

CORPUS-BASED CONTRASTIVE ANALYSIS AND TRANSLATION UNIVERSALS: A TOOL FOR TRANSLATION QUALITY ASSESSMENT ENGLISH→SPANISH¹

Rosa Rabadán, Belén Labrador and Noelia Ramón

Affiliation: University of León (Spain) (same for the three authors)

Address (same for the three authors):

Departamento de Filología Moderna

Facultad de Filosofía y Letras

Campus de Vegazana s/n

Universidad de León

24071 León

Spain

Abstract:

Assessing translation quality is generally seen as a difficult 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 (ACTRES² Project) developed at the University of León (Spain) for identifying instances of low-quality rendering of grammatical features when translating from English into Spanish using translation universals. The analysis provides information about: i) the resources available (or absence thereof) in each of the languages to express a given meaning and their relative centrality; ii) the solutions favored by translators to bridge the cross-linguistic disparities and/or gaps; iii) the erroneous or non-existent uses and structures transferred from the source language into the target language. These results can be systematized in terms of simplification, interference, or unique grammatical features. Additional areas that can benefit from this type of research are translation practice, translator training and foreign language teaching (FLT), among others.

Résumé:

L'évaluation de la qualité des traductions est généralement considérée une tâche difficile à accomplir à cause de l'inadéquation des instruments disponibles actuellement. L'objectif de cet article est de démontrer l'utilité d'une méthodologie contrastive basée sur corpus (Projet ACTRES) développée à l'Université de León (Espagne) qui emploie des universels de traduction pour identifier des cas de basse qualité dans des traductions de l'anglais à l'espagnol. L'analyse apporte de l'information sur : i) les ressources disponibles (ou l'absence de ressources) en chaque langue pour exprimer une signification particulière et sa centralité relative; ii) les solutions favorisées par les traducteurs pour surmonter les disparités et/ou carences entre les deux langues; iii) les structures et les utilisations erronées ou inexistantes qui ont été transférées de la langue source vers la langue cible. Ces résultats peuvent être systématisés dans les termes de simplification, interférence, ou traits grammaticaux uniques. La pratique de la traduction, la formation des traducteurs et l'enseignement des langues étrangères sont d'autres disciplines qui peuvent bénéficier de ces résultats, entre autres.

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 of a translation procedure tends to depend excessively on the social, political and even ethical stand of whoever is making the evaluative judgment. It

seems imperative to emphasize scientific objectivity and reliability as standard criteria so as to curb unverifiable value judgments. This is particularly relevant when published translated materials frequently show grammatical uses that turn the text difficult to understand or even partially meaningless in the target language (TL), causing a deficient flow of the text and a perception of overall low-quality.

This paper offers a sound theoretical background to the concepts of translation universals (TUs) and translationese, i.e. features of translated language that can be attributed to the influence of the source language (SL). The empirical analysis included here has provided three distinct types of translationese which are identified and described. The various advantages derived from the combined use of different types of corpora in translation research in general, and in translation quality assessment in particular, are also addressed and commented upon in detail, and the main features of the comparable and parallel corpora used in this paper are briefly summarized. The contrastive methodology employed for the case studies in this article is outlined in Section 5. The *ACTRES* project 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). This particular method is then illustrated with three different case studies that represent various ways of describing deviations in translated Spanish. Finally, these differences are systematized in such a way that they can be used in combination to assess translated texts.

Our contention is that corpus-based research can offer evaluators objective data on which to build reliable and usable evaluation methods, and that the ensuing empirically-based tools are necessarily linguistic and textual. This paper argues that corpus-verifiable grammatical usage in certain problem-areas may be used (alone or in

conjunction with other discriminating criteria/tools) as an indicator of translation quality for non-specialized translation English→Spanish.

2. Translation quality, translationese and translation universals.

Translation Quality Assessment (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 language-pair specific contact (see Mauranen 1999; Baker’s ‘third code’ 1998, Toury’s ‘interlanguage’ 1980: 71 & ff., Toury’s ‘translation-specific lexical items’ 1995: 2006-20). The term ‘translationese’ is regularly used in connection with the distribution of lexical items, although there are works (Santos 1995) that quite aptly use it to indicate ‘grammatical translationese’ and no reason prevents it from being applied to ‘syntactic’ or ‘rhetorical translationese’ as well.

Recent research has also brought the question of translation universals into translation quality research. These are hypotheses on language and textual tendencies that are a recurrent feature of all translated language, irrespective of the languages involved. Among these tendencies and features are: simplification (Baker 1993, Laviosa 1996); explicitation (Olohan and Baker 2000); interference (Toury 1995); under-representation of unique TL items (Tirkkonen-Condit 2002, 2004). Although the goal

of most research in translation universals (TU) has been to find ways to discriminate translations from non-translations by focusing mainly on lexical aspects, some of these translation universals are particularly well suited to serve as tools in order to identify grammatical misuses in translations from English into Spanish. They are: the ‘simplification hypothesis’ (Baker 1993), the ‘law of interference’ (Toury 1995; Mauranen 2004) and the ‘unique items hypothesis’ (Tirkkonen-Condit 2002).

The ‘**simplification hypothesis**’ partially overlaps with normalization and standardization (Toury 1995: 267-274) and suggests that translations tend to boost the use of typical features of the target language, which can be also understood as an underuse of the linguistic resources offered by the TL (Reiss 1971) by concentrating on a small number of them, as it happens in case study I below.

The ‘**law of interference**’ has been understood as a ‘non-universal’ (Baker 1993), as a prime universal (Toury 1995) or as transfer (Mauranen 2004: 79). Interference is considered as the deviation from TL norms towards the SL norm, i.e. ‘dispreferred features’ in the TL, such as the pre-modifying adjectives in Spanish in case study II.

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 features hypothesis**’ since these are also special in terms of their translation potential, as will be shown by case study III below.

In short, TUs would refer to properties of translated language, which differ from those of original language, and that happens irrespective of the languages involved (see Baker 1993: 243), whereas the concept ‘translationese’ is a general term for language-specific features that typically occur in translated language or whose frequency in translated texts differs significantly from their frequency in TL originals. These can and obviously do reflect those universal tendencies in particular language-pair-bound areas of grammar. The types of translationese described in this paper are the following:

1. simplification of TL choices means that high-frequency grammatical/syntactic resources tend to be preferred as translation solutions at the expense of other TL possibilities, e.g. quantifiers in case study I.
2. SL-specific interference in translated Spanish refers to grammatical and/or syntactic uses that have been ‘borrowed’ from English and that are not corroborated by corpus data of original Spanish, e.g. pre-modifying adjectives in case study II, and
3. ‘unique grammatical feature’ is used to identify grammatical/syntactic uses which according to corpus data are exclusive of the TL, Spanish in the case of our language pair, e.g. the ‘perfective imperfect’ in case study III.

Our claim is that these three characteristics of translated language (Spanish), when considered against the corpus-based results of original language (English), can be useful to measure language correctness and sophistication in translated texts and therefore can be seen as a tool to help assess translation quality.

3. Why use corpora to assess translations

There is a well documented literature of the uses of corpora in translation-related endeavors, among others Baker 1996, 1998, 1999, 2004; Kenny 2001; Granger et al. 2003; Zanettin et al. 2003; Santos 2004; Lewandowska-Tomaszczyk 2004, Mauranen 2000, 2004, etc. Some of the reasons for this interest are the quick access to empirical evidence and the immediate feedback that may be obtained, taking into account that the usefulness of corpora is greatly dependent on the type of hypotheses that are going to be demonstrated.

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 of corpora. Whereas much of the work done has concentrated on building the most appropriate corpus for each specific case, 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), which is the final goal of all applied research. This limited exploitation of corpus-based research has important implications for TQA, which is essentially applied in nature. The present paper aims at filling this gap using a specific corpus-based contrastive methodology.

A further drawback is the unpredictability of the results of searches when the corpus user is an applied professional (i.e. a professional translator, a reviewer, etc.) (Wilkinson 2005). Some researchers have appropriately dubbed the process of looking

for applied information directly in the raw corpora ‘serendipity’ (Bernardini 2000); some even go as far as to show how to increase the likelihood of finding relevant information (Bowker and Pearson 2002: 200-202). Recent work is trying to end this state of affairs. Most of the applied proposals address evaluation needs in translator education and in the broader curriculum of the prospective language service providers (Zanettin et al. 2003: 1).

Bowker (2001) has put forward one of the most articulated and realistic proposals to date. Her evaluation corpus is conceived specifically for specialized translation and is organized in a flexible way, making it a really 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 this evaluation corpus does not seem to travel well into other educational contexts. Can teachers/researchers/reviewers afford to devote time to building expert evaluation corpora (Varantola 2003)? Will the benefits of building them and using them exceed the effort of tool-building, or will they not? Why should not a service provider expect to be supplied with tools to do his/her job straight away? Are translation reviewers familiar enough with corpora to correct and improve translations?

Corpora, of whichever type, do not provide answers and/or solutions to their intended users; thus 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 pivotal translationese areas in a given language pair and a given direction.

There is the possibility to use already existing corpora which can be further exploited in combination with other resources for a variety of intended applied goals. In

other words, we do not think it is necessary to compile 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 use corpora in order to corroborate disparities between original and translated language by focusing on three ‘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: a) quantifiers reveal a tendency to simplification in the different distribution of choices when considered in translated Spanish as compared to original language, b) in nominal characterization the data shows the overuse of some of the ‘dispreferred options’ available in the target language, which suggests that there is interference, and c) one of the more salient and idiomatic meaning encoding capabilities of the Spanish *imperfecto* ‘imperfect’ –unique feature - are simply missed when translating Simple Past forms.

4. Data: combining comparable and parallel corpora

In recent times, a considerable amount of research has focused on the various aspects of translation studies using different types of corpora as a source of data (Bowker et al. 1998, Laviosa 1998, 2003, Olohan 2004). Some pieces of research are clearly aimed towards translator training (Bernardini and Zanettin 2000), whereas others analyze the features of translated language as opposed to spontaneously produced language (Baker 2001, Laviosa 1998) or, as already mentioned, issues related to translation quality assessment (Bowker 2001). Depending on the aim, some of these

studies make use of comparable corpora - “original texts in each language, matched as far as possible in terms of text type, subject matter and communicative function” (Altenberg and Granger 2002: 7-8) and others make use of parallel or translation corpora, which consist of “original texts in one language and their translations into one or several other languages” (Altenberg and Granger 2002: 8). In addition, some other studies use both corpora at the same time, as in the case of the English-Norwegian Parallel Corpus (ENPC) (Johansson 1998: 2003).

The research reported here is based on the combined use of three different corpora:

1. Cobuild’s Bank of English³, a large general language monolingual corpus of contemporary English,
2. CREA⁴ - a large general language monolingual corpus of contemporary Spanish. Cobuild’s Bank of English and CREA are used in a joint way as a comparable corpus. We acknowledge the fact that total comparability is difficult to achieve (Laviosa 1997), but the degree of comparability in this case was considered sufficient for our purposes.
3. and the ACTRES parallel corpus of English original texts and their corresponding Spanish translations (P-ACTRES)⁵, which is being compiled at the University of León (Spain).

Both monolingual source corpora (i.e. large corpora from which smaller, phenomenon-specific corpora can be extracted) include over one-hundred million words of running text each and have a similar internal structure concerning intralinguistic varieties, register distribution, mode and statistical dimensions. Each of the two corpora

acts as a source corpus, since restrictive choices have been made concerning language variety, mode and size. For convenience, the varieties chosen are UK English and European Spanish. Because of its applied aim, the mode is written. Books, magazines, newspapers and ephemera were the subcorpora chosen; the total number of words used for this paper amounts to slightly over 30 million words in each language. 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 (see below).

P-ACTRES mirrors the qualitative construction criteria of both the Bank of English and CREA, i.e. subcorpora, register distribution, mode, etc. It differs from them in two respects: instead of being a complete text corpus, P-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 English language materials are not restricted to materials produced in the UK, as choice of SL variety was deemed to be irrelevant when the directionality is from English into Spanish. Since the aim of this paper is to carry out translation quality assessment of English texts translated into Spanish, the diatopic variety of English used in the source text is not considered a discriminating factor. P-ACTRES is open corpus and contains over 2 million words, evenly distributed between the two languages. This allows for studies that are representative of the translation phenomenon between English and Spanish, on the one hand, and it provides material for studies comparing original and translated 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 ‘source corpus’ from which to extract different ‘sample corpora’:

a) a traditional approach is taking a random portion of materials as a sample corpus and search for item ‘x’ – case studies 1 and 2 below.

b) another strategy frequently used is selecting a hundred random text pairs focusing on the grammatical phenomenon being analyzed (past tense, modal verbs, etc) – case study 3.

As corpus management tools P-ACTRES uses the Translation Corpus Aligner (TCA) for sentence alignment (Hofland and Johansson 1998) and the Translation Corpus Explorer (WebTCE) as a browser (Ebeling 1998), developed and constantly refined in Norway for the English-Norwegian Parallel Corpus Project.

5. Method and procedure

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

First, empirical data are extracted from the two monolingual comparable corpora – Cobuild and CREA - on the basis of cross-linguistic similarity perception, and analyzed following the sequence: selection, description, juxtaposition and contrast. The *tertium comparationis* is set up at the descriptive stage and consists of semantic cross-linguistic labels relevant for our language pair, e.g. [TH] for ‘temporary habit’, (Rabadán 2005), ‘descriptive’ (Ramón García 2003), etc. 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 for in the parallel corpus in order to obtain a diagnostic sample of the rendering of a particular grammatical feature into the target language; this provides a list of the actual translational solutions taken for the different uses of the formal structure analyzed. 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 evidence – original Spanish from CREA - with the diagnostic data obtained from P-ACTRES. This allows us to identify differences between original and translated Spanish, which may be due to a particular norm of translation (Toury 1995, Chesterman 1998, Schäffner 1999), to the influence of the SL (Toury 1995: 275; Mauranen 2004), to universal features of translated language (Baker 1993), or simply to incompetent translating. This intralinguistic contrast will eventually highlight the differences between the grammar of original and translated Spanish and the extent to which translationese applies.

6. Case studies

Case studies I and II make use of the comparable monolingual corpora together with a small P-ACTRES sample as a diagnostic tool to obtain examples of translations. This parallel corpus has been aligned on sentence level using the Translation Corpus Aligner. It contains 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 of this sample parallel corpus:

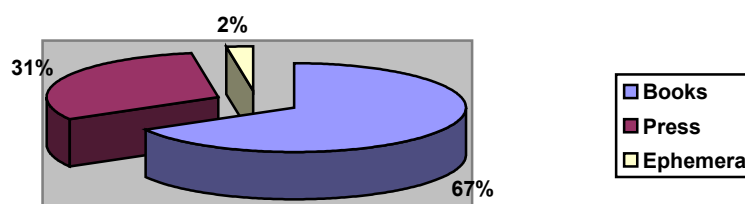


Figure 1: Register distribution of the P-ACTRES sample.

The following table summarizes the number of words included in each subsection, and for each language English and Spanish:

Table 1: Number of words in the P-ACTRES sample⁶.

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

6.1. Case Study I: Intensified quantification.

In a first large-scale contrastive study on quantification (Labrador de la Cruz 2005), 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 (1985), Downing and Locke (1992), Berry (1997) and Biber et al. (1999) and Bello (1981), Alarcos (1994),

Matte Bonn (1995) and Bosque and Demonte (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 CREA; only those subcorpora that represent British English and European Spanish were consulted.

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.

Taking into account the large size of our two subcorpora, the population of concordances to study was generally too large; as a consequence, it was necessary to take a sample of a reduced but still sufficiently representative number of occurrences for each quantifier. However, the frequency rates varied considerably among the different quantifiers, and so it was not possible to study a fixed number of occurrences for all quantifiers. Taking 300 out of 90,000 did not seem to be as equally representative as taking 300 out of 500, for instance. The following statistical formula was applied in order to ascertain how many concordances should be analyzed in each case: $n = N / ((N - 1)E^2 + 1)$ where 'n' is the sample to be analyzed, 'N' the population, i.e., the total number of occurrences yielded by our searches, and 'E' (0.05) for an estimative error of 5%. Thus, for example, the word *none*, which occurs 3,029 times in COBUILD, was studied in 353 of its 3,029 cases. Finally, the total number of concordances to be analyzed 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 interlinguistic contrast between the formal realization 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 intensified.

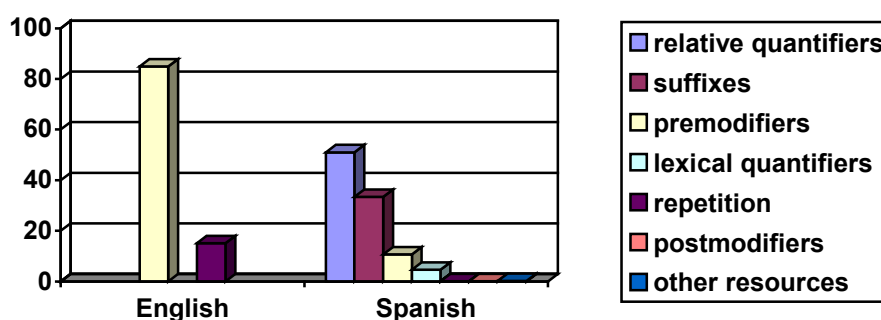


Fig. 2: Intensified quantification in English and Spanish. Comparable data.

As can be seen in Figure 2, English mainly makes use of premodification, - 84.96% of the occurrences, as shown in examples (1):

(1) *I think that's causing quite a lot of concern; We have been through so much together, we will always be friends*

and secondly, repetition is used with this purpose – 15.03%, as in examples (2):

(2) *He is picking many many places where he wants to move; I went off and did loads and loads of interviews.*

Spanish also uses premodifiers (3) and repetition of quantifiers (4) to intensify quantification:

(3) *Esta vez arrancó echando un buen montón de humo y aceite a la cara de Paco,*

‘This time it started by puffing a fair amount of smoke and oil into to Paco’s face.’

(4) *es digna de mucha, mucha consideración,*

‘he/she deserves much, much consideration’

However, these resources occupy lower positions in the rank scale –
premodification: 10.66% of the occurrences and repetition, as seldom as 0.08%.

Relative quantifiers (5) and suffixes (6) are the main formal devices used to express intensified quantification in Spanish, with percentages of 51.02% and 33.39% respectively:

(5) *La existencia de tantos sistemas añade nuevas dificultades,*

‘The existence of so many systems adds new difficulties.’

(6) *Sin embargo, tardó poquísimo en volver.*

‘However, he was back in no time.’

Other minor resources found are postmodification (7) (with a frequency rate of 0.04%), and lexical quantifiers (8) (with a rate of 4.64%):

(7) *que hay un montón exorbitante a eso nadie le pone reparo*

‘the fact that there is an exorbitant amount is something no one objects to’

(8) *aquel maletín parecía de suma importancia para él.*

‘that briefcase seemed of great importance to him’

With such a divergence in the English-Spanish contrast of the formal representation of this function (intensified 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 could be attributed to the influence of English grammar on Spanish translations.

The analysis of the sample parallel corpus reveals a higher than usual rate of premodification and repetition- typical resources of the English language- to the detriment of other more idiomatic ways of intensified 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).

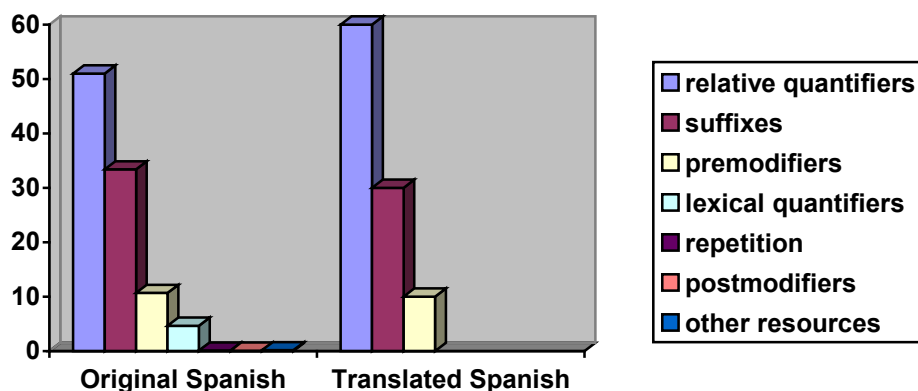


Fig. 3: Intensified quantification in Spanish originals and translations. Diagnostic data.

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 quantification in English; on the contrary, they

stick to a rather limited series of idiomatic and natural resources in Spanish for keeping the same function across the two languages. One of 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.

While this behavior guarantees correction, it plays in detriment of the wealth of resources offered by the target language. It seems the ‘**simplification hypothesis**’ is at play in reducing the range of options and thus narrowing the inventory available in translated Spanish.

6.2. Case Study II: Nominal Characterization

The modification of nouns within the boundaries of the noun phrase (NP) 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 (not oral texts) from 1990 onwards and in the European varieties of English and Spanish were used, amounting to slightly over 30 million words in each case.

Bearing in mind that this contrastive study has taken a semantic function as the starting point – characterization -, and considering the fact that the use electronic corpora requires a formal input, a specific search strategy had to be devised in order to obtain relevant data for the analysis of nominal modification from the two corpora. The

solution taken was the use of a list of very common nouns in the two languages as entries for our corpora in order to analyze their syntactic environment in the search for instances of modification. This option is supported by the consideration that “semantically, (...) the noun appears to play the leading role and the predication, whether adjectival or verbal, is subordinated to it.” (Aarts and Calbert 1979: 137).

Frequency lists in the two languages were used in order to find the most common nouns in English and Spanish. The Cobuild corpus provides frequency lists of all parts of speech, and the ten most common nouns in English were selected for the study: *time, year, world, way, day, man, home, life, night, week*. The Spanish corpus CREA does not provide this type of information, which had to be gathered from other corpus-based sources for this language (Alameda and Cuetos 1995). The ten most common nouns in Spanish were also chosen for the analysis, irrespective of the fact that not all of them were referential equivalents of the English nouns: *vez, parte, tiempo, vida, caso, día, año, forma, mundo, momento* (‘instance’, ‘part’, ‘time’, ‘life’, ‘case’, ‘day’, ‘year’, ‘form’, ‘world’, ‘moment’). Curiously enough, seven out of the ten most common nouns in each language happen to be at least partial equivalents.

There are various additional frequency lists available for both English (British National Corpus⁷) and Spanish (Corpus del Español⁸) with slight differences in the ten most frequent nouns. The list obtained by Alameda and Cuetos (1995) was selected because it was based on a large corpus of mainly peninsular contemporary Spanish including a register distribution similar to the one present in CREA. The aim was not to carry out a lexical contrastive study, but rather reveal the links between syntax and semantics wherever a particular semantic function occurred, no matter what head noun was affected by the modification.

Using the statistical formula above (see 6.1), 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: *of/ de*-phrases and pre-modifying adjectives, where the divergences in use are significant.

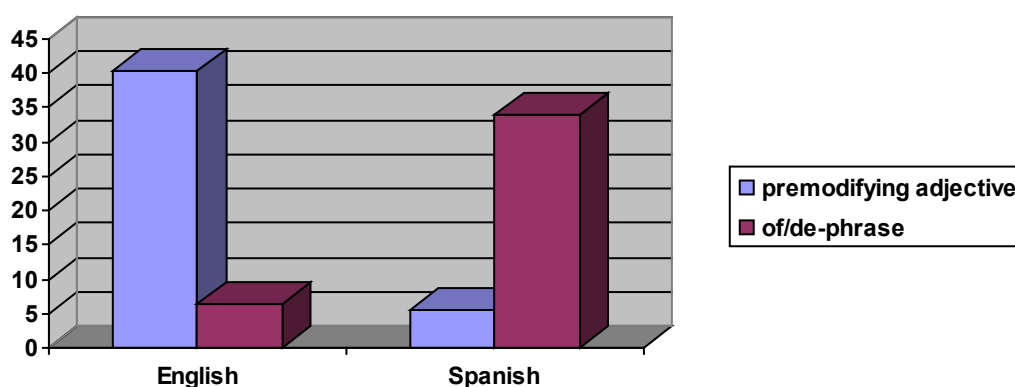


Fig. 4. De-phrases and pre-modifying adjectives with a descriptive function in English and in Spanish. Comparable data.

Fig. 4 illustrates that native speakers of English make a heavy use of premodifying adjectives with descriptive meanings (9) with about 40% of cases, whereas prepositional phrases headed by the preposition *of* (*of*-phrases) (10) occur only in slightly over 5% of cases with this meaning:

(9) a wonderful time, a great year

(10) 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 purely descriptive meanings, occurring in over 30% of instances:

(11) *el tiempo de la fiesta, un año de temperatura social elevada*

‘the time of the party’, ‘a year with great social agitation’

Premodifying adjectives are also an option in Spanish, but native speakers use them with descriptive meanings in only about 5% of cases:

(12) *su turbulenta vida, un buen momento*

‘his/her turbulent life’, ‘a good moment’

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 between the native and translated uses in Spanish.

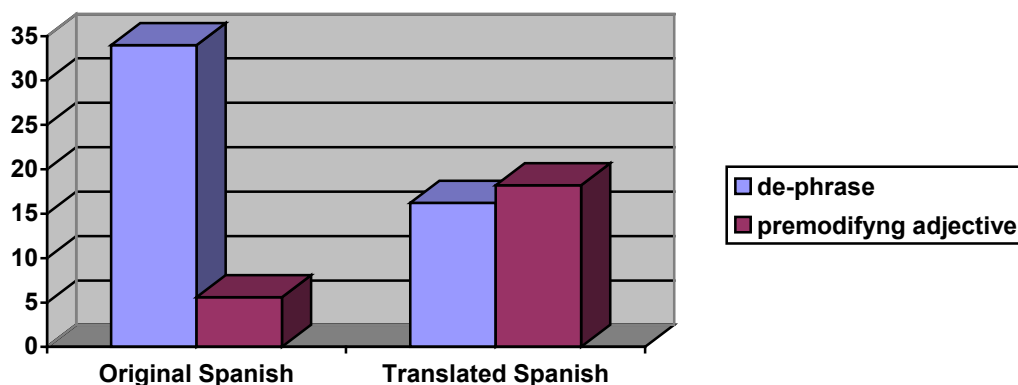


Fig. 5: De-phrases and pre-modifying adjectives with a descriptive function in original and translated Spanish. Diagnostic data.

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:

(13) *el pelo de un rojo intenso, un hombre de buen tamaño, individuos de aspecto enfermizo, un enchufe de inadecuada conducción eléctrica,*

‘deep-red hair’, ‘a well-built man’, ‘ill-looking individuals’, ‘a plug with deficient power supply’

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:

(14) *un grave problema, la extraña criatura, una enorme pirámide*

‘a serious problem’, ‘the strange creature’, ‘a huge pyramid’

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:

(15) *mala espina, puro teatro*

‘bad vibes’, ‘absolute sham’

However, translators clearly overuse pre-modifying adjectives with a descriptive meaning in translations from English into Spanish, leading to a high frequency of rather unidiomatic expressions such as:

(16) *la plateada criatura, este eficaz sistema, este notable informe*

‘the silvery creature’, ‘this efficient system’, ‘this important report’

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.

All this strongly suggests that the overuse of pre-modifying adjectives with a descriptive function in translated Spanish can be considered as symptomatic of **interference** when the SL is English. This particular feature is actually highly characteristic of Spanish translations from English, making them easily identifiable as such. The study has also quantified the overuse of pre-modifying adjectives and the underuse of *de*-phrases in translations from English into Spanish. These quantitative data suggest that figures in excess of the percentage typical of original Spanish may be used as a tool to evaluate translated texts. Hence, this would be an objective way to assess the quality of translations: the lower the discrepancy, the more similar the TT is to naturally occurring Spanish and, consequently, the higher the quality of the translation.

6.3. Case study III: The English Simple Past and the Spanish *imperfect/preterite* option.

The translation of the English Simple Past into Spanish is a typical problem area because of the different ways the grammars of each language handle the expression of past time. English offers an unmarked past form whereas Spanish requires an obligatory choice between the preterite (*pretérito*) and the imperfect tense (*imperfecto*).

As in the previous case studies, empirical data were obtained from three different corpora: the Bank of English, CREA and P-ACTRES. As the translation

problem stems from the Spanish part, we adopted a target-based perspective to start our search for corpus-based evidence of the uses of the preterite and the imperfect in this language using a frequency list from the BDS⁹ to establish the input forms. Ten high-frequency verbal lemmas were randomly singled out from the top 100 and used as input (Rabadán 2005) for separate searches of the Spanish preterite and the imperfect. As CREA still offers quite restrictive querying options, it proved necessary to search all inflected forms and make sure that person and number variation were duly represented in the sample. Our search forms yielded 41,483 occurrences for the preterite and 20,678 for the imperfect, totaling 62,161 cases in Spanish. After applying the statistical formula in 6.1 we ended up with a sample universe of 396 preterite cases and 392 for the imperfect. The procedure to extract the English-language data was determined by the size of the population in the Spanish part. We started by searching The Bank of English for Simple Past forms using ten high-frequency verbal lemmas as well. The output was much smaller than the combined outputs of the two searches in Spanish and a decision was made to go on adding top frequency querying nodes to our input list until a population size comparable to the Spanish one was reached. After searching 20 input items, we reached a population size of 62,108 cases of the English Simple Past. The sample was established at 397 cases.

The following step was to establish the cross-linguistic semantic labels that would function as *tertium comparationis* in the contrast. Drawing on the works by Leech (1987), Huddleston and Pullum (2002), García Fernández and Camus Bergareche (2004) and Rojo and Veiga (1999), among others, we ended up with the following semantic characterization: (a) ‘absolute past’ (i.e. past action/event, with an end-point requirement, Rojo and Veiga 1999), e. g.

(17) *Natasha came forward straight away to be filmed;*

(18) *En su viaje, el alcalde durmió en hoteles y comió en restaurantes, según propia confesión.*

‘In his journey the mayor spent the night in hotels and ate in restaurants, according to his own confession.’

(b) ‘anaphoric past’ (i.e. when the action/event is linked to another action, fact, event, situation, etc., and has no end-point requirement), as in (19) and (20) below:

(19) *An early shift meant he had to leave home at 4am, only returning 11 hours later;*

(20) *Mientras se leían, Rodolfo Martín Villa miraba hacia lo alto, hacia el cielo del hemiciclo,*

‘While they were being read, Rodolfo Martín Villa was looking upwards, at the top of the dome of the parliament’

(c) ‘past habit’ as in

(21) *If the correct combination of little fruit came up, you won; if not, you lost;*

(22) *Los tomaban a la brasa y, según los fósiles descubiertos por ahora, poco hechos*

‘they ate them roasted, and according to the fossils found up to now, rare’

(d) ‘hypothetical past’ as in

(23) *According to the story, Neil reckoned Ravanelli wasn't fit and could lose Middlesbrough the cup if he played at Wembley (see Rabadán 2005);*

(24) *La diputada Rosa Martí anunció en abril pasado que el PSC presentaría un recurso si se tomaba una decisión de este tipo*

‘the MP Rosa Martí announced last April that the PSC would present an appeal if a decision of this type was taken’

(e) ‘progressive’

(25) *El conjunto manresano, que se presentaba ante su afición, perdía en el descanso por 30-37*

‘the Manresan team, playing in front of their followers, was losing 30-27 at half-time’

(f) ‘irrealis’

(26) *Yo creía que esa señora estaba ya enterrada.*

‘I thought that this lady had already been buried.’

Although ‘absolute past’ is generally associated with the preterite, the Spanish imperfect is also able to convey this meaning 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 example (27):

(27) *La Voz de Valencia, Diario de tendencia derechista, próximo a Calvo Sotelo, aparecía el 3 de agosto controlado por Esquerra Republicana.*

‘*La Voz de Valencia*, a right-wing newspaper close to Calvo Sotelo, was published on August the 3rd under the control of Esquerra Republicana.’

This use is generally referred to in the literature as *perfective imperfect* and is seen to be equivalent to a preterite. In our analysis, however, the semantic criteria prevail and this function has been considered as ‘absolute past’.

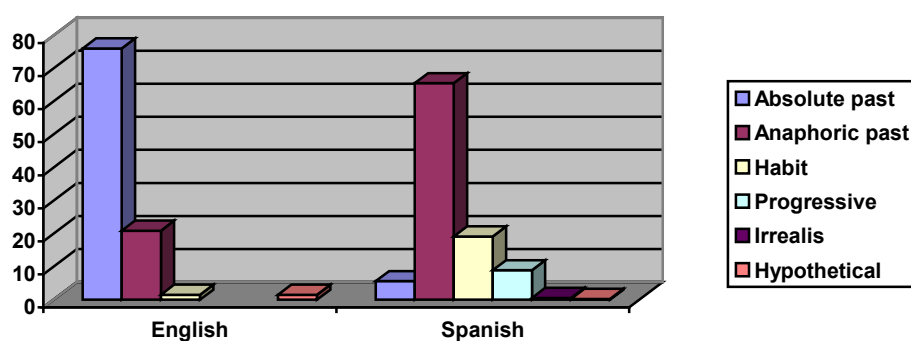


Fig 6. Semantic functions of the English Simple Past and the Spanish Imperfect. Comparable data.

A full-scale inquiry into the translation possibilities of the English Simple Past into Spanish (Rabadán 2005) has yielded the following results (see Fig. 6). The function ‘absolute past’ presents 76.07% of all cases analyzed in English, followed by the ‘anaphoric past’, which comes to 20.9%. ‘Habit’ and ‘hypothetical past’ represent 1.5% of the cases each. There is no evidence in the English language sample of neither ‘progressive’ nor ‘irrealis’ examples. In Spanish, the *preterite* stands for the ‘absolute past’ in all cases recorded (100%), which makes this tense unproblematic and therefore uninteresting for our purposes here. The imperfect, however, covers a much wider range of meanings: ‘Absolute past’ comes just to 5.61% of cases in native usage of the imperfect, whereas ‘anaphoric past’ is the meaning of 65.56% of the cases, followed by a string of other well represented functions such as ‘habit’ 19.13%, ‘progressive’ (8.93%), ‘irrealis’ (0.51%) and ‘hypothetical past’ (0.25%).

A second sampling strategy has been used in this case study. It consisted in selecting 100 random pairs from P-ACTRES containing at least once 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.

Our diagnostic data reveal a radical departure from native usage in the translation solutions chosen (see fig. 7). Except for ‘absolute past’, all the meanings identified in the diagnostic sample have been rendered by an imperfect or, on a few occasions, by other –generally lexical and phraseological- resources, as in example (28):

(28) *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.*

‘Anaphoric past’ is translated by an imperfect in 16% of cases, and ‘habit’ and ‘hypothetical past’ by 3.42% each. There is no evidence of other functions being translated by a Spanish imperfect tense. The most obvious discrepancy between native and translated choices is then the use of the imperfect in native Spanish meaning ‘absolute past’, as shown in fig. 7.

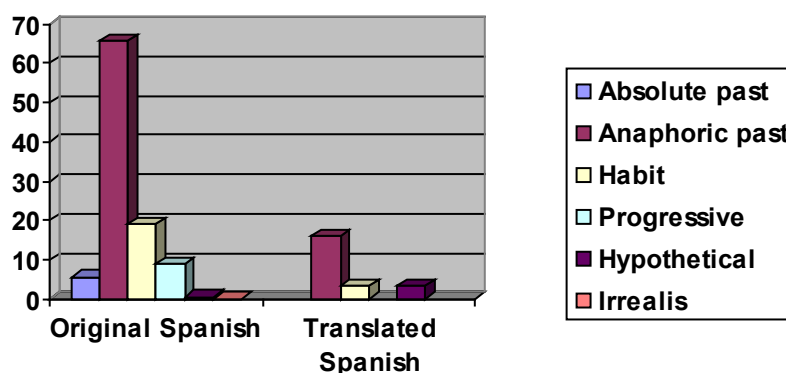


Fig. 7: Semantic functions of original and translated Spanish *imperfect*. Diagnostic data.

The results indicate that there is a raw underuse of the imperfect as a translation of the meaning ‘absolute past’. No evidence has been found of this meaning in translated Spanish, which seems to prove the usefulness of the ‘**unique grammatical features hypothesis**’ discussed earlier as a tool to provide empirical data in order 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.

7. Conclusions

Corpus-based studies like the ones presented in this paper provide three types of useful information: i) contrastive data using comparable corpora (comparable data), ii) descriptive translation data using parallel corpora (diagnostic data), and iii) ‘translationese’ in Spanish, comparing original usage with translated usage in the same language. All three areas have implications for translation practice, translator training and translation quality assessment. 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: intensified quantification, descriptive characterization, and the translation of the absolute past. These discrepancies have been typified by means of the following translation universal hypotheses:

1. The **simplification hypothesis**: In the case of intensified quantification it was found that translators tend to use the top high-frequency Spanish resources for expressing the same function in detriment of some less frequent resources available in the target language. This trend does not involve interference and is

totally acceptable in Spanish; however, it does not fully exploit the wide range of possibilities existing in the target language. This results in a lack of variety and a more homogeneous and uniform type of language.

2. **Interference:** The analysis of nominal characterization has revealed two clear instances of interference between the language pair English-Spanish. Single pre-modifying adjectives were found to occur four times more often in Spanish translations than in texts written originally in Spanish, and this can only be attributed to the influence of the SL English, where the unmarked position of adjectives is the premodifying one. On the other hand, Spanish translators seem not to exploit the potential of *de*-phrases with descriptive meanings, which occurred in only approximately half the times, when compared to original Spanish. Again, this difference can be attributed to the influence of English as the SL, since the formal equivalents – *of*-phrases – are relatively uncommon in this language. In fact, translation from a different SL would probably not yield these particular cases of interference, but others.
3. **The unique grammatical features hypothesis:** The corpus-based analysis of the past tenses has shown that there is at least one function of the Spanish imperfect that typically occurs in original rather than translated language. Original Spanish data indicate that the imperfect can actually be used to convey the meaning ‘absolute past’, whereas the parallel corpus data suggest that translators tend to choose a preterite when rendering this semantic function. This does not mean that the choice is incorrect, rather that the degree of specificity and even accuracy and communicative economy of the translated text is lower than those of the original.

We have argued and demonstrated here that there are several degrees of difference in the usage of the same resource between original and translated language:

- a) the translated language may **overuse** a particular formal resource
- b) the translated language may **underuse** a particular formal resource
- c) the translated language may **lack** a particular formal resource
- d) the translated language may present a **similar** frequency of occurrence of a particular formal resource.

We claim that these differences may be quantified to a certain extent and that, combined with other tools, they can contribute to the systematization and objectivization of translation assessment.

In order to use these results as TQA tools they have to be conceptualized in some way (Rabadán 2007). As they are, they combine both quantitative and qualitative findings which we believe can be used to advantage to evaluate non-specialized translated texts. Our proposal, tentative for the time being, is to fashion them into low-level language-pair specific conditioned statements inspired by Toury's formulation of general laws of translation (cf. Toury 2004: 25-28), as in

- (a) The lower the number of formal options chosen from those available in Spanish to translate intensified quantification, the higher the degree of simplification and the less accurate the translation, and vice versa;
- (b) The lower the number of *de*-phrases/the higher the number of pre-modifying adjectives in translated Spanish, the higher the degree of interference and the less idiomatic the translation, and vice versa;
- (c) The lower the number of instances of imperfect tenses meaning 'absolute past' in the translated text, the higher the degree of under-representation of

Spanish specific grammatical resources and the less acceptable the translation, and vice versa;

(d) The smaller the disparity between native and translated usage in the use of particular grammatical structures associated with specific meanings, the higher the translation rates for quality.

Contrastive work on further problem areas for our language pair will hopefully yield data leading to the formulation of more (and more refined) statements of the type shown above. The more grammatical features are made available as potential assessment tools, the higher the discriminatory power when evaluating. Having more criteria will also increase the usefulness and usability of tools built on these empirically-based data. There are obviously other factors that intervene in the quality of a given translation and that have to be taken into account at more sophisticated levels of analysis. However, the most tangible, objective and widely accepted criteria seem to be language correctness and acceptability, which of course embodies grammatical correctness and semantic and pragmatic appropriateness. Work in progress aims at developing an empirically-based application aimed at translation reviewers and other language service providers, which would be used ideally in conjunction with other assessment tools.

Notes

¹ Research for this paper has been funded by the research grant HUM2005-01215/FILO from the Spanish Ministry of Education. A shorter version focusing on slightly different issues was presented at the AAACL/ICAME conference; Ann Arbor May 2005. A number of issues raised there helped clarify our study. We are grateful to Roda Roberts and Lynne Bowker for their comments.

² ACTRES stands for *Análisis Contrastivo y Traducción Inglés-Español* (Contrastive Analysis and Translation English-Spanish) <http://actres.unileon.es>.

³ Cobuild's Bank of English <http://www.collins.co.uk/books.aspx?group=153>

⁴ CREA (Corpus de Referencia del Español Actual <http://corpus.rae.es/creanet.html>

⁵ The authors of this paper are grateful to Knut Hofland for his help.

⁶ 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.

⁷ URL: <http://www.natcorp.ox.ac.uk/>

⁸ URL: <http://www.corpusdelespanol.org/>

⁹ <http://www.bds.usc.es/> We are grateful to Guillermo Rojo (RAE and University of Santiago de Compostela) for making this list available to us. Personal communication: 28/09/2004.

References

Aarts, Jan M.G. and Joseph P. Calbert. 1979. *Metaphor and Non-metaphor. The Semantics of Adjective-Noun Combinations*. Tübingen: Max Niemeyer. xii + 240 pp.

Alameda, José Ramón and Fernando Cuetos Vega. 1995. *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: University of Oviedo. 2 vols.

Alarcos, Emilio. 1994. *Gramática de la lengua española*. Madrid: Espasa Calpe. 406 pp.

- Altenberg, Bengt and Sylviane Granger. 2002. "Recent trends in cross-linguistic lexical studies." In *Lexis in Contrast. Corpus-based Approaches* ed. by Bengt Altenberg and Sylviane Granger. 3-48. Amsterdam/Philadelphia: John Benjamins.
- Baker, Mona. 1993. "Corpus Linguistics and Translation Studies." In *Text and Technology. In Honour of John Sinclair*, ed. by Mona Baker et al. 233-250. Amsterdam/Philadelphia: John Benjamins.
- Baker, Mona. 1996. "Corpus-based Translation Studies: The challenges that lie ahead." In *Terminology, LSP and Translation*, ed. by Harold L. Somers. 175-186. Amsterdam/Philadelphia: John Benjamins.
- Baker, Mona. 1998. "Réexplorer la langue de la traduction: une approche par corpus." *Meta* 43: 4. 480-485.
- Baker, Mona. 1999. "The role of corpora in investigating the linguistic behaviour of professional translators." *International Journal of Corpus Linguistics* 4: 2. 281-298.
- Baker, Mona. 2001. "Investigating the language of translation: a corpus-based approach." In *Pathways of Translation Studies*, ed. by Purificación Fernández Nistal and José María Bravo Gozalo. 47-56. Valladolid: University of Valladolid.
- Baker, Mona. 2004. "A corpus-based view of similarity and difference in translation." *International Journal of Corpus Linguistics* 9: 2. 167-193.
- Bello, Andrés. 1981. *Gramática de la lengua castellana*. Tenerife: Instituto Universitario de Lingüística. 267 pp.
- Bernardini, Silvia and Federico Zanettin, eds. 2000. *I corpora nella didattica della traduzione. Corpus Use and Learning to Translate*. Bologna: Clueb. 196 pp.

- Bernardini, Silvia. 2000. "Systematising serendipity: Proposals for concordancing large corpora with language learners." In *Rethinking Language Pedagogy From A Corpus Perspective: Papers From The Third International Conference On Teaching And Language Corpora*, ed. by Lou Burnard and Anthony M. McEnery. 183-190. Frankfurt am Main: Peter Lang.
- Berry, Roger. 1997. *Determiners and Quantifiers*. Glasgow: Collins Cobuild. vi + 186 pp.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman. xxviii + 1204 pp.
- Bondarko, Aleksander. 1991. *Functional Grammar: A Field Approach*. Amsterdam/Philadelphia: Benjamins. 207 pp.
- Bosque, Ignacio and Violeta Demonte, eds. 1999. *Gramática descriptiva de la lengua española*. Madrid: Espasa. 3 vol. cxv + 5351 pp.
- Bowker, Lynne, Michael Cronin, Dorothy Kenny and Jennifer Pearson, eds. 1998. *Unity in Diversity? Current Trends in Translation Studies*. Manchester: St. Jerome. xii + 196 pp.
- Bowker, Lynne. 2001. "Towards a methodology for a corpus-based approach to translation evaluation." *Meta* 46: 2. 345-364.
- Bowker, Lynne and Jennifer Pearson. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge. xiv + 242 pp.
- Chesterman, Andrew. 1998. *Contrastive Functional Analysis*. Amsterdam/Philadelphia: John Benjamins. vii + 230 pp.

- Downing, Angela and Philip Locke. 1992. *A University Course in English Grammar*. New York/ London: Prentice-Hall International. xvii + 652 pp.
- Ebeling, J. 1998. "The Translation Corpus Explorer: A browser for parallel texts". In *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*, ed. by Stig Johansson and Signe Oksefjell. 101-112. Amsterdam: Rodopi.
- García Fernández, Luis and Bruno Camus Bergareche, eds. 2004. *El pretérito imperfecto*. Madrid: Gredos. 615 pp.
- Granger, Sylviane, Jacques Lerot and Stephanie Petch-Tyson, eds. 2003. *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi. 219 pp.
- Hofland, Knut and Stig Johansson. 1998. "The Translation Corpus Aligner: A program for automatic alignment of parallel texts." In *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*, ed. by Stig Johansson and Signe Oksefjell. 87-100. Amsterdam: Rodopi.
- House, Juliane. 2001. "How do we know when a translation is good?" In *Exploring Translation and Multilingual Text Production: Beyond Content*. ed. by Erich Steiner and Collin Yallop. 127-160. Berlin/NY: Mouton de Gruyter.
- Huddleston, Rodney D. and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge, UK: Cambridge University Press.
- Johansson, Stig. 1998. "On the role of corpora in cross-linguistic research." In *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*, ed. by Stig Johansson and Signe Oksefjell. 3-24. Amsterdam: Rodopi.
- Kenny, Dorothy. 2001. *Lexis and Creativity in Translation: A Corpus-based Study*. Manchester: St. Jerome. 254 pp.

- Labrador de la Cruz, Belén. 2005. *Estudio contrastivo de la cuantificación inglés-español*. León: University of León. xiv + 553 pp.
- Laviosa, Sara. 1996. "Comparable corpora: Towards a corpus linguistic methodology for the empirical study of translation." In *Translation and Meaning, Part 3*, ed. by Marcel Thelen and Barbara Lewandowska-Tomaszczyk. 153-163. Maastricht: Hogeschool Maastricht.
- Laviosa, Sara. 1997. "How comparable can 'comparable corpora' be?" *Target* 9: 2. 289-319.
- Laviosa, Sara. 1998. "The English Comparable Corpus: A resource and a methodology." In *Unity in Diversity? Current Trends in Translation Studies*, ed. by Lynne Bowker et al. 101-112. Manchester: St. Jerome.
- Laviosa, Sara. 2003. "Corpora and the translator." In *Computers and Translation: A Translator's Guide*, ed. by Harold L. Somers. 105-117. Amsterdam/Philadelphia: John Benjamins.
- Leech, Geoffrey N. 1987. *Meaning and the English Verb*. London/NY: Longman. viii + 139 pp.
- Matte Bonn, Francisco. 1995. *Gramática comunicativa del español*. Madrid: Edelsa. 2 vol.
- Mauranen, Anna. 1999. "Will 'translationese' ruin a contrastive study?" *Languages in Contrast* 2: 2. 161-185.
- Mauranen, Anna. 2000. "Strange strings in translated language: A study on corpora." In *Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects*, ed. by Maeve Olohan. 119-141. Manchester: St. Jerome.

- Mauranen, Anna. 2004. "Corpora, universals and interference." In *Translation Universals- Do they exist?*, ed. by Anna Mauranen and Pekka Kujamäki. 65-82. Amsterdam/ Philadelphia: John Benjamins.
- Newmark, Peter. 1981. *Approaches to Translation*. London: Pearson Educational. 200 pp.
- Olohan, Maeve and Mona Baker. 2000. "Reporting *that* in translated English: Evidence for subliminal processes of explicitation?" *Across Languages and Cultures* 1: 2. 41-158.
- Olohan, Maeve. 2004. *Introducing Corpora in Translation Studies*. London: Routledge. vii + 222 pp.
- Quirk, Randolph, Sydney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman. x + 1779 pp.
- Rabadán, Rosa, Belén Labrador and Noelia Ramón. 2004. "English-Spanish corpus-based contrastive analysis: Translational applications from the ACTRES project." In *Practical Applications in Language and Computers. PALC 2003* ed. by Barbara Lewandowska-Tomaszczyk. 141-151. Frankfurt am Main: Peter Lang.
- Rabadán, Rosa. 2005. "The applicability of description: Empirical research and translation tools." *Contemporary Problematics of Translation Studies. Revista Canaria de Estudios Ingleses* 51. 51-70.
- Rabadán, Rosa. 2007. "Divisions, description and applications. The interface between DTS, corpus-based research and contrastive analysis." In *Translation Studies:*

- Doubts and Directions*, ed. by Yves Gambier et al. 237-252. Amsterdam/Philadelphia: John Benjamins.
- Ramón García, Noelia. 2003. *Estudio contrastivo inglés-español de la caracterización de sustantivos*. León: University of León. xii + 439 pp.
- Reiss, Katharina. 1971. *Möglichkeiten und Grenzen der Übersetzungskritik*. München: Hueber. 124 pp.
- Rojo, Guillermo and Alexandre Veiga. 1999. "El tiempo verbal. Los tiempos simples." In *Gramática descriptiva de la lengua española*, ed. by Ignacio Bosque and Violeta Demonte. 2867-2934. Madrid: Espasa.
- Santos, Diana. 1995. "On grammatical translationese." In *Proceedings of the 10th Nordic Conference of Computational Linguistics, NODALIDA-95, Helsinki 29-30 May 1995*, ed. by Kommi Koskenniemi. 59-66. Helsinki: University of Helsinki.
- Santos, Diana. 2004. *Translation-Based Corpus Studies. Contrasting English and Portuguese Tense and Aspect Systems*. Amsterdam: Rodopi. ix +173 pp.
- Schäffner, Christina, ed. 1999. *Translation and Norms*. Clevedon, England: Multilingual Matters. 140 pp.
- 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?" In *Translation Universals- Do they exist?*, ed. by Anna Mauranen and Pekka Kujamäki. 177-186. Amsterdam/ Philadelphia: John Benjamins.

- Toury, Gideon. 1980. "Interlanguage and its manifestations in translation." In *In Search of a Theory of Translation*, ed. by Gideon Toury. 71-78. Tel Aviv: The Porter Institute for Poetics and Semiotics.
- Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins. viii + 311 pp.
- Toury, Gideon. 2004. "Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals?" In *Translation Universals- Do they exist?*, ed. by Anna Mauranen and Pekka Kujamäki. 15-32. Amsterdam/Philadelphia: John Benjamins.
- Tymoczko, Maria. 1998. "Computerized corpora and the future of Translation Studies." *Meta* 43: 4. 653-659.
- Varantola, Krista. 2003. "Translators and disposable corpora." In *Corpora in Translator Education*, ed. by Federico Zanettin et al. 55-70. Manchester: St Jerome.
- Wilkinson, Michael. 2005. "Using a specialized corpus to improve translation quality." *Translation Journal* 9: 3. <http://accurarapid.com/journal/33corpus.htm> (12/01/06)
- Zanettin, Federico, Silvia Bernardini and Dominic Stewart. 2003. "Corpora in Translator Education: An Introduction." In *Corpora in Translator Education* ed. by Federico Zanettin et al. 1-14. Manchester: St Jerome.