machine, human aid is also necessary for pre and post-editing tasks in translation projects. We may argue that MT provides undeniably huge future prospects and that already it is modifying and modulating the work of human translators in several ways, and as Pigott (2) asserts “. . . someday the majority of translators could and would become MT post-editors (whether in 10, 20 or 30 years time”. The marketplace for translation expands at huge speed, and the hurry for delivering quick and cheap translations will create post-editors instead of translators one day (Pym 488).

A substantial number of views around this topic agree on this point, and open the door to collaborative work between machine and humans in any translation field. The specialization and training of human translators, as well as the development and improvement of machine translation, are essential in achieving the best performance. For any translation practitioner these days who does not want to see his professional scope limited, enlarging his/her educational background so as to include not just skills in the fields of linguistics (including pragmatics or semantics), interlinguistic and intercultural communication, but also a solid command of technological instruments as a first-order working tool. It goes without saying that in defining the right “mix” of skills, we need to bear in mind the aim or purpose of our intended translation (its “skopos”) as well as the specialized (technical, professional) domain or market sector where it operates. It is to these considerations that we now turn.



### 3.5. Marketing and Advertising Translation (MAT)

Translation is a discipline of knowledge that can potentially encompass all areas of human (and economic) activity. While flexibility and multifacetedness is a highly desirable asset in translation, no single translator, as we may expect, can aspire to attain a full command of the whole repertoire of specialized languages. Ideally, therefore, one of the first decisions prospective translators should make is the field or fields of specialization they want to focus on so as to incorporate academic training in such areas as part of their background.

In the present study, we decided to focus on a very specific branch of translation that is actually ubiquitous without our actually realizing it: marketing and ad translation. Translation and marketing/advertising are each complex professions in their own right. When combined, their output is a highly specific product that requires an equally specialized handling. At this point, it may be worth the while advertising in the context of Spain's national legislation:

Toda forma de comunicación realizada por una persona física o jurídica, pública o privada, en el ejercicio de una actividad comercial, industrial, artesanal, con el fin de promover de forma directa o indirecta la contratación de bienes muebles o inmuebles, servicios, derechos y obligaciones<sup>1</sup>.

The formula contains the key components of marketing and advertising communication between producers and consumers with the final aim of selling particular products. Nevertheless, what really matters is the purpose of such communication, that is to say, the effectiveness of the advert, which will be measured in terms of the number of sales or the reputation of the product being advertised. In this context, we should bear in mind the current phenomenon of market globalization, which in the case of advertising campaigns (and the translation work that they explicitly or implicitly involve) plays a critical role. Indeed, globalisation provides the backdrop for us to discern two major tendencies in product marketing that have a direct repercussion on advertising-gearred translation: internationalization and localization. De Pedro (8) defines it as follows:

---

<sup>1</sup>Any form of communication carried out by a natural or legal person, public or private, in the exercise of a commercial, industrial or artisanal activity, in order to promote, directly or indirectly, the procurement of movable or immovable property, services, rights and obligations. [My translation]



La internacionalización es la producción de un mensaje publicitario que atraiga a los consumidores de los diversos ámbitos culturales a los que vaya dirigido, sin hacer concesiones a las diferentes culturas, y permite crear una imagen de marca global, . . . mientras que la localización es la adaptación del mensaje publicitario a los distintos mercados, teniendo en cuenta sus respectivas idiosincrasias, y se ajusta a las necesidades y preferencias de grupos de consumidores específicos<sup>2</sup>. (8)

This author states that both options have advantages, but that legal framework restrictions of each country are a key point to consider, since a common set of global values is difficult to secure, and even though some topics might be legally accepted in a given country, the message could be at odds with locally established cultural values.

In spite of this, De Pedro (8) suggests that internationalization is both legitimate and effective, but it also requires a significant amount of professional work on the part of the translator, since s/he has to focus on multiple aspects —linguistic and cultural differences— which require a robust background knowledge and intense documentation work, all of which in turn implies a substantial amount of interdisciplinary training. The interface between advertising and culture, and its consequences for translation, cannot be overemphasized.

We have already said that the advertising text requires a complex translation, and we can add another reason in support of such a statement, since according to González (59) in advertising there is a mixture of both verbal and iconic elements that eventually produce a hybrid textual type. The point is accurately made by Rabadán (131):

Pero los anuncios no son únicamente lingüísticos. Casi siempre incorporan imágenes, sean estáticas como en la publicidad impresa o dinámicas, como en televisión, y estas imágenes contribuyen al significado global del texto. Siempre que interviene más de un código en el mensaje, la discusión se centra en el medio en el que sucede la comunicación: oral, visual, escrito, etc. Y es en este tipo textual híbrido donde se centra la controversia entre traducción/adaptación<sup>3</sup>. (131)

---

<sup>2</sup>Internationalisation is the production of an advertising message that attracts consumers from the various cultural backgrounds to which it is addressed, without making concessions to the different cultures, and makes it possible to create a global brand image . . . while localisation is the adaptation to the advertising message to the different markets, taking into account their respective idiosyncrasies, and it meets the specific groups of consumers' needs and preferences. (8) [My translation]

<sup>3</sup>However, adverts are not only linguistic. Most of the time they include incorporate images, whether static as in print advertising or dynamic as in television, and these images contribute to the overall meaning of the text. Whenever more than one code is involved in the message, the discussion focuses on the medium in which communication takes place: oral, visual, written, etc. Moreover, it is in this hybrid textual type where controversy between translation/adaptation is centered. (131) [My translation]



In the present study, we shall be dealing with a very specific informative text-type that plays a key role in the marketing of an equally specific commodity: the descriptive labels of cosmetic products from Amazon's online catalogues, where such a correlation between image and text is indeed present. According to Bueno (59), this correlation is known as interpretative or iconographic code, by contrast with two other codes defined by the same author: the linguistic (carrying verbal content) and the iconic one (which is made up of the visual elements and non-verbal ones). As we will focus on product description, that is the verbal code, we will zoom in on the linguistic code exclusively. In advertising markets, as we have already stated, what matters is engaging the consumers' attention at first sight, so the first important element is the name of the product. While translators do not directly deal with graphic design or typeface, they must be aware of the implications of such non-linguistic choices insofar as they also contribute to the ad's total meaning. The title is the first thing the consumer is going to see, so it has to be eye-catching and stand out among the rest of the text. As Valdés (103) states, "en la traducción publicitaria, los titulares y eslóganes plantean la mayor dificultad debido a la gran cantidad de recursos empleados para atraer la atención del receptor<sup>4</sup>". We may argue that, while also characterized by a strongly descriptive (even technical) verbal component, cosmetic labelling also exhibits some of the suggestive quality of advertising taglines referred to in the above quote. Label translation also have to be effective in economically and tightly conveying the ST central message, while at the same time not omitting any of the required content, so that potential buyers are fully informed of the main characteristics of the product they may purchase.

---

<sup>4</sup>"In advertising translation, headlines and slogans pose the great difficulty due to the large amount of resources used to attract receivers' attention." (103) [My translation]



In order to be successful in their assignments, translators need to consider various strategies and select the one that best suits their work. The first one is adaptation, including the sometimes slippery search for equivalence in the case of idiomatic expressions. A potential choice is also literal word-for-word translation, a less demanding, time-consuming strategy in the case of large advertising campaigns than a total rewriting in the service of cultural equivalence. Also economic reasons may be behind the decision—otherwise quite common these days—to opt for a “no translation” strategy, which as Valdés (256) puts it, “es la elaboración y difusión de un mismo mensaje en múltiples mercados como parte de una única campaña publicitaria<sup>5</sup>”.

While this course of action may imply a number of shortcomings like the loss of meaning or cultural correlation to the target audience, many companies tend to see this option not just as a quick and cheaper procedure and a way to ensure translation errors, but also as an attractive strategy in an ever more globalized world where the source languages (e.g. English or French) in advertising slogans enhances the sophistication and international appeal of the marketed product. On the other hand, a copywriter-translator (or the advertiser that commissions the translation job) may also choose to create a new text, due to the problems that may be encountered when transferring cultural values from one language to the other, but also when the appellative function of the advertising text cannot be fulfilled by means of plain adaptation. Finally, in some cases, it is not necessary to engage in this kind of recreation, but simply to omit such parts as are not strictly necessary to convey the marketing message.

Once we have seen the difficulties involved in such a specific variety of translation and the several strategies available to the translator, we will try to test the latter in a practical case by following all the steps that a professional translator would take. Additionally, we will try to ascertain the reliability of machines or automatic translation in such a specific setting as cosmetic label translation and the kind of aspects that still need to be improved in such a scenario.

---

<sup>5</sup>Is the elaboration and diffusion of the same message in multiple markets as part of a single advertising campaign.” (256) [My translation]



#### 4. Presentation of the data: Corpus on cosmetics

With the purpose of achieving precise and reliable results, the first step we took in this study was the compilation of a corpus on the field of cosmetics. A corpus is a great and useful tool for translators, since it represents real samples of a given language and, therefore, affords the opportunity to work with actual contexts of use. We can define it as a collection of recorder texts that are held to be representative of a given language and enable researchers to draw different conclusions regarding how that language is used according to a descriptive linguistic analysis. Corpora can constitute powerful reference tools and sources of an abundance of data that may be classified and studied for specific purposes. Before compiling a corpus we need to establish its aims and applications in the context of our study goals. Since corpora can serve different research aims, they can be classified in relation to such aims. In the following lines, we will proceed to explain our own goals in relation to the use of this research tool.

Our approach is empirical, since the data has been collected and analysed in order to draw individual conclusions from the observation of this particular compilation rather than from theoretical models. The analysis has been both qualitative and quantitative; the latter implies the gathering and interpretation of information taking into account specific aspects. The numerical results were not our ultimate goal, but they were necessary to achieve an output that could be analysed. The subsequent qualitative analysis relies on the concept of representativeness, since the samples we are working with need to accurately reflect the characteristics of the language register we are exploring, i.e; the language of cosmetology and, more specifically, cosmetics labelling. The corpus we have compiled is representative of one single domain (cosmetic labelling) in two different languages, and it is also balanced because the ration between both sets of data (English and Spanish) is commensurate.

More particularly, and in terms of corpora classification, in this study, we have compiled a bilingual comparable corpus in English and Spanish. In other words, our collection includes items in two different languages upon which we perform comparative analysis. It is also specialized because we are focusing on concrete and particular aspects of the languages use (a specific domain) and it does not intend to represent either language as a whole, since we are only interested in a particular field, cosmetics. This is related to its openness, since its growth is not constrained—it can be constantly expanded as more



products are added to the digital platforms or other sources are tapped. In conclusion, ours is a corpus-based inquiry targeted at studying the behaviour of phraseological units based on their occurrence and patterns of use. Once the data are collected, an essential step is their annotation, that is, the addition of extra information to that data that make up the collection. In this case, we have chosen to carry out a lexical annotation in our corpus as we are working with sets of words (multi-word units) in order to understand the particular object of our research.

Following the definition and classification of our corpus, we will next present our aim and the methodology we employed in order to develop our investigation. The compilation of a bilingual corpus on cosmetics is intended to sustain a study of multi-word units (phraseology) present in the descriptive labels of cosmetic products that can be found in digital catalogues, both in Spanish and in English. We intend to observe the similarities between the more frequent collocations that appear in such descriptions in the two languages involved, which in turn requires an analysis of the grammatical categories (Parts of Speech, POS) that are used in such collocations. We want to establish whether or not the most reiterated words of each of the subcorpora belong to the same classes according to our POS analysis, and if these most frequent grammatical categories are the ones that shape the most often repeated collocations. In other words, the present study is based on the comparison of both languages within a very specific context of use and according to their grammatical properties, all in order to appraise the extent to which the collocations represented in the subsamples are similar.



#### 4.1. Corpus methodology

The first aspect we must reflect here is the structure of the corpus study, how we have planned it. We have divided it into two main parts that consist of two folders named as “SPANISH” and “ENGLISH”, which contain all the information labelled in each language, together with an Excel document (Corpus 0) where all this information is arranged and organised.

Our first task consisted in harvesting the data for our study from a single source, i.e. the internet, and more precisely, the same webpage, Amazon (a digital catalogue that contains products of all kinds), but from two different countries: UK and Spain (<https://www.amazon.es/> and <https://www.amazon.co.uk/>). As our corpus is bilingual, we made searches of cosmetic products from different subfields such as lipsticks, blushes... that we previously organised and whose classification can be seen in Figure 2. In the process of compiling the labels, it was clear to us that we wanted to obtain the same product in English and Spanish, so filtering was the method we followed. We wrote down the exact name of a product that we already had selected in one of the languages, for example, *Maybelline Barra de Labios Color Sensational Vivid Matte 05 Nude* in Spanish, and we sought for the same product but in the other language, so we obtained the same result: *Maybelline LIP CS Vivid Matte 05 Nude Flush* in English.

In order to compile the information thus gathered, we created a Word document where pasted all segments until we reached a total word count of 30.092. Next, we divided them according to their language, so that we created two individual folders respectively named “SPANISH” and “ENGLISH”. Here, we created the tags for each of the labels we had gathered, so we gave the name to each individual product and saved them in TXT format (UTF-8). This task implied the fact of classifying them and according to the genre to which they belong, their field (product type) or the language in which they were written, among others. All together, and by the end of the compilation process, we had 139 searched products, with 12469 words in Spanish and 17632 in English. By way of practical illustration, here is a sample tag that exemplifies the criterion we followed when tagging the products: ***001DCCLSLOAMNESP***

ELEMENTS INCLUDED	DESIGNATION
Number of the product	001
Genre	DC (Description in a Catalogue)
Field	C (Cosmetics)
Subfield	LS (Lipstick)
Brand	LO (L'Oréal)
Webpage	AM (Amazon)
User	NE (Non-Expert)
Language	SP (Spanish)

**Figure 1. Tagging**

All these data was collected in an Excel document (named “Corpus 0”), so that it could be clearly organized and differentiated.

Figure 2 below shows the different products we searched for the purpose of corpus construction with their matching abbreviations. It also contains the several brands we obtained and their corresponding abbreviations.

Subfield abbreviation		Brand abbreviation			
Subfield	Abbrev.	Brand	Abbrev.	Brand	Abbrev.
Lipstick	LS	L' Oréal	LO	TOO	TO
Mask	MK	Maybelline	MB	TheBalm	TB
Eyeliner	EY	Rimmel London	RI	Technic	TH
Blush	BL	Tarte Drench	TD	Bourjois	BJ
Foundation	FO	Revlon	RV	Urban Decay	UD
Highlight	HL	Ultra HD	UH	Channel	CH
Bronzer	BR	Astor	AS	Ellos`re real	ER
		Beauty Goddess	BG	Huda Beauty	HB
		Benefit	BF	Nyx	NX
		Instalash	IL	Mac	MC
		Max Factor	MF	Mac Jacobs	MJ
		Dior	DR	Lancome	LC
		Wet`n Wild	WW	Kat Von D	KV
		Clinique	CQ	Makeup Revolution	MR
		Avon	AV	Barry M	BM

**Figure 2. Abbreviations**



On the basis of the above information, we proceeded to the next step of our corpus compilation, i.e. generating the document Corpus 0.

The information thus gathered provided enough materials to start processing the data by means of several software applications in order to obtain tangible results and reach conclusions. It is to this that we now turn.

## 4.2. Software tools

As already stated, the present analysis mainly focuses on lexical aspects. For this reason, the TagAnt software was used in the first place, since it enabled us to search all the different grammatical categories (POS) that appeared in the subcorpora. The same procedure was implemented twice in the course of the study: first for the Spanish subcorpus and secondly for the English-language dataset. TagAnt is easy to operate: it only requires compiling all the descriptions in a TXT document (in a UTF-8 format) which is then uploaded on the program. The results were shown on the right side of the screen and were copied in another TXT document which was named “corpus tagged” and included the first results.

This software provides the tags with abbreviations, so that it becomes possible to look for such taggers and obtain their meaning. The next step implied the use of the second and last of the software: AntConc. This program allows the user to develop different searches; in our case, we first searched the most frequent POS that appeared in each of the subcorpora by entering the whole “corpus tagged” set that we obtained with TagAnt. We filtered the search through the option “*WordList*” > “*Tool Preferences*”, where we were allowed to enter a TXT document with all the tags and sent the order to the program so that it searches for specific ones in our “corpus tagged” set and obtained the following results:



POS DATA - English		
Rank	Category	Frequency
1	<b>nn</b> ( <i>noun, singular or mass</i> )	3827
2	<b>np</b> ( <i>proper noun, singular</i> )	3146
3	<b>jj</b> ( <i>adjective</i> )	1979
4	<b>in</b> ( <i>preposition/ subord. Conj.</i> )	1715
5	<b>dt</b> ( <i>determiner</i> )	1208
6	<b>to</b> ( <i>"to"</i> )	1023
7	<b>Sent</b> ( <i>end punctuation</i> )	823
8	<b>nns</b> ( <i>noun plural</i> )	741
9	<b>cc</b> ( <i>coordinating conjunction</i> )	650
10	<b>pp</b> ( <i>personal pronoun</i> )	601
11	<b>sym</b> ( <i>symbol</i> )	588
12	<b>vv</b> ( <i>verb, base form</i> )	558
13	<b>rb</b> ( <i>adverb</i> )	505
14	<b>cd</b> ( <i>cardinal number</i> )	441
15	<b>vvz</b> ( <i>verb, present 3dp. Sing.</i> )	368

Figure 3. POS English

POS DATA - Spanish		
Rank	Category	Frequency
1	<b>nc</b> ( <i>Common nouns</i> )	3448
2	<b>np</b> ( <i>Proper nouns</i> )	2369
3	<b>prep</b> ( <i>Preposition</i> )	1284
4	<b>adj</b> ( <i>Adjectives</i> )	1069
5	<b>art</b> ( <i>Articles</i> )	776
6	<b>card</b> ( <i>Cardinals</i> )	534
7	<b>vlfjn</b> ( <i>Lexical verb. Finite</i> )	518
8	<b>sym</b> ( <i>Symbols</i> )	491
9	<b>del</b> ( <i>Complex preposition</i> )	468
10	<b>cm</b> ( <i>Comma</i> )	432
11	<b>fs</b> ( <i>Full stop</i> )	410
12	<b>cc</b> ( <i>Coordinating conjunction</i> )	356
13	<b>vldj</b> ( <i>Lexical verb. Pastparticiple</i> )	252
14	<b>colon</b>	248
15	<b>pdel</b> ( <i>Portmanteau word formed by de and el</i> )	230

Figure 4. POS Spanish

Subsequently, we set out to search the 20 most repeated words (per subcorpus) in descending order of frequency, yet excluding all words that did not have semantic meaning and including all others like nouns or adjectives. We compiled these results in two additional tables where the information can be seen clearly laid out.

AntConc Wordlist frequency - Spanish		
Rank	Wordlist	Frequency
1	color	173
2	labios	86
3	piel	59
4	maquillaje	51
5	natural	48
6	acabado	46
7	mate	42
8	volumen	41
9	ojos	40
10	tono	39
11	textura	37
12	negro	36
13	efecto	34
14	suave	34
15	base	32
16	caja	32
17	contenido	32
18	colorete	31
19	cepillo	30
20	intenso	30

**Figure 5. AntConc wordlist frequency- Spanish**

In this figure we can appreciate that the noun “color” is the most frequent word used in the Spanish corpus, with a frequency of 173 times repeated. In this limited list, the word that has the lowest frequency value (30) is the adjective “intenso”.

The same results but for the English data were the following:

AntConc Wordlist frequency - English		
Rank	Wordlist	Frequency
1	colour	154
2	look	130
3	skin	108
4	matte	106
5	formula	102
6	black	96
7	lashes	94
8	finish	75
9	mascara	75
10	foundation	72
11	blush	71
12	lip	68
13	liquid	67
14	natural	63
15	lipstick	59
16	box	58
17	eye	58
18	brush	55
19	easy	55
20	liner	54

Figure 6. AntConc wordlist frequency- English

By looking at the table, we can see that in English the most frequent repeated word stands for the noun “*colour*”, which matches the Spanish results for its cognate “*color*”; although the frequency rate for the former (154 times) is slightly lower than for the Spanish term. As regards the least repeated word table slot, there is a mismatch between the two subcorpora, since in the English language wordlist this is occupied but the noun “*liner*” with a value of 54, while in the Spanish language table the box included a word which does not even belong to the same grammatical category.

After these preliminary results, we focused on what is actually the main purpose of this study, which is the scrutiny of collocations in the cosmetic labels subcorpora. To this end, we took each of the 20 words in the top-frequency wordlists one by one and studied their more frequent collocations by following the criterion “first word to the right” (1R) or “first word to the left” (1L). The aim was to identify the most frequent collocations

for the words most frequently repeated in each subcorpus. The results of our search were incorporated into two tables, each corresponding to the two languages represented in the corpus as shown in Figure 7:

Collocations formed by the most frequent words	
Spanish	English
de <b>color</b> ( <i>freq. 20</i> )	your <b>skin</b> ( <i>freq. 26</i> )
de <b>labios</b>	<b>look is</b>
la <b>piel</b>	of <b>colour</b>
de <b>maquillaje</b>	<b>matte finish</b>
acabado <b>natural</b>	the <b>formula</b>
un <b>acabado</b>	mascara <b>black</b>
acabado <b>mate</b>	your <b>lashes</b>
de <b>volumen/ volumen de</b>	lasting <b>finish</b>
de <b>ojos</b>	lash <b>mascara</b>
maquillaje <b>tono</b>	<b>foundation with</b>
una <b>textura</b>	powder <b>blush</b>
color <b>negro</b>	<b>lip colour</b>
un <b>efecto</b>	<b>liquid eyeliner</b>
textura <b>suave</b>	a <b>natural</b>
base de	matte <b>lipstick</b>
la <b>caja</b>	<b>box contains</b>
<b>contenido de</b>	<b>eye pencil</b>
<b>colorete en</b>	the <b>brush</b>
el <b>cepillo</b>	<b>easy to</b>
color <b>intenso</b>	super <b>liner</b>

Figure 7. Collocations made up of the most frequent words in both corpora.

Then, we created two lists in TXT format with all the most frequent collocations in both languages. By the time these steps were completed, we had already had stored the POS analysis and the frequency of words and their collocations, but we still needed to conduct the POS analysis of the collocations list that we had obtained as a result of our word-by-word software-driven search. In order to do that, we went back to TagAnt and successively entered the TXT files named “collocations in English” and “collocations in Spanish” so as to obtain the grammatical categories that entered each collocation. The results we obtained were as follows:

POS analysis of the most frequent collocations - Spanish		
Rank	Wordlist	Frequency
1	ART + CN	5
2	PREP + CN	3
3	NC + PREP	4
4	NC + ADJ	3
5	Vladj + ADJ	1
6	CN + CN	1
7	ART + Vladj	1
8	Pdel + NC	1

POS analysis of the most frequent collocations - English		
Rank	Wordlist	Frequency
1	NN + NN	4
2	DET + NN	3
3	ADJ + NN	6
4	NN + PREP	1
5	PREP + NN	1
6	PP\$ + NN	2
7	NN + VBZ	1
8	NN + VVZ	1
9	ADJ + "to"	1

**Figure 8. POS analysis of the most frequent collocations- Spanish/English**

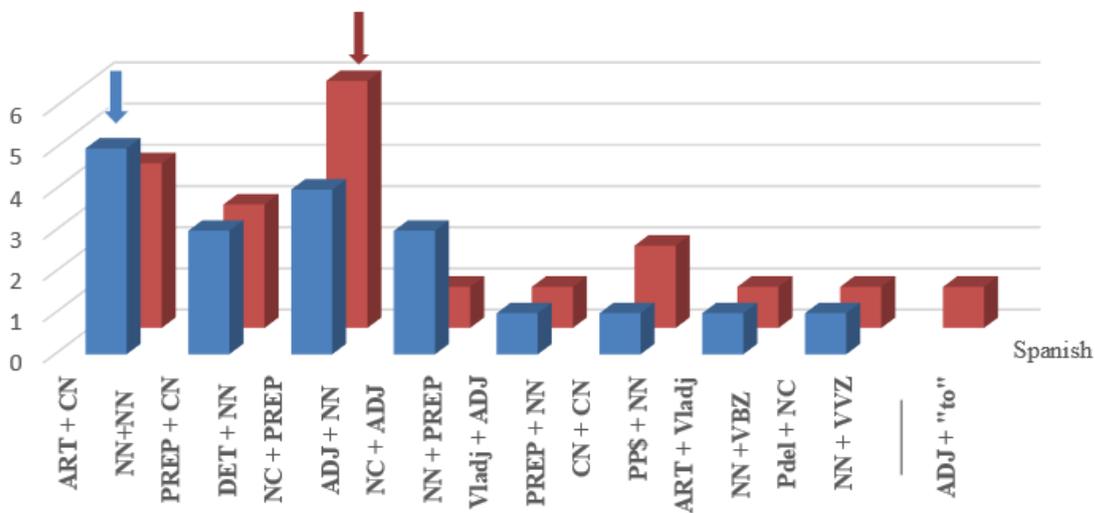
These tables show the most common POS patterns in the identified collations (Spanish and English respectively).

The above figures were helpful in contrasting data and reaching the final step of our analysis: the statement of results and the drawing of conclusions.

### 4.3. Corpus analysis results

From Figure 8 we can infer the following information: the most common collocation in Spanish is made up of Article + Common Noun (5 times), while in English the highest value corresponds to Adjective + Noun (6 times). Accordingly, we can state that there is NO MATCH between English and Spanish as the collocations they use when describing cosmetic products in the corpus under examination. Nevertheless, we cannot claim that they totally differ either, insofar as nouns are present in both sets of most common collocations. It is also worth highlighting that in English there is a higher occurrence of verbal tenses, while in Spanish articles and prepositions are more common. Perhaps the best way to visualise such similarities and contrast is by overlapping the graphs that project the content in Figure 8.

**POS analysis of the most frequent collocations comparison**



**Figure 9. POS analysis of the most frequent collocations comparison**

The blue bars correspond to the Spanish collocation patterns, while the red ones correspond to the English ones.

The above results make it possible for us to determine whether or not we have fulfilled our purpose. We set out to study the collocations that appeared in the descriptions of cosmetic catalogues in English and Spanish in order to gauge the extent of their



linguistic similarities. We also wanted to see if the most frequent words that were repeated in the corpora coincide with the collocations they could enter, in order to see if their frequency rate influenced the use of collocations. As primary finding, we were able to establish that the grammatical category “noun” is the most frequent one in both languages, and at the same time nouns from the two languages studied are cognates since they are genetically connected. When we looked downwards in the list of collocations, we realized that this coincidence did not happen anymore. This fact leads us to conclude that the collocations are not the same in both languages. Another important issue that contributes to this conclusion is the POS analysis because Spanish and English do not use the same grammatical categories to form the collocations. For example, we have already seen that in English there is an abundant use of verbal times, while in Spanish these are partially absent.

The above analysis allows us to conclude that even in such a highly patterned lexical register as that used in cosmetology—a language domain where, additionally, the influence of English is persuasive (Tejedor 319)—the descriptions of the same products in two different languages (English and Spanish) do not employ the same grammatical patterns, there being a number of outstanding differences that we have synthetically presented in the preceding lines. Otherwise, it shows the fundamental role the lexicon plays and how it can vary when the same content is expressed in two different languages, since we can have a same product in two different languages, which share the same function but the means of expressing it can differ significantly. In the next section of this graduation project, we will check if these assumptions are borne out in the translation of cosmetic labelling and if the detected grammatical patterns are more likely to be shared by two languages when translational operations is directly involved. Additionally, we shall test the reliability of machines in the conversion of cosmetic labels from one language into another on the initial hypothesis that the automatic device will not provide accurate results.

---

**4. Presentation of the data: Corpus on cosmetics.** This section is borrowed from the project named *Phraseological study in the field of cosmetics*, developed during academic year 2017-2018 as part of the course “Lingüística Aplicada II”, tutored by Doctor Belén López Arroyo, in the University of Valladolid’s Degree in English Studies.



## 5. Translation analysis of cosmetic labels

With the purpose of comparing and testing the performance of two distinct translation modalities applied to a selection from our bilingual corpus, we next assess translations produced by machines together with those yielded by a human translator (in this case, ourselves). It must be highlighted that in this section we will restrict our sample to one single cosmetic label, while the complete selection and its corresponding translations are included in the “Annexes” to this graduation project.

In order to conduct our trial, we selected randomly ten descriptive labels of one the hundred and thirty-nine that we originally compiled, five belonging to each of the two languages involved in this project (English and Spanish). Subsequently, we explored the several machine translation systems that are available as automatic translation tools, and we selected DeepL and Google Translator as our translation tools. In parallel, we produced our own translated version of the samples. This *modus operandi* eventually yielded bidirectional translations, since we had five samples translated from English into Spanish, while the remaining five underwent the inverse operation. Below we provide four consecutive tables, each consisting: one featuring the original label in its respective language, while the three left display the several translations resulting from the implementation of different procedures.

- **Original label (in English)**

*Rimmel London Scandaleyes Bold Eyeliner, Black, 2.5 ml*

- Intense glossy colour
- Up to 24 hours of wear
- Innovative ball tip applicator
- Waterproof formula
- Lightweight formula with intense colour

Want the London look? Who says a girl can't have it all? ! Finally, a liquid eyeliner that provides a bold look in one quick easy step, and is long-lasting yet easy to remove. The ball-tip applicator allows for a quick and comfortable application. The result? A high-impact look that lasts all day! When you want to stand out from the crowd with a bold and traffic stopping eye-look in a quick stroke! With up to 24 hours of wear, intense colour and a waterproof formula, Scandaleyes Bold Eyeliner is the perfect product for your no-fuss high-impact look! Simply apply our innovative ball tip application along with the upper lash line and follow the roots of your lashes to ensure a gap free line. Flick out at the end to achieve a striking, eye-catching look. It's hard to go wrong with Scandaleyes Bold Eyeliner. This unique product is ideal for that super quick and super bold eye look. The intense colour and the long-wearing formula are a must for the Rimmel girl. Scandaleyes Bold Eyeliner is an invaluable weapon against boring everyday eyeliner. Rimmel London is an inner confidence that how we look is always cool, irreverent, never predictable and always evolving. Rimmel is not perfection or intimidating, it is like London: young, urban, eclectic, edgy. It is expression with no rules, it is real, it is accessible. **GET THE LONDON LOOK!**

## - **Human Translation**

### *Delineador de ojos negro Rimmel London Scandaleyes – 2,5 ml*

- Color intenso y brillante
- Hasta 24h de duración
- Innovador aplicador con punta esférica.
- Fórmula resistente al agua.
- Fórmula ligera, pero con color vivo.

¿Quieres una mirada londinense? ¡¿Quién ha dicho que una chica no puede tenerlo todo?! Por fin existe un delineador de ojos líquido que proporciona una mirada atrevida en un sencillo y rápido paso, y que es duradero, pero a la vez fácil de desmaquillar. El aplicador con punta esférica permite una aplicación rápida y cómoda. ¿Que cuál es el resultado? ¡Una mirada de alto impacto que dura todo el día! ¡Para cuando quieras destacar entre la multitud con un look de ojos atrevido y que para hasta el tráfico con un rápido trazo! Con una duración de hasta 24h, un color intenso y una fórmula resistente al agua, el delineador de ojos Scandaleyes en tono negro para conseguir tu look de alto impacto y sin complicarte la vida. Lo único que tienes que hacer es deslizar el aplicador esférico por encima de la línea superior de las pestañas y asegurarte de que no queda ningún hueco sin rellenar. Continúa la línea hacia el exterior para conseguir una mirada llamativa y vistosa. Es difícil equivocarse con el delineador de ojos de Scandaleyes. Este exclusivo producto es ideal para realizar este look superrápido y atrevido en tus ojos. Un color intenso y una fórmula de larga duración son imprescindibles para una chica Rimmel. El delineador Scandaleyes es un arma inestimable en comparación con el aburrido delineador cotidiano. Rimmel London es una garantía para vernos siempre guapas, irreverentes, nunca predecibles y en constante evolución. Rimmel no es ni perfecto ni intimidante, sino que es como Londres: joven, urbano, eléctrico y atrevido. Es expresión sin reglas, es real y está a tu alcance. ¡CONSIGUE UNA MIRADA LONDINENSE!

## - DeepL translation

### *Rimmel London Scandaleyes Bold Eyeliner, Negro, 2.5 ml*

- Color brillante intenso
- Hasta 24 horas de uso
- Innovador aplicador de punta esférica
- Fórmula impermeable
- Fórmula ligera con color intenso

¿Quieres el look de Londres? ¿Quién dice que una chica no puede tenerlo todo? Por último, un delineador líquido que proporciona un aspecto atrevido en un solo paso rápido y sencillo, y es duradero y fácil de quitar. El aplicador de punta esférica permite una aplicación rápida y cómoda. ¿El resultado? Un look de alto impacto que dura todo el día! Cuando quieras destacarte de la multitud con un ojo audaz y un tráfico que se detiene, ¡míralo de un golpe rápido! Con hasta 24 horas de uso, color intenso y una fórmula impermeable, el delineador Scandaleyes Bold Eyeliner es el producto perfecto para tu look de alto impacto sin complicaciones. Simplemente aplique nuestra innovadora aplicación de punta de bola junto con la línea de las pestañas superiores y siga las raíces de sus pestañas para asegurar una línea sin espacios. Deslícese hacia afuera en el extremo para lograr una apariencia llamativa y llamativa. Es difícil equivocarse con el delineador Scandaleyes Bold Eyeliner. Este producto único es ideal para ese look súper rápido y súper atrevido. El color intenso y la fórmula de larga duración son imprescindibles para la chica Rimmel. El delineador Scandaleyes Bold Eyeliner es un arma invaluable contra el aburrido delineador diario. Rimmel London es una confianza interior de que la forma en que nos vemos es siempre fresca, irreverente, nunca predecible y siempre en evolución. Rimmel no es perfecto ni intimidante, es como Londres: joven, urbano, ecléctico, nervioso. Es una expresión sin reglas, es real, es accesible. **CONSIGUE EL LOOK DE LONDRES!**

## - Google Translator translation

### *Rimmel London Scandaleyes Bold Eyeliner, Negro, 2.5 ml*

- Color brillante intenso.
- Hasta 24 horas de uso.
- Innovador aplicador de punta de bola.
- Fórmula impermeable
- Fórmula ligera con color intenso.

¿Quieres el look londinense? ¿Quién dice que una chica no puede tenerlo todo? Finalmente, un delineador líquido que proporciona una apariencia audaz en un paso rápido y fácil, y es duradero y fácil de quitar. El aplicador de punta de bola permite una aplicación rápida y cómoda. ¿El resultado? ¡Un look de alto impacto que dura todo el día! ¡Cuando quieras destacar entre la multitud con una mirada atrevida y de tráfico que se detiene en un golpe rápido! Con hasta 24 horas de uso, colores intensos y una fórmula a prueba de agua, el delineador de ojos Bold de Scandaleyes es el producto perfecto para tu look de alto impacto sin complicaciones. Simplemente aplique nuestra innovadora aplicación de punta de bola junto con la línea superior de las pestañas y siga las raíces de sus pestañas para asegurar una línea sin huecos. Deslízate al final para lograr un aspecto llamativo y llamativo. Es difícil equivocarse con Scandaleyes Bold Eyeliner. Este producto único es ideal para ese look súper rápido y súper atrevido. El color intenso y la fórmula de larga duración son indispensables para la chica Rimmel. Scandaleyes Bold Eyeliner es un arma invaluable contra el delineador de ojos aburrido todos los días. Rimmel London es una confianza interna de que nuestra apariencia es siempre fresca, irreverente, nunca predecible y en constante evolución. Rimmel no es la perfección o la intimidación, es como Londres: joven, urbano, ecléctico, nervioso. Es expresión sin reglas, es real, es accesible. ¡OBTÉN EL ESTILO LONDINENSE!



In the theoretical part of this project, we already outlined a brief analysis of the automatic and human translation processes, as well as of the specific field we are addressing in this project, marketing and advertising. In the case of specialized translation conducted by humans, we emphasized the importance of having a sound previous background knowledge of the area of specialization the translated text belongs to so as to achieve reliable and accurate outputs. In our own (human) translation, we tried to apply this professional standard by conducting some background research on cosmetology. Additionally, we paid special attention to purely linguistic aspects, particularly typographic syntax, grammar and semantics.

On the other hand, and in relation to typography, English and Spanish do not share identical conventions. The example above illustrates such differences with regard to exclamatory and interrogatory marks: in Spanish they are used both in initial and final position, while in English they only appear as a sign of closure. In relation to grammar, the natural differences between the two languages involve issues like word-order, the pro-drop status of Spanish, a differing verbal paradigm, the possessive meaning of Spanish articles in certain collocations, etc. The analytical study of our corpus in the previous section of this thesis helped us to discern the most frequent grammatical patterns underlying the collocations we identified in our descriptions of cosmetic products. More specifically, we tracked a predominant sequence for Spanish (article plus noun) which did not match the most frequent one in the English sample (adjective plus noun). Arguably, this observation ultimately relates to natural differences between the two languages as regards the placement of descriptive adjectives, which typically postmodify nouns in Spanish by contrast with the clearly prevailing premodifying pattern that characterizes English grammar.

While producing our translation, we bore in mind grammatical patterns that emerged as most prevalent following the study of the corpora and tried to implement them in our rendering. In fact, we used the predominant order of article plus noun when translating from English to Spanish, as well as the common pattern of adjective plus noun when translating from Spanish into English, so as to produce descriptions that were not simply accurate in terms of lexical choices, but grammatically idiomatic.

In the third place, the semantic component embedded in the descriptive labels that constitute the target of our inquiry, equally shaped our translation strategy. The appellative function projected in the marketing and advertising messages demanded what



Peter Newmark called a “semantic translation”, somehow similar to that required in the rendering of aesthetically- marked texts: “Original expression (where the specific language of the speaker or writer is as important as the content), whether it is philosophical, religious, scientific, legal, technical or literary, needs to be translated semantically” (11). The text’s appellative function, in our view, is fully compatible with their partly technical nature and stems from the simple fact (which we explained in our theoretical framework) that the main purpose of marketing and copywriting, and of the translation work that both may entail, is the achievement of the highest number of sales. This introduces an element of style into the picture with demands both a close attention to formal details and a limited amount of creativity. An example of such requirements can be seen in the item that we selected for the purpose of this section, reproduced and translated by us in the first of the above tables. There, the two-word English question “the result?” was translated with slightly more idiomatic overtones as “¿que cuál es el resultado?”, thus prioritising a more conversational tone over an otherwise totally legitimate and understandable, yet possibly weaker translation in terms of marketing impact.

Let us now turn to the discussion of the translation supplied by the MT systems. Using these automatic tools simply required copying and pasting the original texts onto the system’s screens without any kind of modification and storing the results in a Word document. The procedure followed consisted in analysing the same aspects as we focused on while producing our own translation: that is, typographical, grammatical and semantic. Regarding typography, the two automatic devices provided both correct and wrong results, although we observed a better performance with Google translator. If we stick to our single selection from the larger sample, the latter application successfully the initial and final exclamation marks through the target text, whereas DeepL failed in translating the initial sign, as for example in the catchphrase **CONSIGUE EL LOOK DE LONDRES!**

In relation to grammatical structures, both systems sustained correct patterns in most phrases, although they perhaps failed in enriching the text’s cohesion with linking words or slightly deviated from what we perceive as a more canonical and idiomatic handling of the descriptive adjective/noun combination. The segment “intense glossy colour”, for example, was rendered as “color brillante inenso” by Google Translator, perhaps betraying some stiff automatism in the translation process.

The semantic and terminological component was perhaps the least accomplished one, insofar as the above-mentioned appellative impact was considerably weakened. The systems' word-by-word translations, particularly in the case of Google Translator, occasionally produced either incoherent or flat results that were either de-contextualized or not particularly eye-grabbing not in tune with the technical and marketing idiom employed in the sector in Spanish-language channels. Thus the rendering of *ball tip* as *punta de bola* or *waterproof* as simply as *impermeable* contrasts with what we modestly believe is a more fluent and natural human translation. The systematic non-translation of the English word "look" by the machines, on the other hand, is possibly a legitimate choice (although not advised by sources like the *Diccionario panhispánico de dudas* or FUNDEU) since the Anglicism is quite consolidated in present-day Spanish and today's advertising favours the use of foreign (particularly English and French) words for their urban, modern or sophisticated connotations; yet it is worth highlighting that the kind of stylistic variation that human translation is likely to engage in (*look, mirada, estilo*) can both be enriching and introduce shades of meaning and even wordplay (e.g. the association of those synonyms with the female gaze or the female eyes in this case) capable of enhancing the text's marketing appeal. Finally, we also detected plain errors like the repetition of the word *llamativo* in *un aspecto llamativo y llamativo* (in the source text *super quick and super bold eye look*), due to partly similar meanings of two words collected in the dictionary-based component of the system, which in this case failed to discern the context of use.

Ultimately, we can conclude that the machine cannot equal the role of the human translator, since there are several aspects involved in translation that go far beyond a mere one-to-one equivalence in terms of lexis, grammar, meaning and even typographic conventions. Even so, we were able to detect differences between the two devices employed. Generally speaking, we obtained better results with DeepL than with Google Translator. Finally, it is worth highlighting that machine translation tools such as the ones used for the purpose of our trial can play an important role in aiding human translation and saving a considerable amount of time, even though, despite their impressive current development, they are not as yet sufficient and fully reliable when dealing with professional translation tasks.



## 6. Conclusion

Machine translation systems have experienced a phenomenal growth in the current century. So much so that the end of human translation as we have known it is now being increasingly announced. The old human vs machine translation debate is forced to update its arguments at a growing pace and forecasts are voiced regarding the impending prospect of machine systems that will have full capacity and autonomy to perform translation work with no human intervention whatsoever. While the exact future of human translation in the age of artificial intelligence and big data cannot be determined right now, it is reasonable to assume, as we did in the introduction to this graduation project, that absolute parity between humans and machines as regards this activity will not be attained any time soon, particularly where elements like culture, complex linguistic modality, subtle emotions, ambiguity or humour, to mention but a few, play a central part (needless to say, in matters related to time and costs savings, machines will always have the upper hand). In this sense advertising translation and literary translation share some common ground, and that is why both benefit from the unquestionably unique input of individual human translators.

Likewise, the degree of specialization necessary to deal with professional translation assignments like the one illustrated by our sample require a thorough acquaintance with the specific technical or business sector involved—in our case, advertising and marketing translation in the field of cosmetics. Leaving aside the issue of specialized terminology (which can be managed by comprehensive datasets embedded into the dictionary component of MT systems) and the natural grammatical differences between the two languages involved that could be eventually sorted out by a sufficiently streamlined rule-based machine translation—even though we still see a fair amount of grammatical errors in machine translated texts—the more subjective component, the impact of emotions in consumers' decisions as modulated by visual and especially verbal means often involving ambiguity and connotation, is perhaps less susceptible to a fully automated process.

Marketing and advertising is a huge world, and so are its consequences translation-wise. For this reason, we have focused on cosmetics descriptive labels in both English and Spanish and moreover resorted to a limited corpus of utterances (our own compilation) and their underlying grammatical patterns so as to initially discern recurring



structural differences between Spanish and English as reflected in this particular sample. We then attempted to carry over these observations into the translation exercise (and the ensuing comparison with the MT output) so as to support our assumption that the human translator remains so far irreplaceable. Proving such a hypothesis naturally led us to employ machine translation systems that we used to translate the cosmetic labels/descriptions in our selection.

We modestly believe that in such a specialized and at the same time semantically complex domain as the one represented by our compilation, fully automated translation software does not perform to the highest standard: one which, while enormously more costly and time-consuming, can be accomplished by human translation. In our simple experiment, human translation proved more sensitive and adaptive to 1) the natural order of the TT collocations, 2) the extended context of the descriptive labels being translated and 3) the appellative/persuasive nuances of the marketing and advertising idiom; by contrast, the MT output, remarkably correct in a number of instances, was nevertheless unsubtle when it came down to interpreting the labels' contextual frame, the several contexts of use for some of the lexical items involved or the shades of meaning of several descriptive terms—all of which produced a number of dubious renderings.

Ultimately, we can assume that our initial hypothesis has been proved and that the input of human translators is not, at least as yet, replaceable, even though the benefits of MT systems (particularly in partnership and under the supervision of competent human translators) are undeniable: otherwise entire industry sectors and the ICT big companies would not be making such large investments in upgrading the kind of neural machine translation systems that we employed in our trial and which in any case are far more powerful than the old phrasal-based statistical technologies. The importance of advertising and marketing in the purchasing and sales sector and the international reputation of such large business as those in the booming sector of cosmetics cannot be left in the exclusive hands of MT systems that have not yet come to full fruition. High quality results will most likely continue to depend on the contribution of human translation, either on an exclusive basis or in collaboration with the work of ever more powerful automated systems.



## 7. Bibliography

- Alarcón Navío, Esperanza. “Traducción automática versus traducción humana: tipología de errores”. Ed. AIETI. Granada 12-14 Feb. 2003.
- Alonso del Caño, Carmen. “Phraseological study in the field of cosmetics”. Universidad de Valladolid, 2018. Doctor Belén López Arroyo. 2018. Print.
- Boletín Oficial del Estado. Ley General de Publicidad. En BOE núm.274, de 15 Nov. 1988. [En línea]. <https://www.boe.es/buscar/doc.php?id=BOE-A-1988-26156>. 15 Mayo 2019.
- Bueno García, Antonio. *Publicidad y traducción*, Núm. 2. Zaragoza. Diputación Provincial de Soria, 2000. Print.
- Collins Dictionary. “Machine Translation”. Collinsdictionary.com. Web. <https://www.collinsdictionary.com/es/diccionario/ingles/machine-translation>. Accessed 20 May 2019.
- De Pedro Ricoy, Raquel. “Las estrategias de internacionalización en la traducción publicitaria”. *Revista lingüística y lenguas aplicadas*. Edimburgo, 2007. 7-16- Print.
- González, Juan Antonio. *Teoría General de la Publicidad*. México D.F. and Madrid: Fondo de Cultura Económica, 1996. Print.
- Hina, Horst. “Hacia una teoría de la traducción”. ES, 1. *Publicaciones del Departamento de Inglés de la Universidad de Valladolid*. 169-194.
- Hina, Horst. “Presente y futuro de la traducción”. *Revista de Filología Inglesa*. Ed. Universidad de Valladolid. Valladolid, 2016. 29-76. *Dialnet*. Web. 13 May 2019.
- Hutchins, John. “Machine Translation: a brief history”. *Concise history of the language sciences: from the Sumerians to the cognitivists*. Ed. E.F.K. Koernes and R.E. Asher. Oxford: Pergamon Press, 1995. 431-445. Print.
- Kin Quah, Chiew. *Translation and technology*. 1<sup>st</sup> Ed. New York: Palgrave Macmillan, 2006. Print.
- Leiva Rojo, Jorge. *Aspects of Human Translation: the current situation and an emerging trend*. *Hermeners. Revista de traducción e interpretación*, 20 (2018): 257-294.



- Newmark, Peter. *About translation*. Vol. 74. Multilingual matters, 1991. p. 11.
- Pigott, Ian. *Systran Development at the EC Commission*. 1976 to 1992. Luxemburgo: European Commission Translation Service (SDT), 2001. Print.
- Rabadán, Rosa. "Traducción, intertextualidad, manipulación". *Estudios sobre la traducción*. Publicaciones de la Universidad Jaume I. Castellón, 1994.
- Robertshaw, Emily. *Packing translation- 5 elements you need to consider*. *Andiamo! The language professionals*. 10 April 2017. Web. 21 June 2019.
- Somers, Harold. *Machine Translation*. Netherlands: Kluwer Academic Publisher, 1999. Print.
- Valdés, María Cristina. *La traducción publicitaria: comunicación y cultura*. Valencia: Publicaciones de la Universidad de Valencia, 2004. Print.
- Wills, Wolfram. *Knowledge and Skills in Translator Behaviour*. Germany: John Benjamins Publishing Company, 1996. Print.

## CORPUS BIBLIOGRAPHY

- Anthony, Laurence. "Laurence Anthony's Antconc". Laurenceanthony.Net, 2018, <http://www.laurenceanthony.net/software/antconc/>. Accessed 28 May 2018.
- Anthony, Laurence. "Laurence Anthony's Tagant". Laurenceanthony.Net, 2018, <http://www.laurenceanthony.net/software/tagant/>. Accessed 28 May 2018.
- Anthony, Laurence. "Antconc Help". Laurenceanthony.Net, 2018, <http://www.laurenceanthony.net/software/antconc/releases/AntConc357/help.pdf>. Accessed 28 May 2018.
- Anthony, Laurence. "Tagant Help". Laurenceanthony.Net, 2018, <http://www.laurenceanthony.net/software/tagant/releases/TagAnt120/help.pdf>. Accessed 28 May 2018.



- "English - Tree Tagger Tags". Courses.Washington.Edu, 2018, <https://courses.washington.edu/hypertext/csar-v02/penntable.html>. Accessed 28 May 2018.
- "Spanish - Tree Tagger Tags". Cis.Uni-Muenchen.De, 2018, <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/spanish-tagset.txt>. Accessed 28 May 2018.
- "Amazon.Es: Online Catalogue". Amazon.Es, 2018, <https://www.amazon.es/>. Accessed 22 May 2018.
- "Amazon.Co.Uk: Online Catalogue". Amazon.Co.Uk, 2018, <https://www.amazon.co.uk/>. Accessed 23 May 2018.