

Corpus Linguistics for English-Spanish Translation Quality Assessment (TQA): A Case of Applied Corpus Research (ACTRES)*

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

1. INTRODUCTION

This paper sets out to explore the possibilities of designing an effective and efficient tool to contribute to the assessment of translations by using a limited number of language-pair-bound descriptive anchor phenomena, more specifically grammatical, not lexical anchor phenomena. Our proposal applies to narrative texts, although the working hypothesis is that the procedure is useful for any text type provided the data are relevant and appropriate for other textual varieties.

The proposal consists of three parts: the first focuses on tool building, on how to obtain descriptive data which are relevant for our applied aim; the second deals with the verification of applicability of these results and the third presents a prototype of a semi-automatic application. Basically this means that the procedure needs to be useful and usable, that is, it must identify the descriptive knowledge that is relevant for the applied needs it is meant to serve, and, its intended users must find the tools efficient and effective for their needs.

2. APPLICABILITY: USEFULNESS AND USABILITY

Usefulness is a performance indicator associated with the extent to which tools (technological, conceptual or otherwise) are relevant to the actual needs of a user. When research has an applied goal, not every phenomenon that is interesting from a descriptive point of view is necessarily relevant, but those that tend to be associated with frequent problems in cross-linguistic practice are.

The first requirement of an effective and efficient application is its *usability* -it is imperative to show how descriptive findings can work as an efficient tool for applied purposes. Unlike the usual corpus-based descriptive work, which is not directly usable, in this proposal the procedure, the conceptual tools, and the way to interpret the results must be made available to their final users and be usable (Rabadán 2008). In other words, this is meant to be a contribution to applied TS.

3. THE APPLICABILITY OF DESCRIPTION: ANCHOR PHENOMENA

When examined from a cross-linguistic analytical perspective, certain grammatical areas show clear differences in the meanings some of the resources can convey in each of the languages. Empirical data demonstrate that dissimilarity in the way(s) grammatical meanings are conveyed is a constant source of cross-linguistic problems affecting both text processing and production. These features and their unnecessary transference into the other language also mark the difference between translated and non-translated language. These language-specific associations between grammatical meaning and formal resource can be seen as “anchor

* **Acknowledgement**

Research for this article has been undertaken as part of the ACTRES program, partly funded by the Ministry of Education and ERDF [HUM2005-01215]. The acronym stands for *Análisis contrastivo y traducción English-Spanish* “Contrastive analysis and translation English-Spanish”. Dr Ramón- Ángel Fernández is responsible for the computerized application.

phenomena” and can be used as key indicators of the degree of success in cross-linguistic transfer. A performance which is close to non-translated usage of grammatical features would then rate higher for discourse and language quality than another which departs from it.

“Anchor” is used here in the same sense as in “anchor words”: in a parallel corpus, the anchor words are specific words that are defined for the two languages involved and that are related by some type of cross-linguistic equivalence. “Anchor phenomena” would then be those grammatical resources that are perceived as being cross-linguistically equivalent but that tend to and/or do convey partially divergent meanings, for example, the future in English and French (Celle 2005) or progressive forms in English and in Spanish (Rabadán 2005), among others. Since cross-linguistic grammatical meaning dissimilarity cannot be assumed to be the same for different language combinations or in each direction, the form-meaning associations that qualify as anchor phenomena also differ by direction and language combination.

Corpus-based contrastive findings for the language pair English-Spanish show that there are a number of grammatical phenomena, more specifically 6, that rate high for usefulness in this particular applied goal: subject pronouns (1st person sing, yo and 3rd person sing. él), lexical verbs with modal meanings (poder), progressive forms in Spanish, combinations of adj + noun, the definite article (los/las in the plural) and a series of chosen adverbs ending in – mente.

For reasons of effectiveness, in this presentation we will comment on the main features of some of these phenomena. In section 5, we will briefly review the relevance of these corpus-based contrastive findings.

4. ANCHOR STAGE: TOOLS

Monolingual reference corpora in the two languages have been used as comparable corpora (Labrador 2005, Ramón 2003), and more recently a parallel corpus has been compiled containing original English texts and their corresponding Spanish translations, the P-ACTRES corpus. This aligned parallel corpus includes written material from a variety of different registers (fiction, non-fiction, press & miscellanea) published in English in the year 2000 or later, thus representing the contemporary stage of the English language, and translated for the Spanish readership.

4.1. CREA and P-ACTRES

CREA is a very large reference corpus sponsored by the *Real Academia Española* and includes around 175 million words of running texts in a wide range of different registers and geographical varieties of the Spanish language worldwide. On the other hand, P-ACTRES is an open corpus and currently contains over a million words in English with their corresponding translations into peninsular Spanish only. The English-Spanish parallel texts that have been aligned at sentence level and can be searched with the Corpus Work Bench browser (CWB)[†]. This aligned version[‡] contains over one million three-hundred thousand words, distributed among fiction (45.88%), non-fiction (30.23%) and press (newspapers with 13.83% and magazines with 10.04%) as follows:

P-ACTRES	English	Spanish	Total
Books – Fiction	396,462	421,065	817,527
Books – Non-Fiction	494,358	553,067	1,047,425
Newspapers	115,502	137,202	252,704

[†] We are grateful to Knut Hofland for his co-operation in the setting up of the P-ACTRES parallel corpus.

[‡] A small demo is currently available at: <http://actres.unileon.es>.

Magazines	119,604	126,989	246,593
Miscellanea	40,178	49,026	89,204
TOTAL	1,166,104	1,287,349	2,453,453

Table 1: Contents of the English-Spanish Parallel Corpus: number of words (June 2008)

4.2. *Tertium comparationis*: Cross-linguistic labels

A further (conceptual) tool which is necessary in the anchor building stage is the *tertium comparationis* (Krzyszowski 1990:15), which in this proposal consists of a set of cross-linguistic labels that function as the ‘model’ against which the degree of cross-linguistic match is measured. Anchor results, to be applicable, need to be both useful and usable. Our labels are useful for cross-linguistic meaning discrimination in the contrastive process (Rabadán 2005). Their role is to help identify the meaning features that are relevant for applied purposes. This means that general linguistics taxonomies are not necessarily useful here as they are geared primarily towards monolingual description. Application-oriented labels use information from any model of linguistic description and from different levels of analysis. This results in labels that show different statuses, since they account for grammatical, pragmatic, semantic, and even interlanguage information (Chesterman 1998: 27-40). In this study the labels have been set up for YO, ÉL and for PODER. The analysis of – MENTE adverbs and of PREMODIFYING ADJECTIVES are purely quantitative.

Labels for YO and ÉL have been arrived at drawing on Enríquez (1984), Luján (1999) and Fernández Soriano (1999) primarily, and they show a very irregular distribution depending on each particular form. Our choice has been to rework the information into the following labels/categories:

1. *Neutral*. This use corresponds to the basic deictic function and is always compulsory. This function is particularly relevant in the case of YO and TÚ (*tú y yo*).
2. *Emphatic*:
 - a. *Optional emphasis*. It refers to a surplus use of the pronoun which does not affect either deixis or anaphora, but that contributes pragmatic meanings such as ‘marker of formality’ (*usted*), focalization (*yo hago mi trabajo*), etc....
 - b. *Non-optional emphasis*. It is particularly relevant in the case of YO and ÉL (*soy yo, Teresa; ¿No fue él quien le pidió que me recibiera en su nombre?*).
 - c. *Contrastive*. This function is particularly relevant in the case of most pronouns. (*... o el matrimonio decidía salir y él se encargaba de aquella vigilancia...; ¡Eres tú quien debería estar aquí, no yo!*)
 - d. *Formulaic*. This function is particularly relevant in the case of TÚ (*vete tú a saber, allá tú.*)
 - e. *Narrative discourse marker*. A type of discourse reference marking which occurs when the explicit pronoun moves the narrative action forward. It is particularly relevant in the case of YO (*no os peleéis, que se lo cuento yo, y yo decía si no hace falta;*) and USTED and deserves to be considered separately.

Labels for PODER have been taken from Rabadán (2006), those for the progressive forms from Rabadán (2005). In both cases they had been arrived at after a process of language-pair bound contrastive work in which usability conditions were a priority. Premodifying adjectives (Ramón 2003) and -mente adverbs (Ramón and Labrador, forthcoming) have been shown to be useful for contrasting original and translated Spanish from a quantitative perspective.

4.3. Statistics

Statistics are useful when interpreting results. They can provide a welcome link between quantitative and qualitative empirical evidence as they help to focus on those uses or functions that trigger cross-linguistic problems. Yet, quantitative data by themselves do not supply applicable information. Results have to be filtered and their representativeness and suitability for the purposes of this study qualitatively assessed. This involves stating whether your results are statistically significant. Under the conditions of the descriptive data we are dealing with here, it is appropriate to use statistical ‘hypothesis testing for independent proportions’, and particularly two indicators: z-score and p-value. Both measure the difference between the data and what is expected under the null hypothesis (that both translated and non-translated grammatical usage is identical). Calculations have been done for a 95% confidence interval and a significance level of $\alpha = 0.05$.

5. METHOD

The working procedure followed in this paper is based on the combined use of data from a monolingual reference corpus in Spanish (CREA, Corpus de Referencia del Español Actual) and data from the P-ACTRES corpus, i.e., from Spanish translations of original English texts. The procedure is based on Krzeszowski’s model (1990) and comprises the following stages: selection and sampling; description, juxtaposition and contrast.

5.1. Selection and sampling

The subcorpus chosen has been in all cases ‘fiction’. The reason for this is that the use of personal pronouns in these texts is much higher than in non-fiction: for example, the unmarked 3rd person singular pronoun in original Spanish use comes to 55.70% for fiction vs. 30.18% non-fiction. A second reason for preferring fiction to other subcorpora is the fact that the results will be applicable to this particular text form, which accounts for both the usefulness and usability of the data.

5.2. ANCHOR PHENOMENA: SUBJECT PRONOUN *YO*, VERB *PODER* & *-MENTE* ADVERBS

For the pair and direction English→Spanish subject personal pronouns are a good candidate for ‘anchor phenomenon’. Roughly, pronouns have two main functions: deixis and anaphora (Huddleston and Pullum 2002: 1463-82); however, the actualization and distribution of these roles is different in English and in Spanish. English always shows a formal filler in the subject slot, whereas in Spanish subject pronouns are typically omitted in the deictic function, as the information related to person, number, and gender (the latter except in the 1st and 2nd person singular) is already included in the verbal inflections. However, Spanish grants subject pronouns a number of extra uses which add pragmatic meanings such as contrast, narrative marker, etc. (Marcos Marín 1978, Alarcos 1980, Fernández Soriano 1999, Luján 1999). It is on these uses that we are going to focus in order to examine the real usefulness of these features as an anchor for assessment.

The search has focused on singular personal pronouns *YO* and *ÉL* in roles other than those where their omission would lead to ungrammatical sequences. Contexts where the pronouns have potentially additional meanings were selected from both CREA and P-ACTRES.

PODER (lexical equivalent of *can*) is a lexical verb in Spanish which does not regularly take on modal functions other than those derived from its lexical meaning. With the exception of collocations and idiomatic constructions, all contexts were considered. Only the uses in the present indicative tense have been considered in the analysis. *PODER* cases were also obtained from both CREA and P-ACTRES.

Both corpora were sampled using simple random sampling to assure coverage and representativeness.

PRONOUN	POPULATION (N)	CASES STUDIED IN CREA	POPULATION (N)	CASES STUDIED IN P-ACTRES
yo	4,635	368	1,160	298
él	2,927	352	625	244

Table 2. Quantitative data for anchor candidates YO & ÉL

POPULATION (N)	CASES STUDIED IN CREA	POPULATION (N)	CASES STUDIED IN P-ACTRES
2,395	331	538	224

Table 3. Quantitative data for anchor candidate PODER

Because of its intended application this paper has focused exclusively on data from fiction texts, so the searches were restricted in both corpora to the fiction section. In the Spanish monolingual reference corpus CREA the 2000-onwards fiction section (novel) including texts only in European Spanish contains 2,157,056 words. The fiction section in P-ACTRES contains 421,065 words. In both cases the number of personal pronouns is very large, so a statistical formula was employed to extract only the necessary number of concordances for the analysis.

5.3. Description

Case study 1: FIRST PERSON SINGULAR PRONOUN: YO

The 1st person singular pronoun in Spanish *yo* tends to be omitted in most contexts when it is not emphatic due to its redundancy with respect to the verbal morphology indicating person and number. The analysis of the various functions of the pronoun *yo* in original Spanish texts and in translated texts yielded the following results:

CREA		P-ACTRES	
YO			
%	RAW CASES	RAW CASES	%
10.05%	37	62	20.80%
	<i>Neutral</i>		
21.46%	79	82	27.51%
	<i>Optional emphasis</i>		

4.07%	15	<i>Non-optional emphasis</i>	10	3.35%
49.18%	181	<i>Contrastive</i>	50	16.77%
5.43%	20	<i>Formulaic</i>	4	1.34%
9.78%	36	<i>Narrative discourse marker</i>	90	30.20%
100%	368	TOTALS	298	100%

Table 4 . Quantitative data for YO

It can be noted that the distribution of the various functions identified differs considerably between original and translated Spanish. However, being quantitatively different does not necessarily mean that the difference is significant in qualitative terms. To avoid this pitfall, we need to check whether these differences are statistically significant or are just a symptom (Rabadán 2007 in press). Two magnitudes, the z score and the p- value may help here. Both are obtained by calculating the hypothesis test for independent proportions. In both cases the confidence interval is 95% and the margin or error 5%. To be statistically significant the difference between both proportions (translated and non-translated cases) has to lie outside the curve ± 1.96 for the z-score and must be < 0.05 for the p-value. Applying statistical analyses, it was found that only some of the functions were actually significantly different, as shown in the following table:

YO	z-score	p-value
Contrast	8.74	0.00E
<i>Optional emphasis</i>	-1.81	.0698

Neutral	-3.88	.001
Narrative	-6.69	0.0025
Formulaic	2.82	.0048
Non-optional emphasis	0.49	6267

Table 5. Statistical significance of YO values

The main use of the 1st person pronoun in original texts is clearly a contrastive use to emphasize the difference between the 1st person speaker and some other person (2nd or 3rd): *él lo leía y yo le miraba de reojo; usted dijo horror y yo terror*. Curiously enough, this particular function was found in only 17% of cases in Spanish translations, indicating that other uses are favoured by translators.

The category termed as “neutral” here refers to cases where the use of the personal pronoun is obligatory for grammaticality, as in coordinated constructions: *tú y yo*). This function occurs slightly more often in translations than in originals.

The use of the 1st person singular pronoun *yo* for narrative functions occurs in nearly 10% of cases of original Spanish and is typical of fiction texts, in particular of drama. It is often the case that the instance of the pronoun is preceded by the coordinating conjunction *y*, thus indicating a sequence in the action described:

Puri: Y la atamos.

Toña: Y yo le di una hostia, aunque no me arrepiento.

The narrative function is the most common one in Spanish translations with about 30% of the total, three times more common in translations than in original Spanish texts.

The two remaining functions identified are very infrequent in both original and translated texts, and only the formulaic use is statistically significant. The formulaic use of the pronoun *yo* refers to short fossilized expressions including it with a clearly pragmatic meaning as in: *yo qué sé, un qué sé yo, ya lo dije yo*, etc. This function occurs slightly more frequently in original texts (5%) than in translations (1%), which is an expected result, since pragmatic uses are more difficult to convey in translations.

Case study 2: PODER (LEXICAL EQUIVALENT OF CAN, MAY)

The lexical equivalent of modal verbs *can* & *may* in Spanish is *PODER*. Spanish has other resources such as verbal mood and a variety of tenses indicating non factuality to convey the modal meanings that constitute the core meaning of the English modal verbs. When the typical modal uses of English *can* are transferred on to Spanish *poder* new functions that do not exist in English are associated unnecessarily to Spanish *poder*. This tendency to transfer because of this confusing cross-linguistic dissimilarity can also give rise to an excessive use of certain functions conveyed by resources different to those to which it is normally associated in non-translated Spanish.

The analysis of the various functions of PODER in original Spanish texts and in translated texts yielded the following results:

CREA		PODER			P-ACTRES	
%	RAW CASES			RAW CASES	%	
44.1	146	<i>POSSIBILITY</i>		88	39.28	
28.4	94	<i>ABILITY</i>		47	21	
7.85	26	<i>PERMISSION</i>		31	13.83	
--	--	<i>ASPECTUAL</i>		24	10.71	
0.9	3	<i>PREDICTION</i>		4	1.78	
18.73	62	<i>OBLIGATION/ADVISABILITY</i>		30	13.39	
100%	331	TOTALS		224	100%	

Table 6 . Quantitative data for PODER

The most obvious and prototypical functions of *can/poder* are those which are also part of the lexical meaning of the Spanish verb: possibility and ability. Cases of obligation/advisability are not a cross-linguistic problem either, as they generally appear in both languages in negative contexts indicating some ethical reason or some social convention. The main difference between translated and non-translated language is the rise of a new function, ‘aspectual’, which is purely interlinguistic (i.e. owes its existence to the contrast between original and translated language). It is used to indicate those uses of *poder* in

contexts where it is accompanied by a perception verb – either sensorial or intellectual- and which is typical of translated texts. When this collocation happens in non-translated Spanish it means ‘ability to do....’ In a situation where there are difficulties or some obstacle to see, understand...etc.

By applying statistical analyses, it was found that only some of the functions were actually significantly different, as shown in the following table:

PODER	z-score	p-value
<i>POSSIBILITY</i>	<i>1.13</i>	<i>0.2590</i>
<i>ABILITY</i>	<i>1.97</i>	<i>0.489</i>
<i>PERMISSION</i>	-2.28	<i>0.0227</i>
<i>ASPECTUAL</i>	-6.09	####
<i>PREDICTION</i>	-0.91	<i>0.3624</i>
<i>OBLIGATION/ADVISABILITY</i>	<i>1.66</i>	<i>0.971</i>

Table 7. Statistical significance of PODER values.

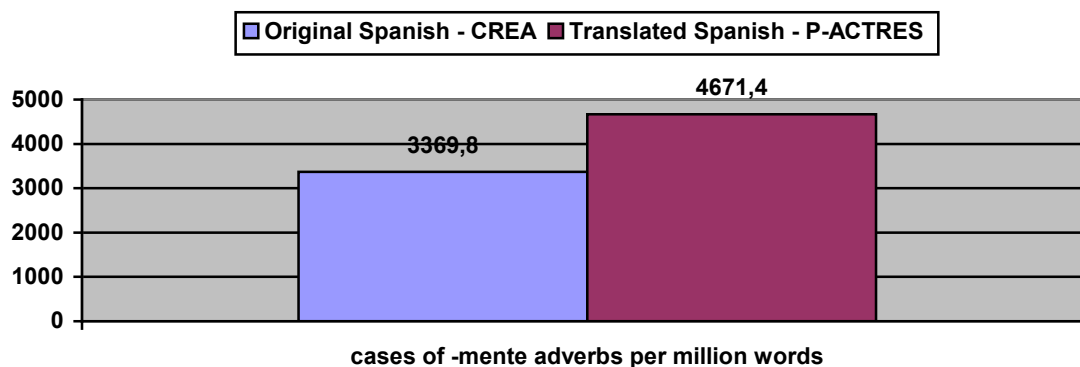
Only ‘permission’ and the new function ‘aspectual’ show statistical significance in translated Spanish, which makes them good candidates to be used as anchor uses in TQA English-Spanish.

Case study 3: ADVERBS ENDING IN –MENTE

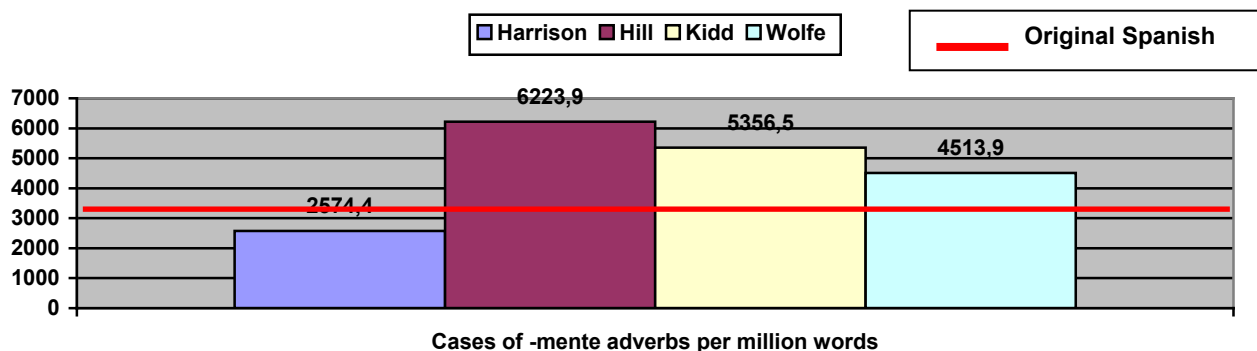
Adverbs ending in –mente are generally considered a good numerical indicator for identifying Spanish translations of English texts. The reason is that most –mente adverbs are far less commonly employed in original Spanish texts than their frequency of occurrence in translations as the equivalent of cognate adverbs in -ly in English texts (Ramón & Labrador, forthcoming).

The quantitative analysis yielded a whole of 7,269 cases of –mente adverbs in the section on novels in CREA, the reference corpus of original Spanish. Taking into account that

this section contains 2,157,056 words, there are 3,369.8 cases of -mente adverbs per million words in original Spanish texts (novels). In contrast, there were 1,967 cases of -mente adverbs in the P-ACTRES corpus of Spanish translations (subcorpus novels, containing 421,065 words). In short, there were 4,671.4 cases of -mente adverbs per million words in Spanish translations of English texts, many of which most probably prompted by an English—ly adverb in the source text. Figure 1 illustrates graphically this difference.



Once we know the general figures for original Spanish, a further test was carried out using 4 extracts from Spanish translations to validate the results. Figure 2 shows the number of cases of -mente adverbs found in each extract in the proportion of cases per million words.



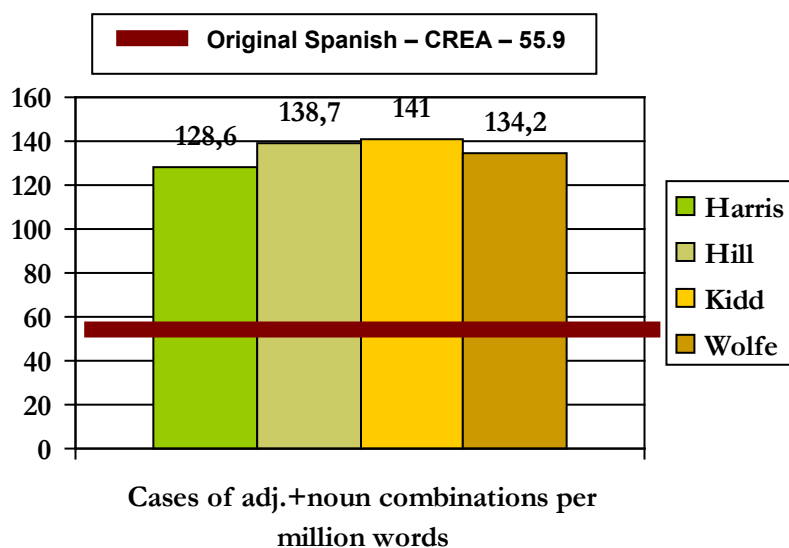
In Hill, Kidd and Wolfe the differences with original Spanish are statistically significant, but in the case of the novel by Harrison, there is no statistically significant difference in this aspect between the translation and original Spanish.

Case study 4: PRE-MODIFYING ADJECTIVES

In Spanish, the unmarked position of adjectives is the postnominal position, but the premodifying position is also possible with certain pragmatic connotations and with certain adjectives in nearly fixed constructions. Considering that the unmarked position of adjectives in English is the premodifying one, it is expected that Spanish translations of English source texts present a much higher number of premodifiers due to the influence of these particular source texts.

For this quantitative study, the data employed have been extracted from CREA for the data of original Spanish. In previous studies (Ramon 2003) the percentage of occurrence of single adjectives as premodifiers in original Spanish was determined with respect to the whole of single word modifiers. The idiomatic frequency is of 55.9 cases of single premodifying

adjectives per million words. The data were then contrasted with the percentage of occurrence found in the four extracts from our validation corpus. The results are shown in Figure 3:



The difference is statistically significant for a p-value under 0.01 in all four cases. The results indicate that the number of premodifying adjectives occurring in Spanish translations of English texts point very clearly towards their translated nature and reflect the influence of the source language. As for the use of this parameter as an indicator of translation quality, all four extracts show a very similar deviation from original Spanish texts.

Case study 5: PERSONAL PRONOUN ÉL

ÉL yields the following data, which prove that original Spanish uses this pronoun neutrally in quite a high number of cases (11,93%), similarly to ‘yo’ (10,05%). It is also significant that ‘él’ is mostly used as narrative discourse marker both in non-translated and translated Spanish (43,46% and 40.57% respectively).

CREA		P-ACTRES	
		ÉL	
%	RAW CASES	RAW CASES	%
11.93	42	23	9.43
	<i>Neutral</i>		
3.69	13	7	2.87
	<i>Optional emphasis</i>		

12.22	43	<i>Non-optional emphasis</i>	17	6.97
27.56	97	<i>Contrastive</i>	97	39.75
1.14	4	<i>Formulaic</i>	1	0.41
43.46	153	<i>Narrative discourse marker</i>	99	40.57
100	352	TOTALS	244	100

Table 8 . Quantitative data for ÉL

Translated usage shows differences mainly in the areas labelled as discourse marker, non-optional emphasis and neutral, which may indicate that a) translated texts are less cohesive, b) translated texts overuse emphatic markers, and c) translated texts are less economic in their use of pronominal resources than original language texts.

The statistically significant results for ÉL are shown in the following table:

ÉL	z-score	p-value
<i>Neutral</i>	0.96	.3346
<i>Optional emphasis</i>	0.55	.5827
<i>Non-optional emphasis</i>	2.09	.0363

Contrastive	-3.12	.0018
Formulaic	0.96	.3390
Narrative reference marker	0.70	.4822

Table 9. Statistical significance for ÉL values

5.4. Juxtaposition and contrast: obtaining the anchor values

At this point we need to determine whether the use of the pronouns in non-translated and translated Spanish is identical, or, if not, which the actual anchor differences are. This tool building process is necessary in order to i) make sure that we are using the strongest possible conclusions from limited amounts of data, that is, that the anchors we are proposing are truly useful for the pair English-Spanish; ii) avoid the ‘confirmation bias’ on our part, that is, the tendency to search for interpretations that confirm our unverified view(s) on the basis of raw quantitative data, and iii) to provide a transition from quantitative data to relevant information usable by our intended final users. In other words, comparing results from each translation with just CREA results can be misleading, more time-consuming and force us to reach wrong assessment conclusions. By contrast, focusing on grammatical uses that have been empirically proven to cause distortion -or even unintelligibility- in translations of English narrative into Spanish can be said to be a real help for the evaluation of the linguistic quality of translated texts. The results in the tables above indicate that the most useful values for our assessment purposes are as follows:

ANCHOR USES	YO	ÉL
<i>Neutral</i>	√ overuse	
<i>Optional emphasis</i>		
<i>Non-optional emphasis</i>		√ underuse

<i>Contrastive</i>	√ <i>underuse</i>	√ <i>overuse</i>
<i>Formulaic</i>	√ <i>underuse</i>	
<i>Narrative reference marker</i>	√ <i>overuse</i>	

Table 10 . Anchor uses for YO and ÉL

YO: There is an overuse of the narrative and neutral functions. However, there is a clear underuse of the contrastive and formulaic functions of this pronoun in translations.

ÉL: Significant differences lie in two main values: contrastive, which is overused in the translations, and non-optional emphasis, which shows an underuse in translated Spanish.

ANCHOR USES	PODER (PRES. IND)
<i>POSSIBILITY</i>	
<i>ABILITY</i>	
<i>PERMISSION</i>	√ <i>overuse</i>
<i>ASPECTUAL</i>	√ <i>overuse</i>
<i>PREDICTION</i>	
<i>OBLIGATION/ ADVISABILITY</i>	

Table 11 . Anchor uses for PODER

5.5. Interpretation of anchor results

As noted by Davies (2001: 7), what is significant is not the size of the significant results, but their effect and consequences on language use (Rabadán 2008). And in our case these are as follows:

- a. An overuse of the **neutral functions** of the subject pronouns indicates a trend towards verbosity in translations, thus leading to a flattening and cluttering effect in translated texts.
- b. An overuse of **non-optional emphasis** reveals an uneconomic use of the expressive (deictic) resources of Spanish or the existence of modulations transferring pragmatic emphasis from other parts of the text. In contrast, the underuse of this function results in a downgrading in idiomaticity.
- c. An overuse of **contrastive cases** makes the texts repetitive leading to a more laborious and complicated reading. It renders a text distasteful, as it constantly explicitates information which is already clear to the reader. An underuse of the contrastive function causes ambiguity and a low degree of cohesion in translated texts.
- d. An underuse of the **formulaic function** was observed for the cases of *yo*, resulting in a lower degree of idiomaticity of the translations, since there are few fixed expressions containing these pronouns in translated texts.
- e. An overuse of **narrative discourse markers** renders a text redundant and repetitive. An underuse of the same reveals that translated texts are less cohesive, which necessarily affects textual anaphora and severely detracts from textual intelligibility.
- f. An overuse of **permission** cases also distorts the flow of the Spanish text, as frequently these contexts amount to a check on the other participant's agreement on something. In addition, this practice introduces unwanted and unnecessary materials in the text which produce verbosity and detract from translation quality.
- g. An overuse of **aspectual poder** indicates a high degree of redundancy and translationese in the Spanish text, which do not need additional modality markers. This practice affects the interpretation of the text, to which it lends and additional **ability** meaning and has consequences for textual cohesion.

6. TESTING FOR QUALITY: CASE STUDIES.

The data obtained from comparing the uses of personal pronouns in original Spanish texts and in translated Spanish texts were used again to attempt and provide assessment of other recent Spanish translations of English texts. These have been obtained from the inventory of TEST_ACTRES

6.1. The texts to be assessed: TEST_ACTRES

TEST_ACTRES is an English-Spanish catalogue and open corpus containing 17 contemporary English language narrative translated into Spanish. Its role in our research is to facilitate the testing and empirical verification of ACTRES applications suitable for this textual register.

Four random extracts from contemporary novels have been chosen for this test: K. Harrison's *The Seal Wife* (2003), T. Wolfe's *I am Charlotte Simmons* (2005), S.M. Kidd's *The Mermaid Chair* (2006) and Joe Hill's *Heart-shaped Box* (2006).

6.2. CASE STUDY I: K. Harrison's *The Seal Wife/ La mujer de nieve*

The Spanish translation of Harrison's novel was published in Spanish with the title *La mujer de nieve* in 2005. The translator was Encarna Castejón, and the book was published in Barcelona by Anagrama. The extract from the translation included 15,537 words, and there

was only one case of the pronoun *yo* and another case of the pronoun *tú*, both in the same sentence and with a contrastive function: *Tú eres mi madre selva, yo soy tu abeja*.

CREA	p-score	z-score	ÉL	TT HARRISON'S NOVEL	YO	z-score	p-score	CREA
				<i>Neutral</i>				
				<i>Optional emphasis</i>				
43	0.0822	1.74	1	<i>Emphatic non-optional</i>				
97	0.0008	-3.34	20	<i>Contrastive</i>	1	-1.02	0.3101	181
				<i>Formulaic</i>				
				<i>Narrative discourse marker</i>				
352			37	TOTALS	1			368

K. Harrison's *The Seal Wife* (2003)/ *La mujer de nieve* shows significant misuses of both pronouns and verb *poder*. It reveals a notorious degree of overuse of subject pronouns with a contrastive function and of 'aspectual poder' (7 cases). The first necessarily affects textual anaphora and reference ties and networks and severely detracts from textual intelligibility because of redundancy. The second produces changes in meaning (ability) as a result of marked translationese and the redundancy of resources affects textual cohesion and as a consequence overall quality.

- ❑ [slide] No significant misuses of most anchors.
- ❑ Considerable overuse of subject pronoun *él* with a contrastive function.
- ❑ This affects textual anaphora and reference ties and networks and detracts from textual intelligibility because of redundancy.
- ❑ Overuse of aspectual & permission *poder*, a high degree of redundancy and translationese

6.3. CASE STUDY II: T. Wolfe's *I am Charlotte Simmons*/ *Soy Charlotte Simmons*

The Spanish translation of Wolfe's novel was published with the title *Yo soy Charlotte Simmons* in 2005. The publishing house was Ediciones B (Barcelona) and the translators were Eduardo Iriarte and Carlos Mayor. The extract of the translation contained 21,489 words.

CREA	p-score	z-score	ÉL	TT WOLFE'S NOVEL YO	YO	z-score	p-score	CREA
				<i>Neutral</i>	13	-5.02	<0	37
				<i>Optional emphasis</i>				
43	0.0817	-1.74	9	<i>Emphatic non-optional</i>				
97	0.1247	-1.54	16	<i>Contrastive</i>	5	3.65	0.0003	181
				<i>Formulaic</i>	3	-0.92	0.3584	20
				<i>Narrative discourse marker</i>				
352			25	TOTALS	32			368

T. Wolfe's *I am Charlotte Simmons* (2005)/ *Soy Charlotte Simmons* rates well for anchor value 'formulaic', which may be interpreted as a healthy sign of an idiomatic translation. By contrast, it shows a curious and significant overuse of neutral values (Yo/I) and worryingly low scores for the anchor value contrastive. The first indicates verbosity in translations, thus leading to a flattening and cluttering effect in translated texts. Underusing the contrastive function (of yo) causes ambiguity and a low degree of cohesion in this text.

This text rates high for overall assessment of linguistic and textual effectiveness. Only yo-management presents some flaws, which not surprisingly mirror the features of translated language as revealed by P-ACTRES.

Wolfe's text shows no significant difference in the uses of 'poder' but for 'aspectual poder' (5 cases), which presents a notorious overuse when compared to non-translated usage. As in the previous novel, this redundant use is responsible for misunderstandings of the translated text, as it adds the 'ability feature' to contexts where there was nothing marked as such. The result of these massive translationese also affects cohesion and therefore overall textual quality.

- ❑ [slide] Significant overuse of neutral values (Yo/I): verbosity in translations, thus leading to a flattening and cluttering effect in translated texts.
- ❑ Underuse of the anchor value contrastive: ambiguity and a low degree of cohesion in this text.
- ❑ Overuse of aspectual *poder*, high degree of translationese and redundancy.
- ❑ This text rates high for overall assessment of linguistic and textual effectiveness. Only yo-management presents some flaws, which not surprisingly mirror the features of translated language as revealed by P-ACTRES.

6.4. Case study III: S.M. Kidd's *The Mermaid Chair*/ *El secreto de la sirena*

Our Spanish translation of S.M. Kidd's novel was published with the title *El secreto de la sirena* in 2007 in the publishing house Edicions B in Barcelona. The translator is Javier Guerrero. The extract contains 15,495 words and there were 57 cases of *yo* and 7 of *tú*. The distribution of the functions of the pronoun *yo* in this extract was the following: narrative (24 cases, 42.10%), optional emphasis (17 cases, 29.82%), contrastive (9 cases, 15.78%), neutral (7 cases, 12.28%).

CREA	p-score	z-score	ÉL	TT KIDD'S NOVEL	YO	z-score	p-score	CREA
				<i>Neutral</i>	7	-0.51	0.6077	37
				<i>Optional emphasis</i>				
43	0.0495	1.96	0	<i>Emphatic non-optional</i>				
97	0.6025	-0.52	9	<i>Contrastive</i>	9	4.72	0	181
				<i>Formulaic</i>	0	1.80	0.0714	20
				<i>Narrative discourse marker</i>	24	-6.52	0	36
352			28	TOTALS	57			368

S.M. Kidd's *The Mermaid Chair* (2006)/ *El secreto de la sirena* rates satisfactorily for anchors neutral and formulaic, which points to both a grammatically correct and idiomatic use of subject pronouns. The ratings for contrastive, narrative discourse marker and non-optional emphasis, however, show a marked departure from efficient usage. Contrastive is underused in the case of *yo*, which creates areas of slow reading. In short, this creates partial ambiguity

and affects text intelligibility. Overusing the anchor narrative discourse marker may affect textual cohesion and roundedness rendering the text redundant and repetitive. The absence of non-optional emphasis also adds to less idiomaticity.

No cases for the anchor values of *poder* were found in this text, which makes it impossible to offer any assessment for this particular text in this respect.

- ❑ [slide] Contrastive values are underused in the case of *yo*, which creates ambiguity and affects text intelligibility.
- ❑ Overusing the anchor ‘narrative discourse marker’ may affect textual cohesion and renders the text redundant and repetitive.
- ❑ The absence of non-optional emphasis also adds to less idiomaticity.
- ❑ No cases of *poder* were found.

7.5. Case study IV: Joe Hill’s *Heart-Shaped Box/ El traje del muerto*

The Spanish translation of Joe Hill’s novel has been published under the title *El traje del muerto* in 2007 by Summa de Letras (Madrid). Julio A. Sierra holds the copyright for the translated text. In this text the extract contains 20,000 words +. The raw results for our usable anchor results are 10 *yo*, 4 *tú*, 13 *él* and 23 *usted*.

CREA	z-score	p-score	ÉL	TT HILL’S NOVEL	YO	z-score	p-score	CREA
				<i>Neutral</i>	0	1.06	0.02911	37
				<i>Optional emphasis</i>	5			
	1.34	0.5309	0	<i>Emphatic non-optional</i>	1			
	-3.25	0.0012	9	<i>Contrastive</i>	3	1.20	0.2311	181
				<i>Formulaic</i>	0	0.76	0.4487	20
			4	<i>Narrative discourse marker</i>	1	-0.02	0.9818	36
			13	TOTALS	10			368

Joe Hill’s *Heart-Shaped Box/ El traje del muerto* rates well for all anchor values except for contrastive. This means that differences in usage for most anchors have not been found significant, which may be interpreted as a healthy sign of an idiomatic translation. There is a notorious degree of overuse of subject pronouns with a contrastive function. This necessarily makes the text more repetitive than it is necessary and it leads to a more laborious and complicated reading as it explicates information which is already clear to the reader. This text rates high for overall assessment of linguistic and textual effectiveness. Only contrastive *él* management presents some flaws, which seem to obey to one of the universals of translation behaviour: explicitation (Rabadán, Labrador & Ramón 2007).

This text shows a significant overuse of ‘aspectual *poder*’ with the same consequences as in previously assessed novels: changes in meaning (ability) as a result of marked translationese, redundancy of resources which affects textual cohesion and, when these phenomena are considered together, their presence in the text detracts from overall quality.

- ❑ [slide] Differences in usage for most anchors have not been found significant: idiomatic translation.
- ❑ Overuse of subject pronouns with a contrastive function: text is more repetitive and more explicit.
- ❑ Significant overuse of aspectual *poder*: changes in meanings.

- This text rates high for overall assessment of linguistic and textual effectiveness. Only contrastive *él* management presents some flaws, which seem to obey to one of the universals of translation behaviour: explicitation (Rabadán, Labrador & Ramón 2007)

SLIDES 17 & 18

As for the numerical values, slide 17 illustrates the quantitative data with respect to -mente adverbs in the four extracts from the test corpus and compares them with the number of cases in original Spanish texts from CREA. A clear overuse (a statistically significant overuse) can be noted in three out of the four cases, while Harrison's translation shows no such difference. In the case of slide 18, the graph compares de the number of cases of single premodifying adjectives in original Spanish and in the four translated extracts. It can be noted very clearly that all four sample texts show clear features of being translations by having around three times more constructions of this type than original Spanish texts from CREA.

7. A SEMI-AUTOMATIC TQA PROTOTYPE

The ACTRES team engineering researcher(s) are currently working on a TQA application interface written in JAVA. So far, 3 different stages have been tackled:

Stage 1: Analyze and tag document to be assessed semi-automatically. Morphosyntactic tags used are EAGLES (V2.0). The lexicon chosen for these first steps has been compiled by M. Marimon at IULA "Institut Universitari de Lingüística Aplicada" at Universitat Pompeu Fabra (http://www.upf.edu/pdi/iula/montserrat.marimon/dict_morph.gz), although any other can be used.

Stage 2: Add up expert knowledge to the tagged document. This means that the EAGLES tags need to be complemented, in our case and for the time being by adding an -XX ending, which will account for different functions in Spanish. Example puedo {[VMIP1S0-PO] poder} (-PO => posibilidad). Currently, this stage has to be supervised by the user, who has to select the appropriate option for each case. Once this has been done, the program automatically factors in the cases for each anchor. The results are automatically supplied both as raw cases and as %.

Up to now, the application includes 5 anchors: yo, poder, Adj + N, estar + gerund (periphrastic uses) and -mente Adverbs.

In the near future work will concentrate on 1) the automatization of expert knowledge tagging so as to minimize user intervention, and 2) provide an automatized way of measuring linguistic quality as compared to adequate grammatical usage. A derived, secondary use of this application would be to identify a text as original or translated.

Usability has been guaranteed by using JAVA, which allows it to run on any computer. The non-stop assessment option completes all the steps automatically. This ensures usability for applied professionals as final users.

8. RESULTS OF TQA APPLICATION

Our results indicate varying degrees of textual and linguistic quality in the four case studies, which range from a remarkably effective use of grammatical resources in the target language in Harrison's and Hill's texts to some intelligibility problems in Wolfe's text and a disappointing overall performance in S.M. Kidd's Spanish text. The reasons lie overwhelmingly in one area: a defective management of the contrastive use of the pronouns, which affects text progression and intelligibility. The consequences are that the perception of the authors and their work by their Spanish language readers can be negatively affected and render low ratings of readers' appreciation creating thus low commercial and market figures.

The results of our verification of applicability seem to corroborate the working hypothesis we put forward at the beginning: that a number of descriptive anchor values,

adequately conceptualized, can be an effective tool for translation assessment. However, to become fully usable by applied professionals different sets of anchors have to be identified and verified for different text varieties and a software tool which makes the process semi-automatic produced. We are refining the set of descriptive anchors and working on the second.

REFERENCES

- Alarcos, E. 1980: *Estudios de gramática funcional*. Madrid: Gredos.
- Byrne, J. 2006. *Technical Translation. Usability Strategies for Translating Technical Documentation*. Dordrecht: Springer.
- Celle, A. 2005. "The French future tense and English *will* as markers of epistemic modality". *Languages in Contrast*, 5 (2). 181–218.
- Colina, Sonia. 2008. "Translation quality evaluation. Empirical evidence for a functionalist approach". *The Translator* 14 (1): 97-134.
- Chesterman, A. 1998: *Contrastive Functional Analysis*. Amsterdam and Philadelphia: John Benjamins.
- Enríquez, E. 1984: *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: CSIC.
- Fernández Soriano, O. 1989: *Rección y ligamiento en español: aspectos del parámetro del sujeto nulo*. PhD Dissertation. Universidad Autónoma de Madrid.
- Granger, S., J. Lerot and S. Petch-Tyson (eds.) 2003: *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi.
- House, J. 2001. "Translation quality assessment. Linguistic description versus social evaluation". *Meta* 46 (2): 243-257.
- Huddleston, R. and G.K.Pullum. 2002: *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- Krzeszowski, T.P. 1990: *Contrasting Languages. The Scope of Contrastive Linguistics*. Berlin and New York: Mouton de Gruyter.
- Labrador, B. 2005: *Estudio contrastivo de la cuantificación inglés-español*. León: University of León.
- Lowry, R. 1999-2007. *Concepts and Applications of Inferential Statistics*. <http://faculty.vassar.edu/lowry/webtext.html/>. Visited November 2006.
- Luján, M. 1999: "Expresión y omission del pronombre personal." En Bosque, I. and V. Demonte. *Gramática descriptiva de la lengua española*. Madrid: Espasa.
- Marcos Marín, F. 1978: *Estudios sobre el pronombre*. Madrid: Gredos.
- Parallel English-Spanish Contrastive Analysis and Translation. P-ACTRES. 2002-2008. <http://actres.unileon.es/corpussearch/> (Password-restricted access) Visited April-May 2008.
- Quesenbery, W. 2001. "What does usability mean: Looking beyond 'ease of use.'" Pre-publication article for the STC 2001 Conference Proceedings. <http://www.cognetics.com/papers/whitney/whitney1.html/>. Visited October 2006.
- Rabadán, R. 2005. "The Applicability of Description. Empirical research and translation tools". In *Contemporary Problematics of Translation Studies*. C. Toledano, ed. Special issue of *Revista Canaria de Estudios Ingleses*, 51. 51-70.
- Rabadán, R. 2006. "Modality and modal verbs in contrast: mapping out a translation(ally) relevant approach English- Spanish". *Languages in Contrast* 6 (2): 261 -306
- Rabadán, R. 2008. "Refining the idea of 'applied extension'". In Miriam Shlesinger et al. (eds). *Beyond Descriptive Translation Studies*. Amsterdam: John Benjamins. 103-118.
- Rabadán, R. 2007 in press. "Tools for English-Spanish cross-linguistic applied research". *JES. Journal of English Studies*.
- Rabadán, R., B. Labrador and N. Ramón. (forthcoming) "Corpus-based contrastive analysis and translation universals: A tool for translation quality assessment English→Spanish?" *Babel: Revue Internationale de la Traduction*.
- Ramón, N. 2003: *Estudio contrastivo inglés-español de la caracterización de sustantivos*. León: University of León.
- Ramón, N. & B. Labrador (forthcoming). "Translations of -ly adverbs of degree in an English-Spanish parallel corpus". *Target*.
- Real Academia Española. Visited March 2007. <http://corpus.rae.es/creanet.html>
- Toury, G. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam and Philadelphia: John Benjamins.