

CORPUS-BASED CONTRASTIVE ANALYSIS OF ENGLISH ADJECTIVES IN SPANISH TRANSLATIONS: A TYPOLOGICAL ISSUE

Noelia Ramón

University of León, Spain

1. INTRODUCTION

Adjective position is one of the typological features where English and Spanish differ more clearly, since the unmarked position of adjectives is the pre-modifying one in English and in Spanish the post-modifying one. This dissimilarity is related to the different origin of these two languages, English being a Germanic language and Spanish a Romance language. However, the pre-modifying position of descriptive adjectives is possible in Spanish too, with a number of semantic connotations, involving mainly affectivity. Spanish translations of English texts may therefore be expected to present a higher rate of pre-modifying adjectives than texts produced originally in Spanish due to the influence of the source language English. Previous studies have shown that this is one important divergence between naturally occurring Spanish and Spanish translations from English (Rabadán et al. 2009). Adjective position can therefore be considered one of the main problems in translations between these two languages, and consequently a key point to take into account in translator training as well as in translation quality assessment.

In this paper I will provide a detailed analysis of some of the most frequent adjectives in English and Spanish, with reference in particular to their position in the noun phrase. The working hypothesis is that there will be a strong trend to locate adjectives in Spanish translations in pre-modifying positions more often than if these same adjectives occurred in original Spanish texts. The empirical material used for this analysis will be extracted from the English-Spanish parallel corpus known as P-ACTRES compiled at the University of León, Spain. This corpus contains 2.5 million words of contemporary English texts and their corresponding Spanish translations. A variety of registers are represented: fiction, non-fiction, press, etc. The CREA (*Corpus de Referencia del Español Actual*), a large reference corpus of Spanish, will be used for the comparison with original Spanish texts.

The aim of the paper is to illustrate how quantitative factors such as the overuse of the pre-modifying positions of adjectives, among others (Ramón 2009, Ramón & Labrador 2009), may be used for rating the quality of Spanish translations and should therefore be considered major aspects in translator training and in translation quality assessment.

2. THEORETICAL BACKGROUND

The term ‘typology’ in linguistics refers to the study of linguistic patterns that are found cross-linguistically, in particular, patterns that can be discovered solely by cross-linguistic comparison. In Croft’s words, “cross-linguistic comparison places the explanation of intralinguistic phenomena in a new and different perspective.” (Croft 1990: 4). Word order has always been one of the major issues in typological studies, although it has rather been focused on the order of constituents in the sentence. As Greenberg (1963) demonstrated, there are correlations between various features of ordering. Here we are particularly interested in universal 17, which states that languages with a dominant VSO word order most often present the adjective after the noun.

Typological differences among Germanic and Romance languages are many and varied. However, the differences relating to adjective position in these two language types are particularly important and have clear implications in the translation process between English and Spanish, for example. To begin with, adjectives are invariable lexemes in English, whereas in Spanish all adjectives vary morphologically to express number, and many of them also to express gender.

As for adjective position, the unmarked position for attributive adjectives is the pre-modifying one in English and the post-modifying one in Spanish. This radical difference is one important source of error in foreign language learning as well as in the translation process involving these two languages. Typologically, this difference is clearly related to the relatively strict word order we find in English as the result of a limited inflectional morphology. In contrast, a very rich morphology in all content words in Spanish allows for a rather freer word order, including the possibility to position descriptive adjectives in premodifying positions too. Classifying adjectives do not have this possibility: *un hombre pobre* and *un pobre hombre* are both possible, even

though with different meanings; but *un oso polar* is possible, whereas *un *un polar oso* is not. Descriptive adjectives may occupy both positions in the noun phrase (NP), though generally with different meanings or connotations, mainly involving affectivity in the premodifying position and a neutral meaning in postmodifying position. Some very common adjectives even have morphologically distinct forms for the premodifying position, such as *gran* for *grande*, or *buen* for *bueno*.

No clear rules exist in Spanish with respect to the use of the premodifying positions. Phonemics seems to play a role, since it is mainly short adjectives that can premodify longer nouns. However, each particular adjective presents its own pattern with respect to dominant pre- or post-modifying positions. I will investigate in this paper the patterns of several of the most common Spanish adjectives to shed some light on this issue.

3. METHODOLOGY

The empirical data for this study have been extracted from two electronic corpora: CREA and P-ACTRES. CREA (*Corpus de Referencia del Español Actual*) is a large monolingual reference corpus of the Spanish language. It contains slightly over 154 million words of texts in all registers, oral and written, and in all geographic varieties of contemporary Spanish, from 1975 until our days. From this large reference corpus we have selected a subcorpus of texts from the year 2000 on, published in Spain only (peninsular variety), and from the written section only. All in all there are 18,500,104 million words. P-ACTRES is an English-Spanish parallel corpus compiled at the University of León, Spain, and which contains over 2 million words of written English in various registers with the corresponding translations into the European variety of Spanish. All original texts were produced in the year 2000 or later. Table 1 below shows the number of words in each subcorpus according to register and language.

	ENGLISH	SPANISH	TOTAL
Books – fiction	396,462	421,065	817,527
Books – non-fiction	494,358	553,067	1,047,425
Newspapers	115,502	137,202	252,704
Magazines	119,604	126,989	246,593

Miscellanea	40,178	49,026	89,204
TOTAL	1,166,104	1,287,349	2,453,453

Table 1: Register distribution of P-ACTRES.

For this paper I will use the whole of the Spanish translation corpus, so all in all 1,287,349 words of translated Spanish.

The analysis will consist of two stages:

1. First, I will carry out a quantitative study on some of the most frequent adjectives in Spanish to determine whether there is a clear underuse or overuse of any of these adjectives in the translations when compared to the use in original texts.
2. Second, I will analyze a representative number of cases of some of these adjectives in both corpora to reveal the trend of usage with respect to the position of these adjectives in the NP.

The starting point for our analysis was the frequency list of the CREA corpus for the Spanish language, which was used to identify the most frequent adjectives. The list of the 25 most frequent adjectives runs as follows: *gran* (97), *general*, *mayor*, *nacional*, *mejor*, *nuevo*, *pasado*, *nueva*, *social*, *grandes*, *posible*, *importante*, *final*, *unidos*, *cierto*, *largo*, *claro*, *español*, *buena*, *internacional*, *igual*, *española*, *interior*, *buen*, *especial*. The 25 most commonly employed adjectives in English, according to the Cobuild Wordbanks Online are: *new*, *good*, *old*, *long*, *little*, *great*, *high*, *best*, *big*, *national*, *small*, *full*, *young*, *free*, *public*, *important*, *white*, *local*, *black*, *able*, *early*, *political*, *real*, *hard*, *available*. We can see that many of them are semantic equivalents, some even cognates.

As for the Spanish list, some of the adjectives included are the reduced forms of other adjectives for the premodifying position (*gran*, *buen*), so they are excluded from our analysis, as they may never occupy the post-modifying position. Other adjectives are of the classifying type (*nacional*, *social*, *internacional*), and must always appear after the noun they modify, so they are not useful in this study either. I have selected 5 adjectives from this list – *grande*, *bueno*, *nuevo*, *importante*, *largo* - to determine the following issues in comparison in original and in translated Spanish: a) overall frequency of use in original texts and in translations, b) statistical significance of the

differences identified using chi-square test. After that, I have extracted concordances from the 3 most common adjectives to analyze their particular patterns with respect to position in the NP.

4. RESULTS AND DISCUSSION

4.1. Frequency analysis

The frequency analysis of the most common adjectives in Spanish revealed interesting differences between original and translated texts. All the morphological forms of 5 Spanish adjectives were analyzed and it was found that in all cases there were statistically significant divergences in at least one of the forms. Table 2 below shows the number of cases in the two corpora used, the p-value of statistical significance (I consider values under 0.01 as significant), and whether the divergence is an overuse or an underuse in the translations.

	CASES CREA	CASES P-ACTRES	p-value	TYPE OF DIVERGENCE
GRAN	11.697	1.047	0	Overuse
GRANDE	1.674	188	0	Overuse
GRANDES	6.135	477	0.019	Not significant
MAYOR	10.941	712	0.08	Not significant
MAYORES	2.508	162	0.35	Not significant
BUEN	3.819	198	0.00005	Overuse
BUENO	1.701	264	0	Overuse
BUENA	4.058	274	0.62	Not significant
BUENOS	1.117	106	0.002	Overuse
BUENAS	962	108	$7e-8$	Overuse
MEJOR	9.141	650	0.59	Not significant
MEJORES	1.939	142	0.55	Not significant
NUEVO	8.863	360	0	Underuse
NUEVA	6.037	402	0.39	Not significant
NUEVOS	3.683	177	0.000001	Underuse

NUEVAS	3.602	186	0.00006	Underuse
IMPORTANTE	6.272	368	0.001	Underuse
IMPORTANTES	2.550	169	0.53	Not significant
LARGO	5.194	321	0.03	Not significant
LARGA	1.436	126	0.012	Not significant
LARGOS	506	48	0.03	Not significant
LARGAS	472	49	0.007	Overuse

Table 2: Number of cases and p-value.

In the case of the adjective *grande* and all its variants, Figure 1 below shows the number of cases per million words in original and translated texts. It can be seen that the short form for the pre-modifying position and the singular form of the post-modifying position are both overused in translations.

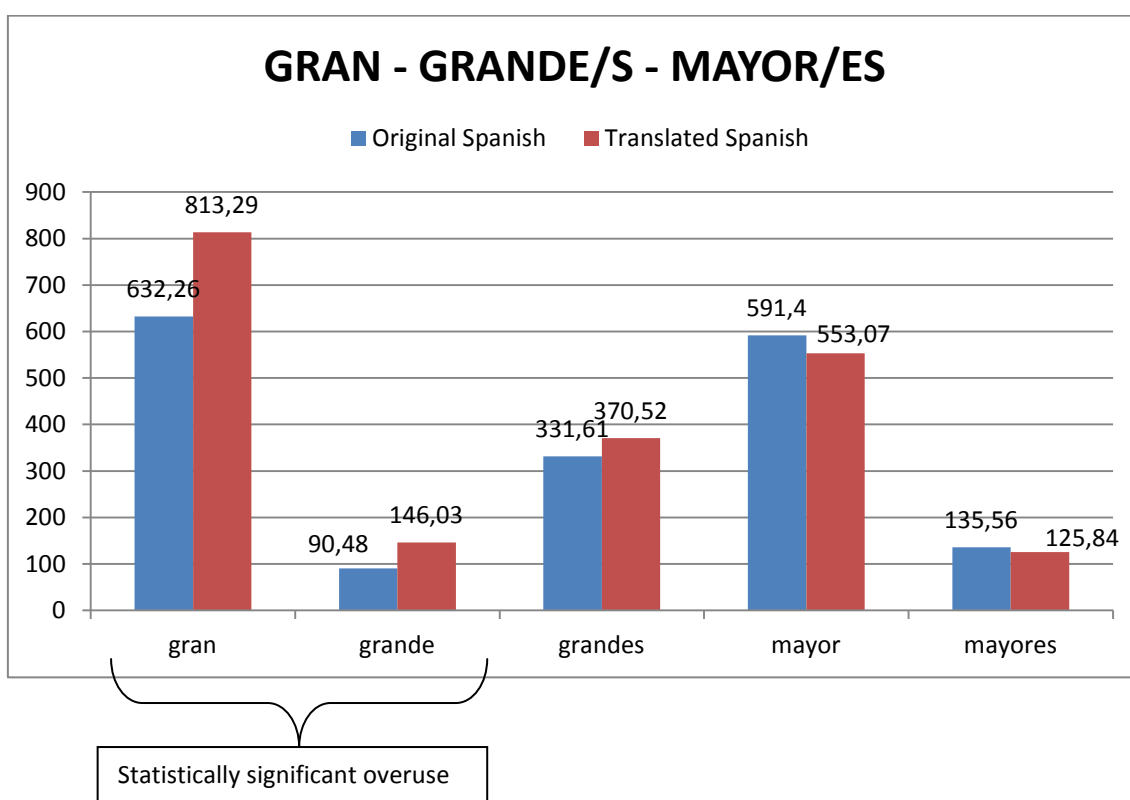


Figure 1. Number of cases per million words in original and translated Spanish.

The general high frequency of this adjective, which is in fact the most frequent one in Spanish, makes us think that the strong trend to overuse it in translations is a feature of the lexical normalization of translated texts in general. This phenomenon is

known as the ‘**simplification hypothesis**’ (Baker 1993, Toury 1995, Laviosa 1996) and suggests that translations tend to boost the use of typical features of the target language.

A similar point can be said of the adjective *bueno* and all its variants. Figure 2 below shows the number of cases per million words of this adjective in original and translated texts.

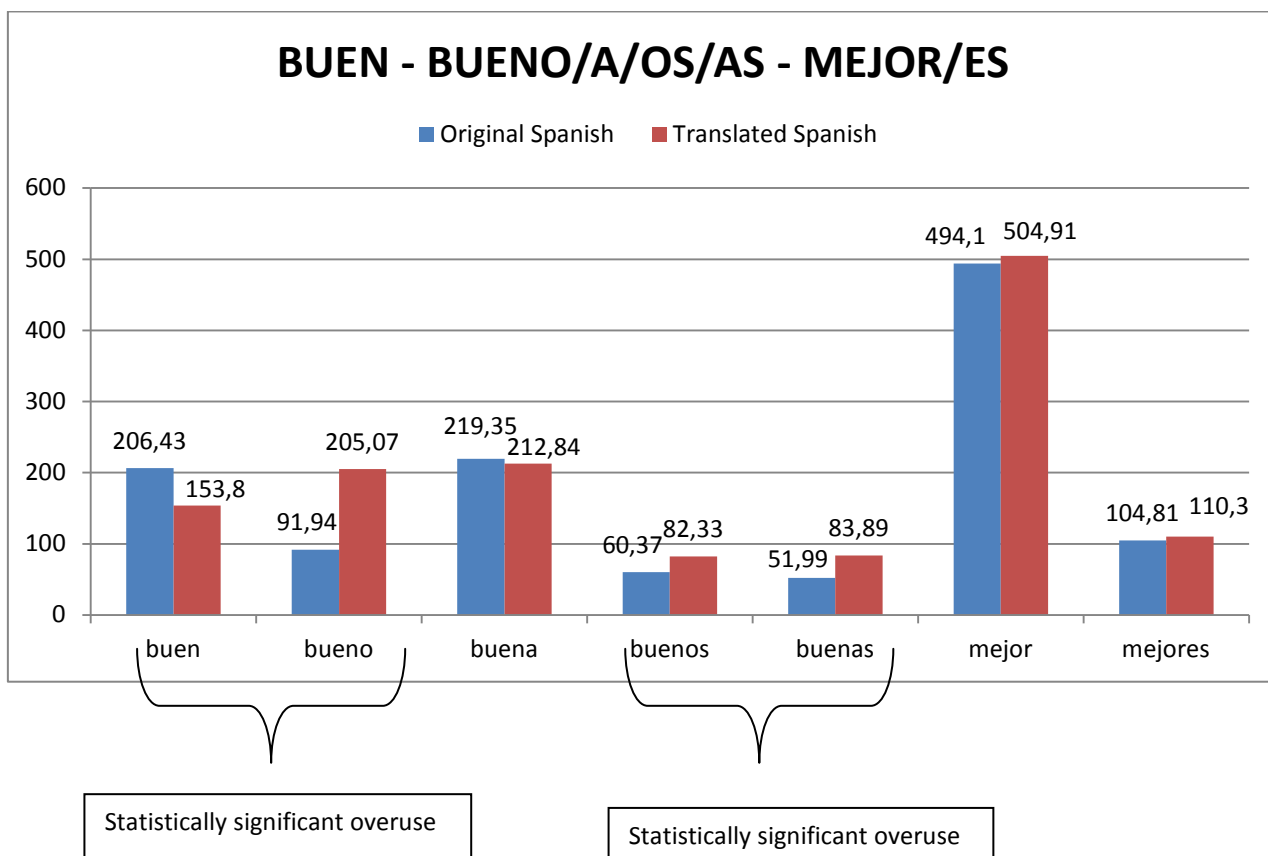


Figure 2. Number of cases per million words in original and translated Spanish.

The short form, the masculine singular and both masculine and feminine plural forms are all overused significantly in Spanish translations when compared with texts written originally in Spanish. Again this indicates a trend towards the normalization or simplification of translated language in general, i.e., high-frequency grammatical/syntactic resources tend to be preferred as translation solutions at the expense of other TL possibilities

In contrast to these two cases of overuse, the adjective *nuevo* and several of its morphological variants appear to be underused in Spanish translations. Figure 3 below shows the number of cases of each form per million words in each of the two corpora.

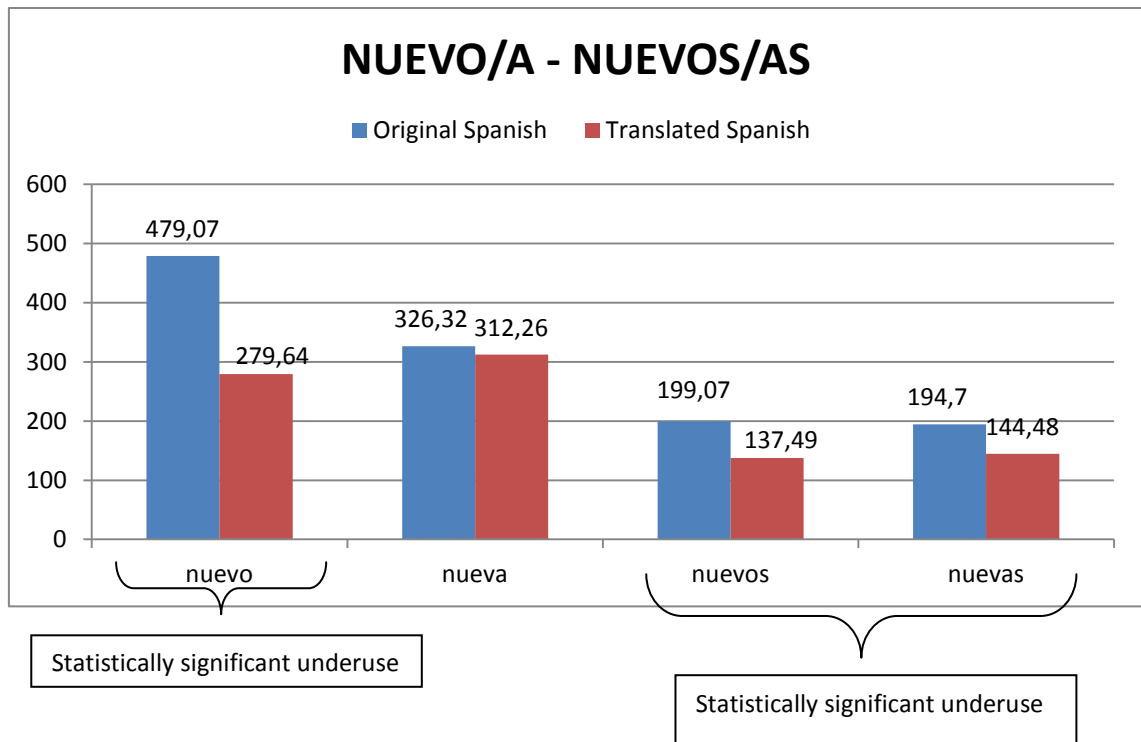


Figure 3. Number of cases per million words in original and translated Spanish.

It can be seen that the masculine singular form as well as both plural forms are significantly underused in translations. The adjective *nuevo* is very frequent in Spanish and has a very general meaning applicable to virtually any noun, so the underuse in translations may be due to the fact that other more precise and less general adjectives are used instead by professional translators to avoid homogeneity, repetition or too much simplicity. The adjective *new* is the most frequent one in English, so the underuse of the formal and functional equivalent in Spanish may be attributed to an attempt to achieve lexical variation by using other adjectives with similar meanings.

As for the adjective *importante*, only two forms have been considered, since this adjective varies only in number, but not in gender. The singular form appears to be underused in translations when compared to original Spanish texts and this underuse is statistically significant. The formal equivalent in English, *important*, is also one of the most frequent adjectives in that language, so the underuse in the Spanish translations of this source language can only be attributed, as in the case of the adjective *nuevo* described above, to an attempt on the part of translators to avoid repetition and simplification by using other less general adjectives. Figure 4 below shows the number of cases of this adjective per million words in original Spanish texts and in translations.

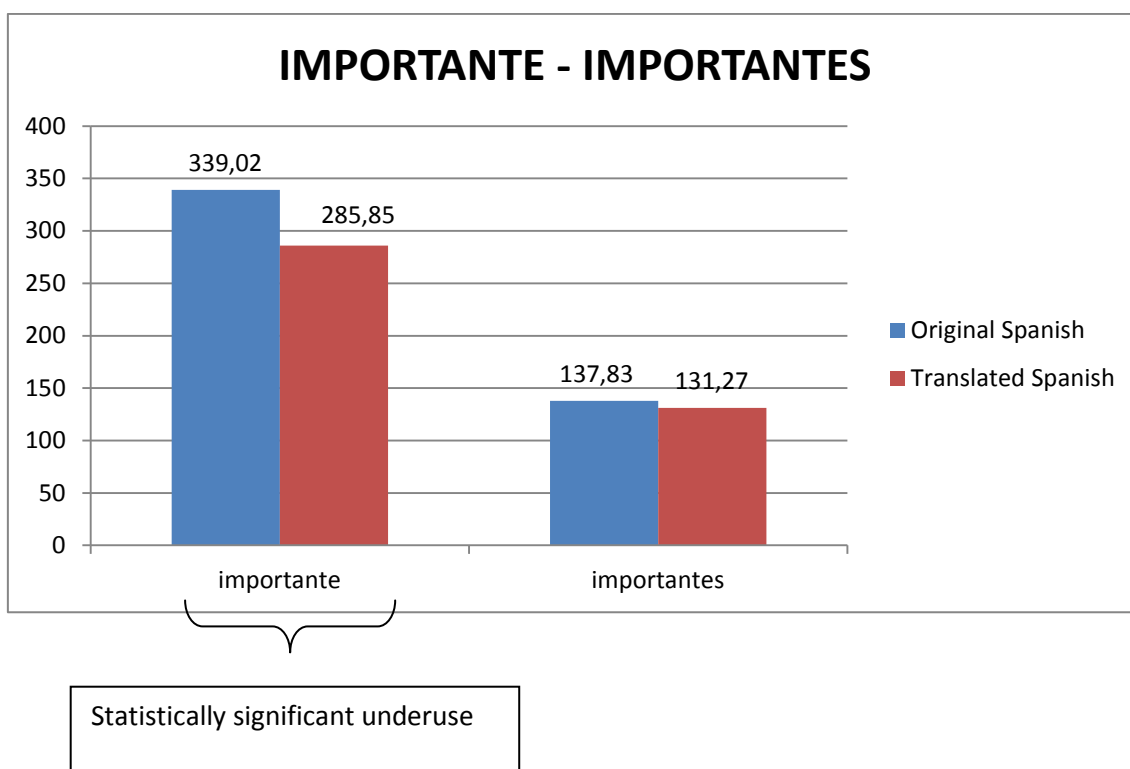


Figure 4. Number of cases per million words.

Finally, I checked the figures for the adjective *largo* and all its morphological variants in Spanish. The English adjective *long*, its functional equivalent, is also one of the most frequent ones in this language. Figure 5 below shows the number of cases per million words of all forms, masculine, feminine, singular and plural, as well as the only case where the differences between original and translated language are statistically significant.

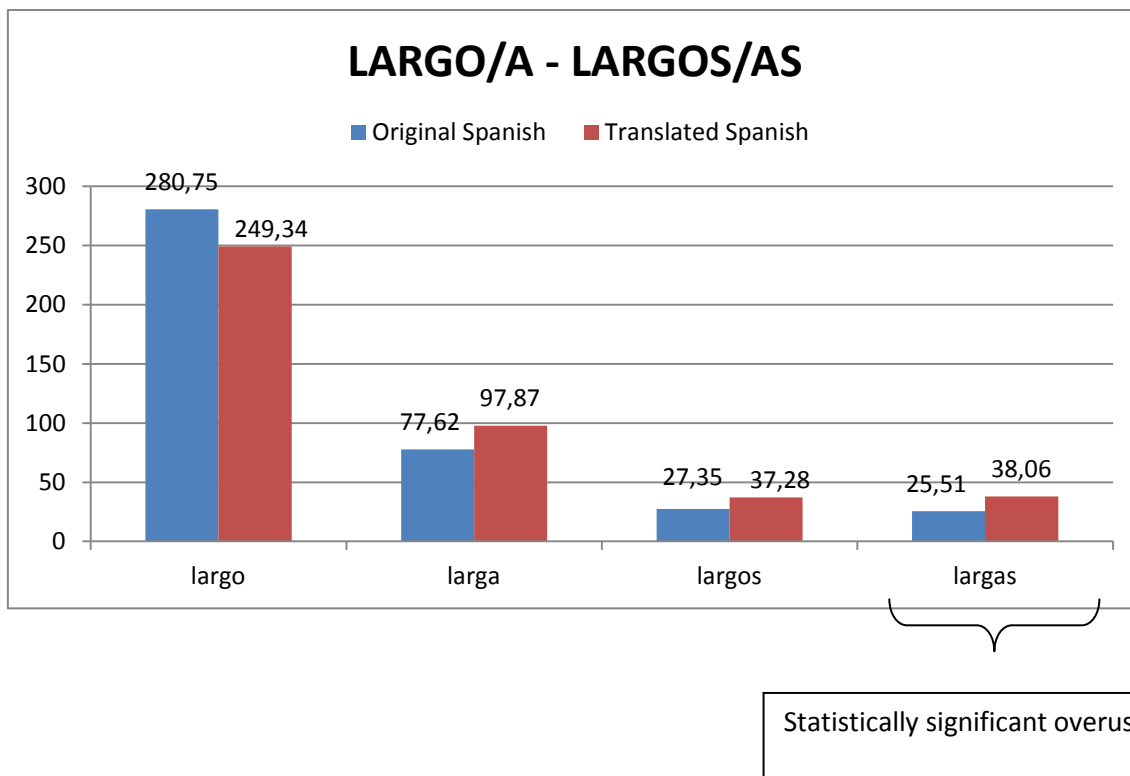


Figure 5. Number of cases per million words.

The only form with a statistically significant overuse is the feminine plural form *largas*, but it is so infrequent in both corpora that this data is not highly relevant. In consequence, this adjective does not seem to be one that is very prone to cause problems in translations between English and Spanish, as it is used with a very similar frequency in most of its forms.

4.2. Pre- and post-modification in translations

The next step in this analysis consist in looking at sufficiently representative numbers of concordance lines of different adjectives to compare the figures of pre- and post-modification in originals and in translations. The adjectives *grande* and *bueno* cannot be used for this purpose, as they have morphological forms specifically designed for the pre-modifying position. I will therefore choose examples of the following adjectives: *nuevo*, *importante* and *largo*, in their masculine singular forms, which are always the most numerous ones.

To be able to study a sufficiently representative number of cases of each adjective in original and translated Spanish, I will use a statistical formula to determine the number required:

$$n = \frac{N}{(N-1) E^2 + 1}$$

The element n is the final sample we will analyze, N is the whole sample of occurrences, and E is the estimated error, in this case 0.05 for a 95% confidence margin. So, from the total of 8,863 cases of the adjective *bueno* in CREA, I will need to analyze 383 cases to obtain representative results. Table 3 below shows the number of instances analyzed of each of the adjectives selected in both corpora, CREA for original Spanish and P-ACTRES for translated Spanish.

	CREA		P-ACTRES	
	Total	Selected	Total	selected
nuevo	8,863	383	360	190
Importante	6,272	376	368	192
largo	5,194	371	321	178

Table 3: Number of cases selected for the study.

4.2.1. The case of *nuevo*

A careful analysis of the concordance of the adjective *nuevo* revealed that this adjective clearly prefers the pre-modifying position, both in original texts as well as in translations. Apart from the use of *nuevo* as a single pre- or post-modifier I have also distinguished other syntactic combinations, in particular those that refer to multiple pre- and post-modification (always involving *nuevo*, of course), or pre- and post-modification. Table 4 below shows the number of cases found in the various positions with the corresponding percentages to make the data comparable in both corpora.

Position in context	Number of cases in	Number of cases in
	CREA	P-ACTRES
Pre- + post-modification	170 – 44.3%	75 – 39.4%
Single pre-modifier	122 – 31.8%	65 – 34.2%
Fixed expression ‘de nuevo’	63 – 16.4%	14 – 7.3%

Single post-modifier	21 – 5.4%	19 – 10%
Adjective used as noun	4 – 1.04%	-
Predicative position	2 – 0.5%	5 – 2.6%
Multiple pre-modification	1 – 0.2%	5 – 2.6%
Multiple post-modification	-	7 – 3.6%
Total	383	190

Table 4: Positions of *nuevo* in original and translated Spanish.

Figure 6 below shows the 4 most common functions identified, those with 10% or more of the cases in each language. It can be seen that *nuevo* is a predominantly attributive adjective, as the predicative position is very infrequent (0.5% and 2.6%, in originals and translations, respectively). The two most frequent positions in both corpora are the pre-modifying positions, with or without additional postmodifiers, accounting together for 75% of cases in original Spanish and for 73% in translated Spanish. Other uses are marginal in both languages, particularly multiple pre- or post-modification. The use of *nuevo* in the fixed expression *de nuevo*, meaning *again*, occurs in 16% of cases in original Spanish and appears to be much less frequent in translations, with less than half the cases (7%).

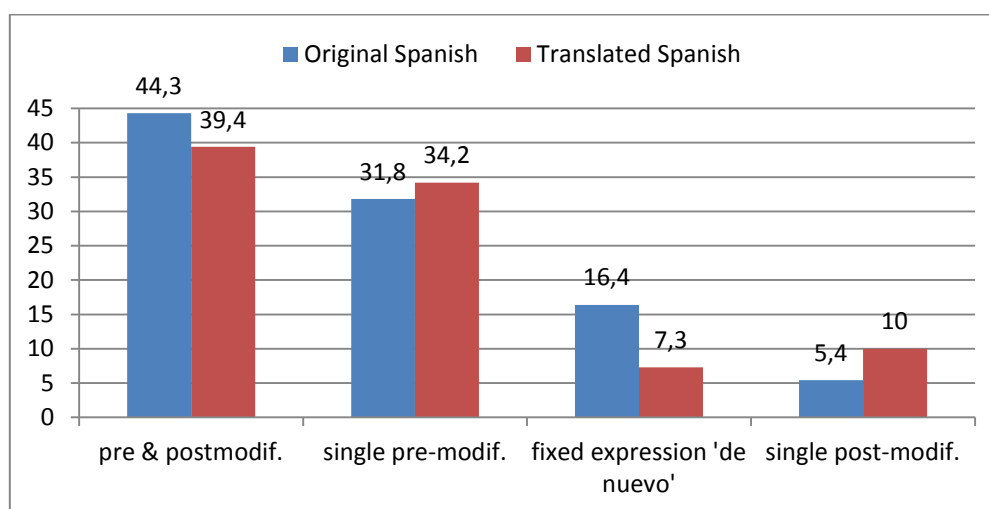


Figure 6: Percentage of occurrence of the various positions of *nuevo* in original and translated Spanish.

In original Spanish the most common use of *nuevo* – 44% of cases - is as a pre-modifier of a noun that also carries post-modification of various types: other adjectives, especially of the classifying type (*el nuevo planeamiento urbanístico consistiría en ...*), and participle clauses (*un nuevo caso revelado ayer*), relative clauses (*un nuevo*

movimiento, que sorteaba la censura), among others. Exactly the same pattern is also the most frequent one in translations, with 39% of cases (*un nuevo plan nacional, un nuevo café que acaba de abrir*). The second most frequent option, single pre-modification, occurs in 39% of cases in original texts (*el nuevo sistema debe compensar a los ayuntamientos, el nuevo servicio se ubicaría en el recinto ...*) and in 34% of cases in translations (*Kurtz había encontrado un nuevo conductor, no le gusta nada mi nuevo trabajo*). Finally, single post-modification is an infrequent option in both corpora, although slightly more frequent in translations than in original texts. The data reveal that there are not many differences between the use of *nuevo* in originals and translations when considering its position in the NP, which is predominantly the pre-modifying one, with or without post-modifiers.

4.2.2. The case of *importante*.

As for *importante*, the analysis has revealed that this adjective is typically of the post-modifying type, with only rare occurrences in the pre-modifying position in both original and translated texts. Table 5 shows the number of cases found in each corpus with the corresponding percentage to make the data comparable.

Position in context	Number of cases in	
	CREA	PACTRES
Single post-modification	112 – 29.7%	59 – 30.7%
Predicative adjective	95 – 25.2%	78 – 40.6%
Multiple postmodification	94 – 25%	27 – 14.06%
Pre- + post-modification	50 – 13.2%	24 – 12.5%
Single pre-modification	24 – 6.3%	3 – 1.5%
Adjective used as noun	1 – 0.2%	-
Multiple pre-modification	-	1 – 0.5%
Total	376	192

Table 5: Positions of *importante* in original and translated Spanish.

Figure 7 below shows graphically that there are important divergences in the use of this adjective in original and translated texts. The single post-modifying position is the most frequent one in original texts, with nearly 30% of cases (*fue una decisión importante, poseía una experiencia importante*) and a very similar percentage was

found in translated texts (*corro un riesgo importante, una variable importante*). However, the predicative position of this adjective is much more frequent in translations (40% of cases: *es importante recordar aquí los argumentos de Kant*) than in original texts (25%: *le afectan numerosos factores y es importante analizarlos*), thus pointing probably towards the influence of the English source texts. The English adjective *important* is also very frequent and may appear more often in the predicative position than its Spanish cognate, leading to this clear difference in use. Original Spanish texts present another 25% of cases of *importante* in a multiple post-modifying structure: *información estadística importante, un salto cualitativo importante*. Translated texts present this position in only 14% of cases: *un valor simbólico importante*.

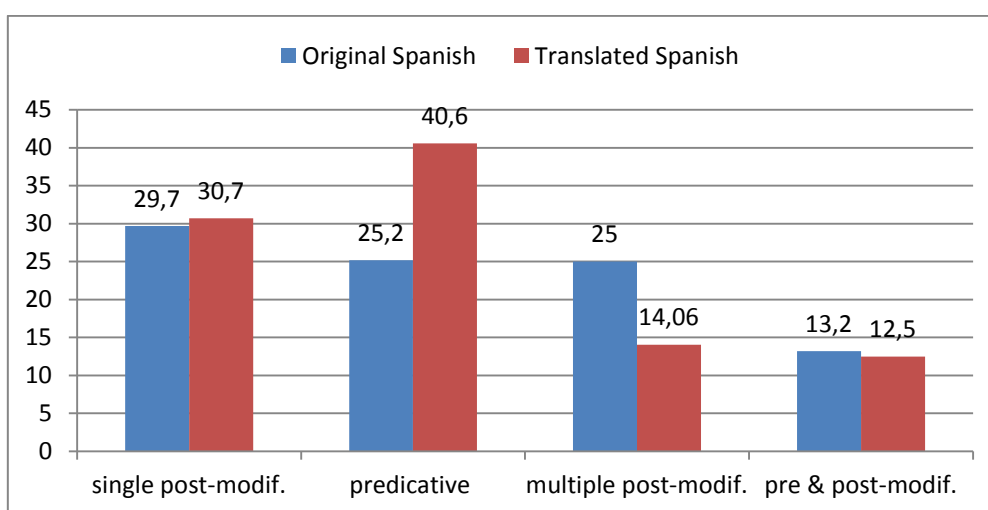


Figure 7: Percentage of occurrence of the various positions of *importante* in original and translated Spanish.

To sum up, 54% of cases in original Spanish are single or multiple post-modifiers, and the predicative position accounts for another 25%. The remaining cases are marginal. In translations only 44% of cases refer to single or multiple post-modification and another 40% to the predicative position. As far as the remaining combinations are concerned, there are no major differences between original and translated texts. Here we can see a trend towards fewer post-modifying positions in translations, but more predicative uses of this particular adjective than original texts, a deviation that can only be explained by the influence of the source language.

4.2.3. The case of *largo*.

Finally, the Spanish adjective *largo* shows a clear preference for the pre-modifying position in both originals and translations. Table 6 below shows the number of cases in each corpus with the corresponding percentage of occurrence:

Position in context	Number of cases	Number of cases
	in CREA	In PACTRES
Fixed expression 'a lo largo de ...'	195 – 52.5%	8 – 4.4%
Single pre-modification	85 – 22.9%	74 – 41.5%
Pre- & post-modification	35 – 9.4%	29 – 16.2%
Noun	26 – 7%	24 – 13.4%
Single post-modification	8 – 2.1%	21 – 11.7%
Predicative	8 – 2.1%	6 – 3.3%
Multiple post-modification	7 – 1.8%	14 – 7.8%
Multiple pre-modification	5 – 1.3%	2 – 1.1%
Numeral	2 – 0.5%	-
TOTAL	371	178

Table 6: Positions of *largo* in original and translated Spanish.

The analysis has revealed that in original texts the adjective *largo* appears mostly (52% of cases) as part of the fixed expression *a lo largo de ...* (*a lo largo de la historia*), meaning *along*. This is interesting because only 4% of cases in translations show this option. It is difficult to try and explain this trend. Of course translated texts are conditioned by their source texts, but translators seem to use other expressions with similar temporal meanings instead of *a lo largo de ...* This fact conditions the rest of the positions identified, as they will always be much more frequent in translations.

In translations, the most frequent use of the adjective *largo* is as a single pre-modifier (41%: *su largo cuerpo*) and this is also the second most common use in originals, but with only half the cases (22%: *recogió un largo aplauso, el largo verano*). The combination of pre- and post-modification occurs in nearly 10% of cases in original Spanish texts (*el largo enfrentamiento que mantienen ambas administraciones, iniciar un largo proceso evolutivo*) and in 16% of cases in translated texts (*un largo camino que descende hacia el río*). The use of the word *largo* as a noun, mainly in expressions of measure indicating length, occurs twice as often in translations as in original texts

(*tiene más de 4.500 kilómetros de largo*), with 13% and 7% of cases, respectively. The use of *largo* as a single post-modifier is also much more frequent in translations (*memorizar un poema largo*) than in original texts (*tenían el morro largo*) (11% versus 2%). The remaining cases are very infrequent in both corpora. Figure 8 below shows the most frequent positions in original and translated texts in percentages.

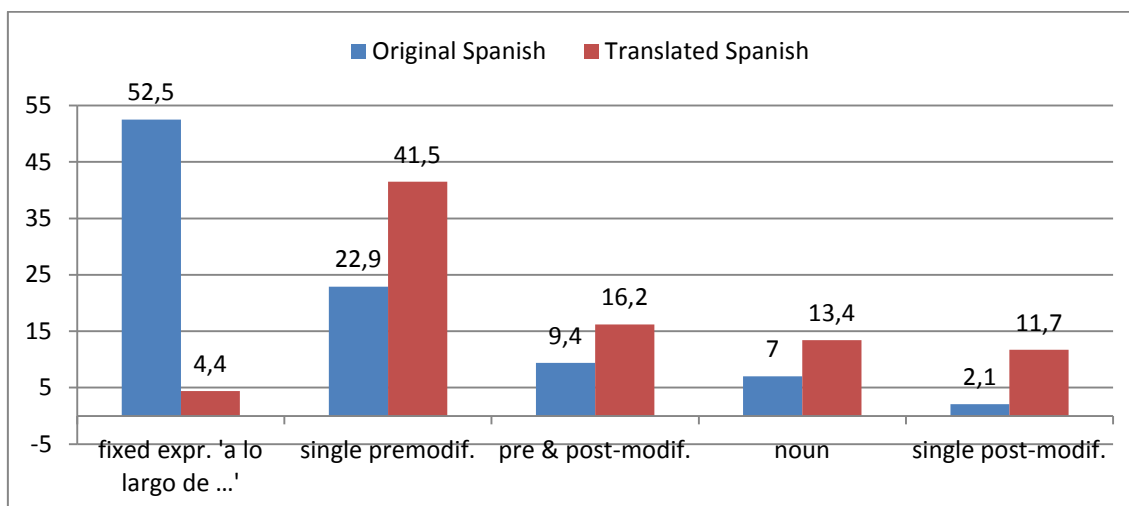


Figure 8: Percentage of occurrence of the various positions of *largo* in original and translated Spanish.

Summing up, the adjective *largo* shows a strong trend towards the pre-modifying position, with or without additional post-modifiers in the same NP. The data show that, apart from the great difference with respect to the fixed expression *a lo largo de ...*, there are no major divergences in the positioning of this adjective in original and translated texts.

5. CONCLUSIONS

From a typological perspective English and Spanish are very different as far as adjective position is concerned and previous studies have shown a generalized overuse of the pre-modifying position in translated Spanish when compared with original Spanish texts. In this paper I have analyzed some of the most common Spanish adjectives in original texts and in texts translated from English. Empirical data for the analysis were extracted from two large corpora: CREA for original Spanish and P-ACTRES for translated Spanish. The aim was to reveal patterns of use and unveil any divergences that might be useful in translator training and in translation quality assessment.

The analysis consisted of two separate sections: a) a quantitative study for determining overuse or underuse of the most frequent adjectives in translations when compared to original texts, and b) a qualitative analysis of the actual use of the most common adjectives in context.

The quantitative analysis actually revealed interesting differences between original and translated language, with a statistically significant overuse of the adjectives *grande*, *bueno* and *largas*, but also a significant underuse of *nuevo* and *importante* in translations. The cases of overuse may be explained by the simplification hypothesis, whereas the cases of underuse point towards an attempt to avoid homogeneity.

The qualitative analysis, on the other hand, has not been conclusive. The adjectives *nuevo* and *largo* show a clear preference for the pre-modifying position in original Spanish, and no important differences were found in their use in translations as far as their position in the NP is concerned. However, they appear to be used less frequently in fixed expressions in translations than in originals (*de nuevo*, *a lo largo de*), thus illustrating a lower degree of typicality in translated language. On the other hand, the adjective *importante* appears to be used much more often in the predicative position in translations than in original texts, and less often in multiple post-modifying structures. These differences can only be attributed to the influence of the English source text, as well as to a trend towards simplification in translated language.

In conclusion, each adjective shows different usage patterns in original texts and different divergences with respect to their translations. No general rule has been noticed, but this may be due to the small number of adjectives studied. It seems to be clear that the point here is lexical: each lexical item behaves in a different way.

REFERENCES

- Baker, M. 1993: "Corpus Linguistics and Translation Studies." In Baker et al. (eds.) *Text and Technology. In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins. 233-250.
- Comrie, B. 1981: *Language Universals and Linguistic Typology*. Chicago: Chicago University Press.
- Croft, W. 1990: *Typology and Universals*. Cambridge: CUP.
- Greenberg, J.H. 1963: Some universals of grammar with particular reference to the order of meaningful elements. In: Greenberg, J.H. (ed.) *Universals of Language*. Cambridge, Mass.: MIT Press. 73-113.
- Laviosa, S. 1996: "Comparable corpora: Towards a corpus linguistic methodology for the empirical study of translation." In Thelen, M. and Lewandowska-Tomaszczyk, B. (eds.) *Translation and Meaning*, Part 3. Maastricht: Hogeschool Maastricht. 153-163.
- Mallinson, G. & Blake, B.J. 1981: *Language Typology*. Amsterdam: North Holland.
- Rabadán, R., B. Labrador & N. Ramón. 2009: Corpus-based contrastive analysis and translation universals: a tool for translation quality assessment English-Spanish? *Babel* 55: 4, 303-328.
- Ramón, N. 2009: Translating Epistemic Adverbs from English into Spanish: Evidence from a Parallel Corpus. *Meta* 54: 1, 73-96.
- Ramón, N. & B. Labrador. 2009: Translations of –ly Adverbs of Degree in an English-Spanish Parallel Corpus. *Target* 20: 2, 275-296.
- Toury, G. 1995: *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins