

# THE *ACTRES* PROJECT: USING CORPORA TO ASSESS ENGLISH-SPANISH TRANSLATION

R. Rabadán, B. Labrador & N. Ramón  
University of León, Spain.

## **Abstract:**

Assessing translation quality is generally seen as a difficult and elusive task because of the inadequacy of the tools available. The aim of this paper is to demonstrate the usefulness of a corpus-based contrastive methodology developed at the University of León (Spain) for identifying instances of translationese. The *ACTRES* project functional framework draws on the work by Bondarko (1991) and Chesterman (1998) and has been designed for translation-oriented cross-linguistic analysis (Rabadán et al. 2004). The long-term study focuses on those semantic areas that are typically problematic for our language pair (modality, quantification, modification, aspectuality, etc). The contrast features a two-step procedure: contrasting typical ways of expressing similar meanings in English and Spanish, and spotting differences between original Spanish and translated Spanish. First, empirical data are extracted from two monolingual ‘comparable’ corpora -The Bank of English and the Corpus de Referencia del Español Actual (CREA); secondly, these results are compared with data from a custom-made translation corpus containing English original texts and their corresponding Spanish translations. The three sets of results provide different types of useful information- i) the resources available (or absence of) in each of the languages to express a given meaning and their relative centrality, ii) the solutions favored by translators to bridge the crosslinguistic disparities and/or gaps and iii) the erroneous or non-existent uses and structures transferred from the source language into the target language. Translation practice, translator training and translation quality assessment (TQA) are the main areas that can benefit from this type of research.

**Keywords:** translationese, comparable corpus, translation corpus, contrastive studies.

## **1. Introduction**

Assessing translation quality is generally seen as a difficult and elusive task because of a lack of conceptual clarity, and the inadequacy of the tools available. How to evaluate the result(s) of a translation procedure tends to depend excessively on the social, political and even ethical stand of whoever is making the evaluative judgement. It seems imperative to emphasize scientific objectivity and reliability as standard criteria so as to curb unverifiable value judgements. Our contention is that this can only be achieved by offering evaluators objective grounds on which to base them, and that these grounds are necessarily linguistic and textual. Recent proposals (notably Bowker 2001) have shown that corpus-based research can provide this type of data. We will argue that corpus-verifiable grammatical translationese may be used (alone or in conjunction with other discriminating criteria/tools) as an indicator of translation quality.

## **2. Translation quality, translationese and translation universals.**

TQA is concerned with judging and evaluating the degree of excellence of translations. Its goal has been summed up by House (2001:156) as revealing “exactly where and with which consequences and (possibly) for which reasons (parts of) translated texts are what they are in relation to their ‘primary texts’”. In short, to find where, how, and possibly why the target textual and linguistic make-up departs from its source.

‘Translationese’ refers to differences between original and translated text/language which cannot be attributed to misrepresentation, but rather to interference of the source language/textual system with the target language/system. It is regularly used in connection

with the distribution of lexical items, although there are recent works that quite aptly use it to indicate ‘grammatical translationese’ (and no reason prevents it from being applied to ‘rhetorical translationese’).

Recent research has also brought up the question of translation universals into translation quality research. These are hypothesis on language and textual tendencies that are a recurrent feature of all translated language, irrespective of language. Among these tendencies and features are simplification, normalization, greater explicitness and conservatism (Baker 1996, Laviosa 1996). The regular version of the **‘normalcy hypothesis’** contends that translations tend to boost the use of typical features of the target language, which can be also understood as an underutilization of the linguistic resources offered by the TL (Reiss 1971) by concentrating on a small number of them. One of our case studies will provide evidence for this.

A further interesting TU hypothesis is **‘the unique items hypothesis’** (Tirkkonen-Condit: 2002: 209). Translated texts would show lower frequencies of linguistic elements that are specific of this target language, i.e., that do not have a ‘similarly perceived’ equivalent. Although generally applied to lexical strings, there is no good reason why this hypothesis cannot be rephrased as the ‘unique grammatical feature hypothesis’, as they are also special in terms of their translation potential. In Spanish, progressive aspect can be expressed by a rather largish inventory of formal resources. In English, the progressive tenses tend to be regarded as the central resource to express this meaning.

In short, TU would refer to unwillingly unavoidable properties of translated language, which differ from those of native produced language, and these happen irrespective of the languages involved, whereas ‘translationese’ is a general term for the properties of translation specific to a particular language pair. These can and obviously do reflect those universal tendencies in particular language-pair-bound areas of grammar.

### **3. Why use corpora to assess translations**

There is a well documented literature of the uses of corpora in translation related endeavours (Zanettin, Bernardini & Stewart 2003). Some of these reasons are the access to empirical evidence and the immediate feedback (provided the questions are adequate).

The pros and cons of whether to use bilingual/multilingual comparable or just parallel corpora have been discussed extensively, questions of design and directionality have also been addressed and problems of applicability in these areas identified. However, when reviewing all these valuable contributions, one cannot avoid the feeling of being treated to a rather vague inventory of the ‘potential applications’ and ‘capabilities’ of corpora. Whereas much of the work done has concentrated on getting the corpus right, it is not so clear that enough attention has been paid to how to actually bridge the very real gap that separates getting descriptive corpus-based work done and putting the results to work (Tymoczko 1998) i.e, the final goal of all applied research.

Recent work is trying to end this state of affairs. Most of these ‘applied’ proposals address evaluation needs in translator education and in the broader curriculum of the prospective ‘language service providers’ (Zanettin, Bernardini & Stewart 2003a:1).

To my knowledge Bowker (2001) has put forward the most articulated and realistic proposal to date. Her ‘evaluation corpus’ is conceived specifically for specialised translation and is organized in a flexible way, making it a real collaborative tool. It would be obviously useful outside the teaching environment, but it faces, as most corpus-based so-called utilities a nearly insurmountable problem, time, and does not seem to travel well into

other ‘education contexts’. Can teachers/researchers/reviewers afford to devote time to ‘building’ expert evaluation corpora (Varantola 2000)? Will the benefits of building it and using it exceed the effort of tool-building, or rather they do not pay off? Why should not a ‘service provided’ expect to be supplied with tools to do his/her job straight away? Are the ‘translation reviewers’ ‘savvy enough in corpus use to make a good job of it?’

Corpora, of whichever type, do not provide answers and/or solutions to its intended users, and that further work between description and its application is needed. This should provide time-saving ready- to-use data to feed the final user tool. In order to be efficient it has to address pivot translationese areas in a given language pair and a given direction.

Another aspect is the fact that already existing corpora can be further exploited in combination (partially) with other resources for a variety of intended applied goals. In other words, we do not think it is necessary to complete corpora anew for each new evaluation process. The same ‘source corpora’ can be used satisfactorily for a number of activities, among them assessment.

The purpose of this paper is to show how to identify these areas by focusing on 3 ‘grammatical translationese-prone’ areas in English-Spanish translation. The selected features tend to be problem triggers in English-Spanish translation: quantifiers, modifiers of nouns and the translation of the English Simple Past form. Each of them illustrates a different actualization of translationese: quantifiers reveal different distribution of choices when considered in native and translated Spanish, in nominal characterization there are clear trends towards the overuse or underused of some of the options available in the target language, and some of the more salient and idiomatic meaning encoding capabilities of Spanish ‘imperfecto’ are simply missed when translating Simple Past forms.

The research reported here is based on three corpora: two large general language monolingual corpora, the Bank of English and the CREA<sup>1</sup>, and the ACTRES parallel corpus.

Both monolingual ‘virtual’ corpora (Varantola 2003) include many million words of running text and have a similar internal structure concerning intralinguistic varieties, register distribution, mode and statistical dimensions. The two corpora act as source corpus, since restrictive choices have been made concerning language variety, mode, subcorpora and size. For convenience the varieties chosen are UK English and European Spanish. Because of its applied aim, the mode is ‘written’. Books, mags, newspapers and ephemera were the subcorpora chosen, which means that the total size is over 30 m words for the English corpus and around 37m words for CREA. Both monolingual corpora have their own built-in tagging, parsing and querying systems, which differ substantially, but nevertheless enable the user to retrieve the same type of information. They have been used as the source for ‘comparable data’ (original language in English and in Spanish) in the contrastive stage.

The ACTRES parallel corpus of English originals and their corresponding Spanish translations mirrors the qualitative construction criteria of both the Bank of English and CREA, i.e. subcorpora, register distribution, mode, etc. It differs in two respects: instead of being a complete text corpus, ACTRES consists of extracts of between 5,000 and 15,000 words from books (fiction and non-fiction), the press (newspapers and magazines) and ephemera. The reason is quite simply, copyright. The English language materials are not

---

<sup>1</sup> Cobuild’s the Bank of English (<http://titania.cobuild.collins.co.uk>). CREA (Corpus de Referencia del Español Actual <http://www.rae.es>)

restricted to materials originated in the UK, as choice of SL variety was deemed to be irrelevant when the directionality is English into Spanish. ACTRES is still under construction and includes approximately 800,000 words evenly distributed between the two languages. The aim is to reach a size of approximately 2 million words, one million per language. This would allow for studies that are sufficiently representative of translation and translated language between English and Spanish. In the meantime, materials are used in different ways as a ‘diagnostic tool’ and always in conjunction with comparable data. One of the strategies is to use the parallel corpus as a ‘virtual corpus’ from which to extract different ‘sample corpora’: one frequently used is selecting ten random pairs focusing on the grammatical phenomenon being analysed (past tense, modal verbs, etc). A more classic approach is taking a random portion of materials as a sample corpus and search for item ‘x’. (Both approaches have been used in the case studies that illustrate this paper). As corpus management tools ACTRES uses Translation Corpus Explorer (WebTCE)<sup>2</sup>, developed and constantly refined in Norway for the English-Norwegian Parallel corpus Project (Hofland & Johansson 1998).

#### **4. Method and procedure: using comparable and parallel corpora**

The ACTRES project typical research line is based on a three-step methodology, an interlinguistic analysis, a cross-linguistic translation analysis, and a subsequent intralinguistic analysis:

First, empirical data are extracted from two monolingual ‘comparable’ corpora – Cobuild and CREA on the basis of cross linguistic similarity perception, and analysed following the sequence selection, description, juxtaposition and contrast. The semantic cross-linguistic *tertium comparationis* is set up at the descriptive stage. The aim is to find evidence –both quantitative and qualitative- of the resources available to express a given meaning in English and Spanish and their distribution. The results of the interlinguistic contrast include both similarities and differences in the formal realization of a particular semantic function.

In the second stage, the same ‘input’ is searched in the parallel corpus in order to obtain a ‘diagnostic sample’ of the rendering of the grammatical feature into the target language. These results are then analyzed for ‘meaning’ (i.e. the ‘tertium comparationis’ labels) so as to obtain the distribution of translated usage.

The third and final analytical stage compares the ‘original language(s)’ evidence with the ‘diagnostic data’ obtained from the ACTRES parallel corpus containing English original texts and their corresponding Spanish translations. This allows us to identify differences between original and translated Spanish, which may be due to a particular translation practice/culture, due to the influence of the source language, in this case English, or simply to incompetent translating. This intralinguistic contrast will eventually highlight the extent to which translationese applies.

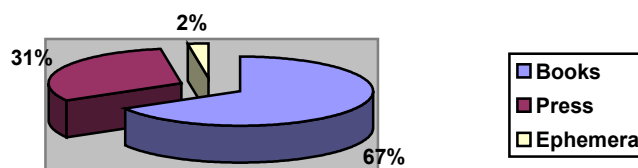
Three different case studies will illustrate this process: one apparently without traces of translationese (normalization strategies), the second exemplifies a clear case of translationese (transfer of grammatical usage), and the last one shows the rejection of a particular resource available in translated language (negative use of unique feature).

---

<sup>2</sup> This tool can be examined at <http://khnt.hit.uib.no/webtce.htm> (01/05/05).

## 5. Case studies

Case studies 1 & 2 utilize the comparable monolingual corpora together with a ‘sample parallel corpus’ as a ‘diagnostic tool’. This sample parallel corpus includes nearly 40,000 words in each language and includes texts from each of the subsections to be represented in the larger corpus. Figure 1 shows the register distribution:



**Figure 1: Register distribution of the sample parallel corpus.**

Table 1 below summarises the quantitative data in each subsection, and for each language English and Spanish:

	ENGLISH	SPANISH
<b>BOOKS</b>	24,747	25,437
<b>PRESS</b>	11,448	11,961
<b>EPHEMERA</b>	881	823
<b>TOTAL</b>	<b>37,076</b>	<b>38,221</b>

**Table 1: Number of words in the sample parallel corpus<sup>3</sup>.**

### 5.1. Case Study 1: Intensive quantification.

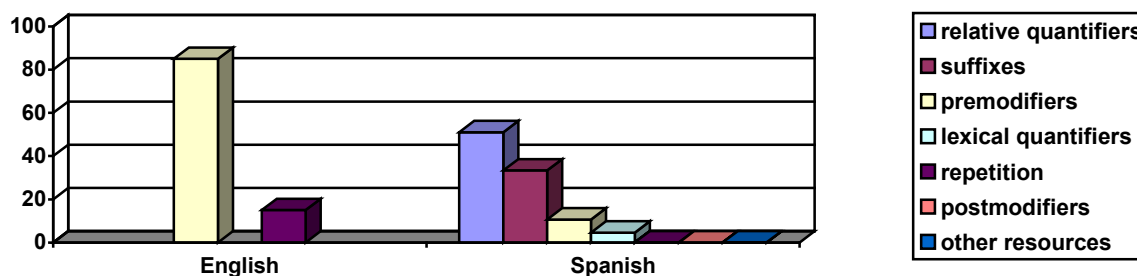
In a first large-scale contrastive study on quantification (Labrador de la Cruz 2002), a list of quantifiers was selected as the object of study. This list was compiled using a number of English and Spanish grammars - Quirk et al (1989), Downing et al (1992), Berry (1997) and Biber et al (1999) and Bello (1981), Alarcos (1994), Matte Bonn (1995) and Bosque et al (1999) respectively, as well as our own intuition and the opinion of several native informants. These lexical items were searched for in Cobuild's Bank of English and RAE's CREA; only those subcorpora that represent British English and Peninsular Spanish were consulted, which account for more than 42 million words in English and 49 million words in Spanish.

Those quantifiers with fewer than 10 occurrences were not included, so finally 188 word forms were studied, 78 of which were English and 110 Spanish. The reason for the higher rate in Spanish is mainly its morphological richness – sometimes one lexeme has four, five or even more word forms. The total number of concordances to be analysed amounted to 48,875 (21,491 of which were English and 27,384 Spanish).

After the analysis and classification of all those concordances, we found that these quantifiers express 56 different functions, 33 of which are inherently quantifying. The

<sup>3</sup> It may be noticed that the translations into Spanish are generally somewhat longer than their corresponding English originals, except in the case of ephemera, where omission is frequent.

interlinguistic contrast between the formal realisation of these functions in English and Spanish shows similarities and differences concerning the type of quantifying resources employed, their distribution and their frequency rates. One of the functions in which these languages most differ is intensification – the way English and Spanish intensify quantification. It is an important function, and it ranks the fifth of the 56 functions in terms of frequency - 5.45% of the uses of quantifiers are intensive.



**Fig 2: Intensive quantification in English and Spanish. Comparable data.**

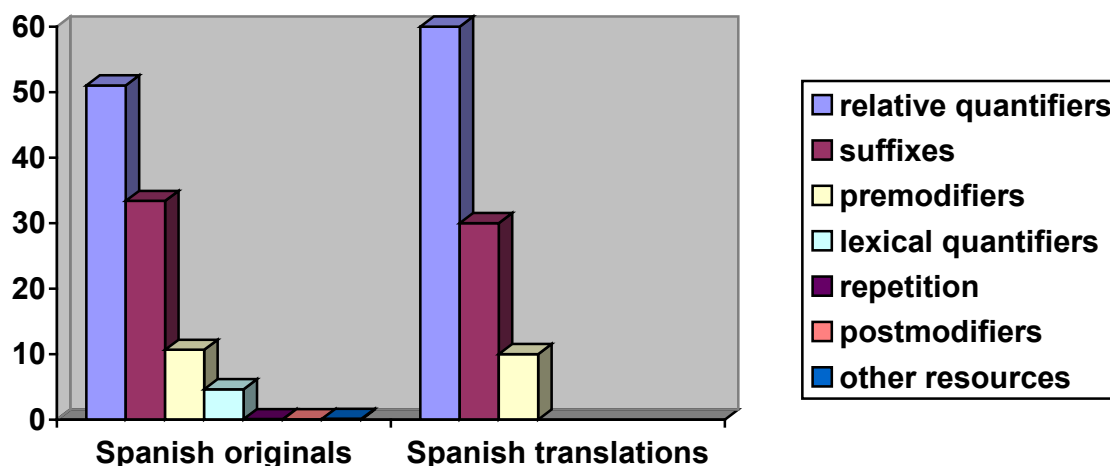
As can be seen in figure 2, English mainly makes use of premodification, - 84.96% of the times, e.g. ‘I think that’s causing *quite a lot* of concern’, or ‘We have been through *so much* together, we will always be friends’ and secondly, repetition is used with this purpose – 15.03%, e.g. ‘He is picking *many many* places where he wants to move’ or ‘I went off and did *loads and loads* of interviews’. However, although Spanish also uses premodifiers, e.g. ‘Esta vez arrancó echando *un buen montón* de humo y aceite a la cara de Paco’ and repetition of quantifiers, e.g. ‘es digna de *mucha, mucha* consideración’ to intensify quantification, they occupy lower positions in the rank scale – premodification: 10.66% of the times and repetition, as seldom as 0.08% of the times. Relative quantifiers, e.g. ‘La existencia de *tantos* sistemas añade nuevas dificultades’ and suffixes, e.g. ‘Sin embargo, tardó *poquísimo* en volver’ are the main formal devices used to express intensive quantification in Spanish, with percentages of 51.02% and 33.39% respectively; other minor resources found are postmodification (with a frequency rate of 0.04%), e.g. ‘que hay un *montón exorbitante* a eso nadie le pone reparo’ and lexical quantifiers (with a rate of 4.64%), e.g. ‘aquel maletín parecía de *suma* importancia para él’.

With such a divergence in the English-Spanish contrast of the formal representation of this function (intensive quantification), it is a good candidate for translationese. we searched for possible discrepancies between native and translated usage, focusing particularly on those instances where the mismatch can be attributed to the influence of English grammar on Spanish translations.

The analysis of the sample parallel corpus reveals a rate higher than usual of premodification and repetition- typical resources of the English language- to the detriment of other more idiomatic ways of intensive quantification in Spanish, namely the use of relative quantifiers and suffixes.

As figure 3 shows, only the three most important resources in Spanish originals have been found in the Spanish translations and the ranking remains the same: first, relative quantifiers (with a 60% of the times – a slightly higher proportion than in Spanish

originals); in a second place, suffixes (with a 30% - a slightly lower proportion) and premodifiers (with a 10%, approximately the same rate as in Spanish originals).



**Figure 3: Intensive quantification in Spanish originals and translations. ‘Diagnostic data’.**

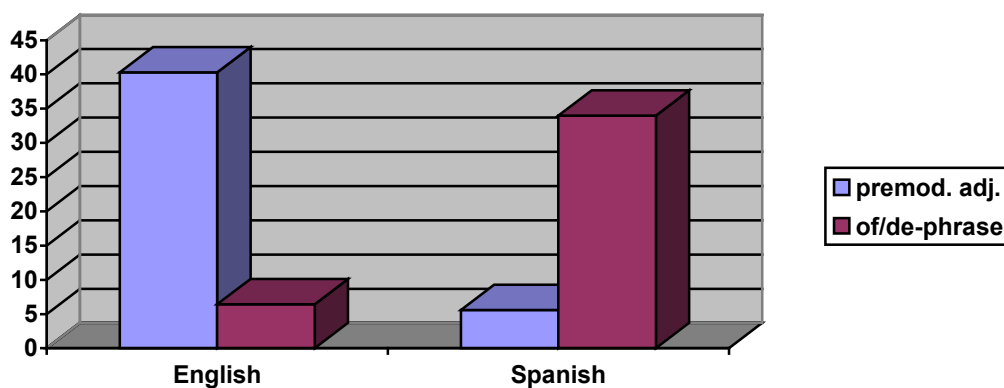
The size of the parallel corpus may account for the fact that the minor resources used to intensify quantification in Spanish originals have not occurred in the Spanish translations but as for the major resources, the results are likely to be extrapolated. When compared with Spanish, one of the most striking features of English is the use of quantifiers in conjunction with intensifiers forming long phrases- long chains of premodifiers attached to the head of the phrase. However, translators do not seem to be tempted to transfer this typical way of intensive quantification in English; on the contrary, they stick to the most idiomatic and natural Spanish resources for keeping the same function across the two languages. While this behaviour guarantees correction it plays in detriment of the wealth of resources offered by the target language. It seems the **‘normalcy hypothesis’** is at play in reducing the range of options and thus narrowing the inventory available to translated Spanish.

The reasons why translators do not fall into some sort of interlanguage here may be the extent to which Spanish sets restrictions on the use of premodifiers, especially with some quantifiers (e.g. *muy poco* is grammatical but *muy mucho* is not), or perhaps to how deeply suffixes are rooted in our minds, as native Spanish speakers. In the case of the quantifier *mucho*, for instance, only the suffix *-ísimo* can be attached, but in the case of the quantifier *poco*, as many as 17 different variants of the lexeme can be formed, due to the variation in gender and number and the possibility to use suffixes like *-ito*, *-itito* or *-itín*, apart from *-ísimo*. Another reason for this lack of traces of translationese is the lexicalisation of relative quantifiers in Spanish – we express in a lexical term, e.g. *tanto*, *cuánto*, what English does in a collocation made up of a quantifier and a premodifier, e.g. *so much*, *how much*.

## 5.2. Case Study 2: Nominal Characterization

The modification of nouns within the boundaries of the noun phrase is a particularly problematic issue in English-Spanish translation. The two languages have opposite

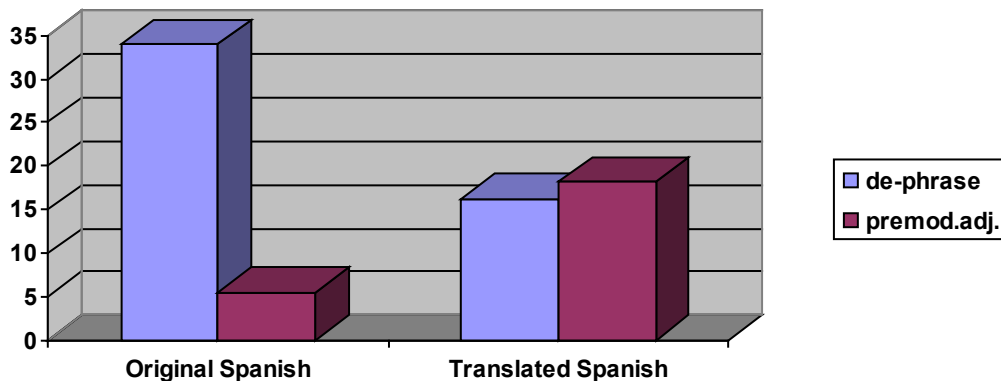
unmarked positions for adjectives, the most common noun modifier, with English locating adjectives mostly in prenominal positions and Spanish in postnominal position. In addition, both languages have available a wide range of formally similar structures to express modifying meanings, but the use and distribution of these structures differs greatly. A large-scale contrastive study (Ramón García 2003) was carried out using data from Cobuild and CREA. Only written texts from the European varieties of English and Spanish were used, amounting to slightly over 30 million words in each case. A whole of 7,882 concordances were extracted from the ten most common nouns in each language, 3,939 in English and 3,943 in Spanish, and their syntactic surroundings analyzed in search of instances of nominal modification. The resources isolated were subsequently classified semantically. Eleven broad semantic functions were identified in the field of noun modification. The descriptive function was found to be the most common one in the two languages. This case study will focus on the function ‘descriptive’ as conveyed by two single-item modifying structures: de-phrases and pre-modifying adjectives, where the divergences are significant and may be traced back to the influence of the source language.



**Fig. 4. De-phrases and pre-modifying adjectives in English and in Spanish. ‘Comparable’ data.**

As shown in Fig 4. native speakers of English make a heavy use of premodifying adjectives with descriptive meanings (*a wonderful time, a great year*), with about 40% of cases, whereas prepositional phrases headed by the preposition *of* (*of*-phrases) occur only in slightly over 5% of cases (*this man of only 22, a night of moonlit romance*). In contrast, the Spanish language seems to rely heavily on prepositional phrases headed by the preposition *de* (*de*-phrases), the formal counterpart of *of*-phrases, for expressing this meaning, occurring in over 30% of instances (*el tiempo de la fiesta, un año de temperatura social elevada*). Premodifying adjectives are also an option in Spanish, but native speakers use them with descriptive meanings in only about 5% of cases (*su turbulenta vida, un buen momento*). These fundamental typological differences hint at possible sources of problems in translations from English into Spanish. When ‘diagnostic’ data are brought into the picture, we obtain the discrepancies –the evidence of translationese - between the native and translated uses in Spanish.





**Fig. 5: De-phrases and pre-modifying adjectives in original and translated Spanish. ‘Diagnostic’ data.**

Figure 5 shows that de-phrases are used with descriptive meanings in 33.97% of cases of single descriptive modification within the boundaries of the NP in original texts written in Spanish, whereas only 16.23% of cases were found in the translations from English. Some examples extracted from the parallel corpus are: *el pelo de un rojo intenso, un hombre de buen tamaño, individuos de aspecto enfermizo, un enchufe de inadecuada conducción eléctrica*, etc. The smaller number of de-phrases in Spanish translations from English may be attributed to the fact that this use does not occur very often with formally parallel of-phrases in English texts.

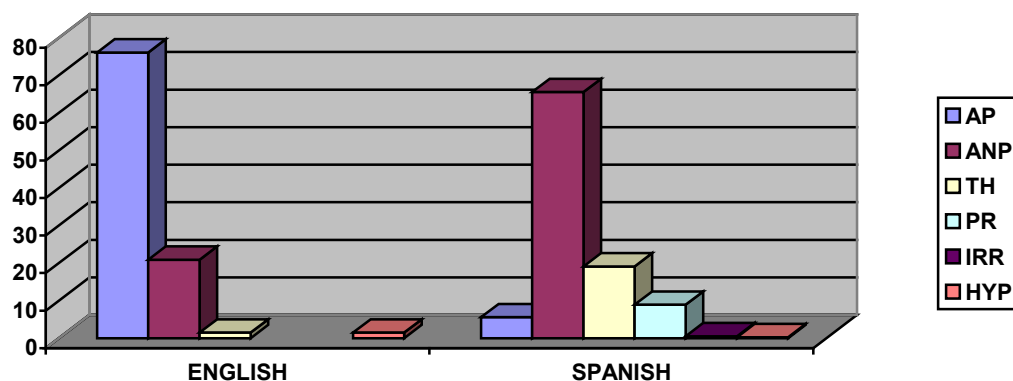
There is also evidence that single pre-modifying adjectives occur with a descriptive meaning in only 5.59% of cases in Spanish original texts, but this figure soars to 18.21% of cases in translations from English original texts. Examples from the Spanish translations are: *un grave problema, la extraña criatura, una enorme pirámide*, etc. The Spanish grammar allows for this option, although native speakers make scarce use of it and mainly restrict it to highly connotative cases or fixed expressions, some of which also occurred in our parallel corpus: *mala espina, puro teatro*. However, translators clearly overuse pre-modifying adjectives with a descriptive meaning in translations from English, leading to a high frequency of rather unidiomatic expressions such as: *el pesado baúl, la plateada criatura, este eficaz sistema, este notable informe, un complicado reloj*, etc. In addition, the parallel corpus included many instances of multiple modification where a pre-modifying adjective was part of the chain. This overuse is most probably due to the influence of the unmarked position of adjectives in English, which is the pre-modifying position.

### **5.3. Case study 3: The ‘absolute past’ and the Spanish *imperfecto/pretérito* option.**

A second sampling strategy has been used in this pilot study. It consisted in selecting 100 random pairs featuring the language feature under scrutiny in the SL part. This has proved particularly useful when the querying item is not or cannot be a lexically defined item, as with past tense forms. Since the data obtained function as a ‘working hypothesis’ which will need further extensive testing, the results- as in the previous case studies- are not to be taken as final.

A full-scale inquiry into the translation possibilities of the English Simple Past into Spanish (Rabadán 2005) has yielded the results in figure 6. The function ‘absolute past’

represents 76.07% of all cases analysed in English, followed by the ‘anaphoric past’, which comes to 20.9%. In Spanish, however, these two primary meanings are normally associated with different tense forms: the *pretérito* stands for the ‘absolute past’ in all cases recorded (100%). ‘Absolute past’ comes just to 5.61% of cases in native usage of *imperfecto* whereas ‘anaphoric past’ is the meaning of 65.56% of the cases, followed by a string of other well represented functions such as ‘temporary habit’ 19.13% vs. 1.5% in the English corpus, ‘progressive’ (8.93% vs. no representation in English), ‘irrealis’ (0.51 vs no case in English) and ‘hypothetical (0.25% vs. 1.5% in English).

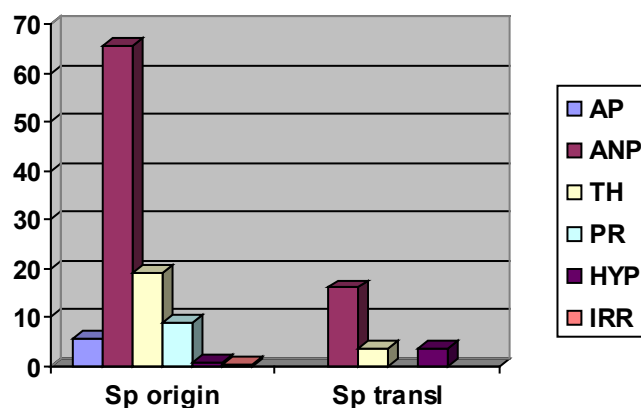


**Fig 6. ‘Absolute past’ [AP] and ‘anaphoric past’ [ANP] representation in English and in Spanish. ‘Comparable’ data.**

Our ‘diagnostic data’ reveal a radical departure from native usage in the translation solutions chosen. [AP] is generally translated by a *pretérito* (77.14%), which agrees with the native usage; all the other meanings identified have been rendered by an *imperfecto* or, on a few occasions, by other –generally lexical and phraseological- resources, as in 39P below, 39P. It was Father Martin’s idea that I should write an account of how I found the body.// Fue idea del padre Martin que yo pusiera por escrito mi experiencia del hallazgo del cadáver.

[ANP] is translated by an *imperfecto* in 16% of cases, and [TH] and [HYP] by 3.42% each. The most obvious discrepancy, as shown in fig. 7 is the use of the *imperfecto* in native and translated Spanish. The Spanish form is able to convey ‘absolute past’ when it is employed as a narrative device in literary (and journalistic) language in order to focus on a specific action or event. As in [AP/322] La Voz de Valencia. Diario de tendencia derechista, próximo a Calvo Sotelo, **aparecía** el 3 de agosto controlado por Esquerra republicana.

This use, labelled by many Spanish linguists as ‘perfective imperfecto’ tends to appear accompanied by a time adverbial in the contexts mentioned above and they can be substituted by a *pretérito*.



**Fig. 7: ‘Absolute past’ [AP] and ‘anaphoric past’ [ANP] in original and translated Spanish *imperfecto*. ‘Diagnostic’ data.**

The results indicate that there is a raw underutilization of the *imperfecto* when translating the meaning ‘absolute past’ (the so-called perfective *imperfecto*). No evidence has been found of this meaning in translated Spanish, which seems to prove the usefulness of the ‘**unique grammatical feature hypothesis**’ discussed earlier as a tool to both identify the text as a translation or an original and provide empirical data to produce an informed quality assessment report. In other words, the absence of this ‘original language’ feature would detract from the ‘quality’ of the translated text, whereas its presence would be an indicator of ‘higher quality’. The closer to the original language distribution, the higher the translation would rank in terms of quality.

## 6. Conclusion

The data shown in these case studies clearly illustrate the type of mismatches that may be found between original texts and translations in three particular semantic areas: intensive quantification, descriptive characterization and the expression of the absolute past. We have argued and demonstrated, that in some specific cases, like modification, the frequency (or lack of) distribution of particular grammatical structures associated with specific meanings in the target language may be used as an indicator of translation quality, and therefore, as a criterion for assessing the quality of translations. The smaller the disparity between native and translated usage in the use of these structures, the higher the translation rates for quality. Although there are other factors that intervene in the quality of a given translation, surely the most tangible and most widely accepted is language correctness, which of course embodies semantic and pragmatic appropriateness. Work in progress aims at developing an empirically-based tool aimed at translation reviewers and other language service providers, which ideally would be used in conjunction with other translation aids.

## REFERENCES

- Alarcos, E. 1994. *Gramática de la lengua española*. Madrid, Espasa Calpe
- Baker, Mona. 1996. Corpus-based Translation Studies: The Challenges that Lie Ahead. In H. L. Sommers (ed). *Terminology, LSP and Translation*. Amsterdam/Philadelphia: Benjamins. 175-186.
- Bello, A. 1981. *Gramática de la lengua castellana*. Tenerife, Instituto Universitario de Lingüística
- Berry, R. 1997. *Determiners and Quantifiers*. Glasgow, Collins Cobuild
- Biber, D. Johansson, S. Leech, G. Conrad S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London, Longman
- Bondarko, A. 1991: *Functional Grammar: a Field Approach* Amsterdam / Philadelphia: Benjamins (tr. from Russian original by I. S. Chulaki).
- Bosque, I & Demonte, V. (eds.) 1999. *Gramática descriptiva de la lengua española*. Madrid, Espasa
- Botley, S., McEnery, A. and Wilson, A. 2000. *Multilingual corpora in teaching and research*. Amsterdam: Rodopi.
- Bowker, Lynne. 2001. Towards a methodology for a corpus-based approach to translation evaluation. *Meta* 46(2): 345-64. Also online at <http://www.erudit.org/meta/2001/v46/n2/index.html>
- Chesterman, A. 1998: *Contrastive Functional Analysis*. Amsterdam/Philadelphia: John Benjamins.
- Downing, A. & Locke, P. 1992. *A University Course in English Grammar*. New York/London, Prentice-Hall International
- Gellerstam, Martin. 1986. Translationese in Swedish novels translated from English. Lars Wollin and Hans Lindquist (eds). *Translation Studies in Scandinavia*. Lund: CWK. 88-95.
- Hofland, Knut and Stig Johansson. 1998. The Translation Corpus Aligner: A Program for Automatic Alignment of Parallel Texts. Johansson, S. & S. Oksefjell (eds.) *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*. Amsterdam: Rodopi. 87-100.
- House, Juliane. 2001. How do we know when a translation is good? Erich Steiner & Colin Yallop (eds). *Exploring Translation and Multilingual Text Production: Beyond Content*. Berlin/NY: Mouton de Gruyter. 127-160.
- Labrador de la Cruz, Belén. 2002. *Estudio contrastivo de la cuantificación inglés-español*. En *Tesis Doctorales 2000*. Universidad de León. CD-ROM. León: Secretariado de y Medios Audiovisuales de la Universidad de León.
- Laviosa, Sara. 1996. Comparable corpora: Towards a corpus linguistic methodology for the empirical study of translation. In M. Thelen & B. Lewandowska Tomaszczyk (eds). *Translation and Meaning, Part 3*. Maastricht: Hogeschool Maastricht. 153-163.
- Matte Bonn, F. 1995. *Gramática comunicativa del español*. I & II. Madrid, Edelsa
- Quirk, R. Greenbaum, S. Leech, G. & Svartvik, J. 1985/1989. *A Comprehensive Grammar of the English Language*. New York, Longman
- Rabadán, R. et al. 2004: English-Spanish Corpus-based Contrastive Analysis: Translational Applications from the ACTRES Project. Lewandowska-Tomaszczyk, B. (ed.) *Practical Applications in Language and Computers. PALC 2003*. Frankfurt am Main: Peter Lang. 141-151.

- Rabadán, Rosa. 2005. The Applicability of Description: Empirical research and translation tools. Carmen Toledano (ed) *Contemporary Problematics of Translation Studies*, RCEI 51 (forthcoming).
- Ramón García, Noelia. 2003. *Estudio Contrastivo Inglés-Español de la Caracterización de Sustantivos*. León: Universidad de León.
- Reiss, Katharina. 1971. *Möglichkeiten und Grenzen der Übersetzungskritik*. München: Hueber.
- Santos, Diana. 1995. On grammatical translationese. Kommi Koskeniemi (ed). *Proceedings of the 10<sup>th</sup> Nordic Conference of Computational Linguistics, NODALIDA-95, Helsinki 29-30 May 1995*. Helsinki: Dpt. Of General Linguistics, University of Helsinki. 59-66. Also online at <http://linguateca.pt/Diana/download/nodalida96.rtf> Accessed April 20, 2005.
- Tirkkonen-Condit, Sonja. 2002. Translationese- a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target* 14(2): 207-220.
- Tirkkonen-Condit, Sonja. 2004. Unique items--- over- or under-represented in translated language?. Anna Mauranen & Pekka Kujamäki (eds). *Translation Universals- Do they exist?* Amsterdam/ Philadelphia: Benjamins. 177-186.
- Tymoczko, Maria. 1998. Computerized corpora and the future of Translation Studies. *Meta*, XLIII (4): 653-659.
- Varantola, Krista. 2003. Translators and disposable corpora. In F. Zanettin, S. Bernardini & D. Stewart (eds) *Corpora in Translator Education*. Manchester, UK: St Jerome. 55-70.
- Zanettin, Federico, Silvia Bernardini & Dominic Stewart .2003a. Corpora in Translator Education: An Introduction. In F. Zanettin, S. Bernardini & D. Stewart (eds) *Corpora in Translator Education*. Manchester, UK: St Jerome. 1-14.