

Translation description for assessment and post-editing

The case of personal pronouns in translated Spanish

Noelia Ramón and Camino Gutiérrez-Lanza

University of León

This paper presents a corpus-based descriptive research procedure for the identification of significant divergences between original Spanish and Spanish translated from English. When considering the language pair English-Spanish, personal pronouns seem to be good markers of significant differences (anchor phenomena), since they must obligatorily occur in English, but not in Spanish. To test this hypothesis, empirical data have been extracted from a large reference corpus in Spanish (CREA) and from an English-Spanish parallel corpus (P-ACTRES), in both cases from the fiction subcorpora. Statistically significant differences have been found in some of the uses of personal pronouns, having textual and pragmatic implications in the target texts. The aim is to use the results obtained in the case of personal pronouns, together with results from other linguistic areas, to build a semi-automated tool for the post-editing of Spanish translations of texts written originally in English.

Keywords: translation post-editing, anchor phenomena, corpora, personal pronouns

1. Introduction

Translation Studies as a distinct scientific discipline has brought us academic recognition in the past few decades. Many theoretical and descriptive studies have been carried out by different authors and in a wide range of socio-linguistic and cultural backgrounds. However, the close utilitarian links that should be established between theoretical and descriptive studies, on the one hand, and the applied activities that derive from them, on the other, are still not clearly developed (Toury [1995] 2012, 11). In many environments, tasks such as translation assessment, proofreading, post-editing,

etc. are carried out by a type of professional on the rise, the language-services provider, who can benefit from translation research in a number of ways.

In Translation Studies, as in all fields of expertise, theoretical knowledge and descriptive studies may be expected to produce applied technological tools that will eventually benefit actual translation practice. The dream of developing fully automated machine translation, however, has vanished, since it has been demonstrated that in most contexts machines cannot translate without the help of human beings, or at least do not provide the same quality standards. In the case of the post-editing of translations, most research until now has focused on machine translation output (Somers 1997; Krings 2001; Vieira and Specia 2011; O'Brien et al. 2014). In contrast, this paper sets out to explore the possibilities of designing an effective and efficient tool to contribute to the assessment and post-editing of human translations; it will do so by using a limited number of language-pair-bound descriptive anchor phenomena which are specifically grammatical and not lexical in nature. We will describe the procedure followed to employ Spanish personal pronouns as indicators of translation quality. The results of the study will contribute to the improvement of post-editing techniques for the language pair English-Spanish.

Personal pronouns have been found to occur much more often in fiction than in non-fiction. This is why, for practical reasons, our proposal applies to Spanish translations of English fiction. However, the procedure is useful for any language pair and any text type or register, provided the data are relevant and appropriate for those other languages or textual varieties. The procedure needs to be user-friendly, so that service providers can incorporate it easily to their daily work routine.

When examined from a cross-linguistic analytical perspective, certain grammatical areas show clear differences in the functions some of the resources can convey in the languages being considered. Empirical data demonstrate that dissimilarity in the ways grammatical functions are actualized is a constant source of cross-linguistic problems affecting both text processing and production, and marks the difference between translated and non-translated language (Tirkkonen-Condit 2002). Language-specific associations between grammatical function and formal resource are defined as “anchor phenomena”: “those language-pair specific resources that can be empirically singled out as recurrent problem-triggers in cross-linguistic communication” (Rabadán 2010, 10). In the case of the language pair English-Spanish, the list of anchor phenomena includes the following grammatical items: personal pronouns, modal verbs

(Rabadán 2006), adverbs (Ramón and Labrador 2008), verbal tenses (Rabadán 2009), demonstratives (Labrador 2011), negation (Rabadán and Izquierdo 2013), adjective-noun combinations (Ramón 2015), etc.

Anchor phenomena are perceived as being cross-linguistically equivalent, but tend to convey partially divergent functions. Since cross-linguistic grammatical function dissimilarity cannot be assumed to be the same for different language combinations or in each direction, the form-function associations that qualify as anchor phenomena also differ by direction and language combination. Anchor phenomena can be used as key indicators of the degree of success in cross-linguistic transfer. A translation 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.

There is a general assumption that the language of translations differs from the language in original texts as a result of the translation process itself (Baker 1993). Translations do not necessarily have to abide by all target-language norms, and author and translator style may be the cause of a number of deviations from target-language standard writing. In general, more flexibility is allowed and readers are more tolerant with respect to peculiar uses of certain linguistic items which may be triggered by the source-language texts or authors. However, it is also true that a reference point is required and the native usage of the target language has generally been considered a valid reference for translation quality. As claimed by Toury (2004, 28), “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.” We believe that it is possible to identify at least some of those particular grammatical structures where translation performance could actually be improved in the post-editing process to make them more similar to native usage. This is why our paper supports the idea that “it is the linguistic analysis which provides grounds for arguing an evaluative judgement” (House 2001, 254). Previous attempts in this direction with the language pair English-Spanish can be found in Rabadán, Labrador and Ramón (2009).

In order to examine the real usefulness of personal pronouns as anchor phenomena for assessment and post-editing, we need to analyse both Spanish original texts and Spanish translations of English. First, empirical data will be extracted from the main reference corpus of original European Spanish texts CREA (Corpus de Referencia del Español Actual) to determine the native usage of personal pronouns in Spanish fiction. Secondly, this information will be compared with data from the English-Spanish

parallel corpus P-ACTRES, which contains contemporary original texts in English and their corresponding translations into Spanish. Since both corpora are not identical in size, we have employed tests for statistical significance to ensure the reliability of the data. In particular, we have used a statistical test called “hypothesis testing for independent proportions”.

The results obtained from the corpus-based description will clearly show the divergences between original and translated Spanish with respect to the usage of personal pronouns. A statistically significant difference in the number of pronouns in the translations and in the distribution of the functions associated to each particular personal pronoun will suggest transfer from the source language, indicating poor management of expressive resources in the translated texts.

Together with other anchor phenomena, the findings obtained for personal pronouns have been implemented in a computer programme for the assessment and post-editing of Spanish translations of texts written originally in English (PETRA 1.0©, LE-222-13) (Rabadán et al. 2014). From a broad perspective, data obtained from the analysis of anchor phenomena are easy to handle by final applied users and may contribute, at least, to the following applied activities: translator training, translation-quality assessment, post-editing, identification of cross-linguistic plagiarism, and detection of pseudo-translation in Spanish language texts.

2. Object of study: personal pronouns in English and Spanish

In the case of English and Spanish, a number of general contrastive grammars have listed the main typological differences between these two languages with language teaching, translation practice, and translator training as the main applied purposes in mind: Vázquez Ayora (1977), Martínez Vázquez (1996), Whitley (2002), López Guix and Minett Wilkinson (2009), and Eldredge and Rodríguez (2015). In these works, pronouns are always considered a problematic aspect in this particular language pair due to the fact that their occurrence is compulsory in English, whereas the subject may be omitted in Spanish in most contexts. Because of this major difference between the two languages, our working hypothesis is that personal pronouns are good candidates for anchor phenomena.

The main communicative function of personal pronouns is personal reference (Halliday and Hassan 1976), and pronouns always refer to given information, thus

having thematic meaning. Pronouns have two main functions in English: deixis and anaphora (Huddleston and Pullum 2002, 1463-1482). However, the actualization and distribution of these roles is different in English and in Spanish. English is a subject-dominant language, whereas in Spanish subject pronouns are typically omitted in the deictic function, as the information related to person and number is already included in the verbal inflections. In addition, Spanish grants subject pronouns a number of extra uses which add pragmatic and textual meanings such as contrast, emphasis, thematization, narrative marking, etc. (Marcos Marín 1978; Alarcos 1980; Fernández Soriano 1989; Luján 1999). All of these differences will hypothetically lead to a) more personal pronouns being used in translations through the influence of the source language (already found by Maia [1998] for the language pair English-Portuguese), and to b) deviations in certain uses of Spanish personal pronouns in translations.

Previous English-Spanish contrastive studies dealing with pronouns focus primarily on the issue of thematization, and mostly with foreign-language learners in mind: “Spanish pronouns also give problems to English speakers because of their strikingly different sentence position and combinations with each other” (Whitley 2002, 184). To our knowledge, there is no research yet on Spanish pronouns from the point of view of translations from English source texts. No major diachronic variations have been accounted for so far in the use of native Spanish personal pronouns due to the influence of translations, but the differences identified in this paper between original and translated usage of these pronouns may eventually lead to such diachronic variations.

3. Method and tools

The long-running research project ACTRES (<http://actres.unileon.es/>), based at the University of León, has been working on corpus-based English-Spanish contrastive analysis and translation for several years. This particular study is an empirical analysis of one specific linguistic phenomenon, namely, Spanish personal pronouns, following a descriptive corpus-based contrastive method. Our working procedure is loosely based on Krzeszowski's (1990) model and consists of the following stages: selection and sampling, description, juxtaposition and contrast. In this section we will describe in detail the method employed to carry out the analysis using three main tools: cross-linguistic labels, corpora, and statistics.

3.1 *Tertium comparationis*: cross-linguistic labels

An essential conceptual tool 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. Our labels are useful for cross-linguistic meaning-discrimination in the contrastive process (Rabadán 2005). Their role is to identify the meaning features that are relevant for applied purposes. This means that general linguistic 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 on the functions of Spanish pronouns in original and translated texts, the labels have been set up drawing on Enríquez (1984), Fernández Soriano (1989), and Luján (1999) primarily, and they show a very irregular distribution depending on each particular form. The information has been conveniently rearranged into the following categories and all of the examples in this section have been taken from CREA. The relevant personal pronouns have been underlined in each example:

1. Neutral. This use corresponds to the main pragmatic function of pronouns: person deixis. In this function pronouns cannot be omitted for syntactic reasons. This function in Spanish is particularly relevant in the cases of *yo* and *tú*. Ex. (1): tú y yo (me and you).
2. Emphatic:
 - a. Optional emphasis. This pragmatic function refers to a surplus use of the pronoun which does not affect either deixis or anaphora, but which contributes textual meanings such as marker of formality, as in the case of the pronoun *usted*, focalization on the subject, as in example (2) yo *hago mi trabajo* (*I do my job*), etc.
 - b. Non-optional emphasis. This pragmatic function is particularly relevant in the cases of *yo* and *él*, and occurs when the pronoun is compulsory for

- grammaticality because of syntactic reasons. Ex. (3) *soy yo, Teresa* (it's me, Teresa). Ex. (4) *¿No fue él quien le pidió que me recibiera en su nombre?* (Wasn't he the one who asked you to receive me in his name?).
- c. Contrast. This pragmatic function is particularly relevant in the case of most pronouns and indicates comparison and distinction between one particular pronoun and other pronouns and noun phrases in the same sentence. Ex. (5) *el matrimonio decidía salir y él se encargaba de aquella vigilancia* (the couple decided to leave and he was in charge of that observation). Ex. (6) *¡Eres tú quien debería estar aquí, no yo!* (It is you who should be here, not me!).
- d. Formulaic function. A pronoun has a formulaic function when it occurs in fixed expressions where it is never omitted in idiomatic Spanish. This function is particularly relevant in the case of *tú*. Ex. (7) *vete tú a saber* (who knows?). Ex. (8) *allá tú* (it's up to you).
- e. Narrative discourse marker. This is a type of discourse reference marking which occurs when the explicit pronoun moves the narrative action forward, changing the point of view from one character to another. This function is particularly relevant in the cases of *yo* and *él*. Ex. (9) *no os peleéis, que se lo cuento yo* (don't fight, I'll tell him). Ex. (10) *No hay tu tía, él ha dicho que viene* (No way, he said he'd come).

In this study, each personal pronoun has been considered separately, analysing its various pragmatic functions in original and translated Spanish. Whenever several functions co-occur, the dominant function has been selected for the analysis.

3.2 Corpora: CREA and P-ACTRES

The analysis relies on the combined use of data from a monolingual reference corpus in Spanish (CREA) and data from the English-Spanish parallel corpus P-ACTRES. CREA is a very large corpus sponsored by the *Real Academia Española* which includes around 175 million words of running texts in a wide range of different registers and geographic varieties of the Spanish language worldwide. In this paper we will focus on the fiction section of the European Spanish subcorpus only, which contains 2,379,249 words for the time range from 2000 until today.

P-ACTRES has been compiled at the University of León. It is an English-Spanish parallel corpus containing 2.5 million words of original English texts and their corresponding translations into peninsular Spanish. This parallel corpus includes written material from a variety of different registers (fiction, non-fiction, press, and 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. The English-Spanish parallel texts have been aligned at sentence level and can be searched with the Corpus Work Bench browser (CWB) (Izquierdo, Hofland and Reigem 2008). The contents of P-ACTRES are distributed as follows:

Table 1. Contents of the English-Spanish Parallel Corpus: number of words

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

As we will focus exclusively on Spanish translations of fiction, our empirical data will be extracted from the fiction subcorpus, which contains 421,065 words.

3.3 Statistics

Once the final samples of concordances have been extracted and classified according to their pragmatic functions, the numerical data obtained have to be interpreted. Statistics provide a welcome link between quantitative and qualitative empirical evidence as they help focus on those uses or functions that trigger cross-linguistic problems. 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 results are statistically significant or not.

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, which have been computed with the software MegaStat. Both indicators measure the difference between the data and what is expected under the null hypothesis: that both translated and non-translated grammatical usages are identical. Since the difference between two proportions follows a normal distribution when the sample size is big enough, the appropriate statistic for testing the null hypothesis is the z-score, whose significance is evaluated by the p-value. In other words, the test clarifies whether the differences found between the original and translated usages of pronouns are due to chance or not. This means that if any statistically significant differences are found, they can only be attributed to the fact that one sample contains originals and the other sample translations. Calculations have been done for a 95% confidence interval, with an estimated error of 5%. To be statistically significant, the difference between the two proportions (translated and non-translated cases) has to lie outside the curve ± 1.96 for the z-score (either higher than 1.96 or lower than -1.96) and the p-value must be lower than 0.05. Only the statistically significant differences have been considered for further analysis and interpretation.

4. Results

4.1. Numerical results

A preliminary numerical analysis has been carried out on the whole samples of original (2,379,249 words) and translated Spanish (421,065 words) in order to determine trends in overuse or underuse of personal pronouns in Spanish translations of English fiction. Because our two corpora differ in size, to make the data comparable, we have used the number of cases per million words of all the singular personal pronouns in original and translated Spanish, as shown in Figure 1.

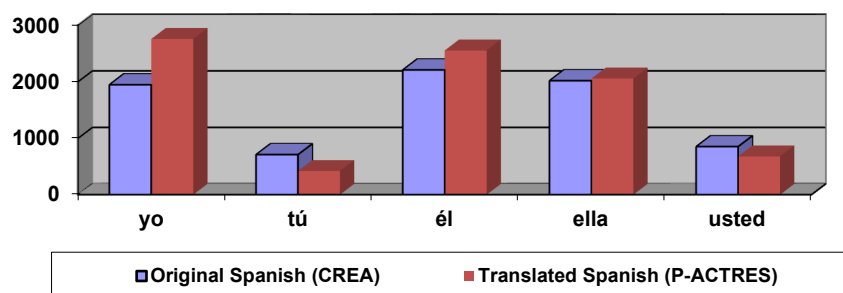


Figure 1. Number of cases per million words of singular personal pronouns in original and translated Spanish

The first thing that catches the eye in Figure 1 is the huge difference in frequency of use between the 1st and 3rd person singular pronouns (*yo*, *él* and *ella*), on the one hand, and the 2nd person singular pronouns (*tú* and the formality marker *usted*), on the other. This difference may be due to the fact that in fiction the 2nd person mainly occurs in dialogue, whereas the 1st and 3rd persons occur much more often in narration. Figure 1 also illustrates that *yo*, *él* and *ella* are overused, whereas *tú* and *usted* are underused in the translations. These results are consistent with the normalization hypothesis, which states that the typical uses of a particular item are boosted in translations (Laviosa 1998, 565): very common items are even more common in translations (*yo*, *él*, *ella*), and infrequent items are even less frequent in translations (*tú*, *usted*). To know whether these differences are statistically significant or not, we have used the raw number of cases of each pronoun to compute the z-score and p-value, as shown in Table 2 below.

Table 2. Statistical significance of the differences in the number of pronouns in original and translated Spanish

PRONOUNS	Population in CREA	Population in P-ACTRES	z-score	p-value
Yo	4,635	1,160	-10.6193	0.0000
Tú	1,691	177	-6.7266	0.0000
Él	2,927	625	-4.2703	0.0000
ella	4,804	866	-0.4998	0.6172
usted	1,675	198	5.4082	0.0000

P-values smaller than 0.05 mean statistically significant differences. As can be seen in Table 2 above, all personal pronouns, except *ella*, show statistically significant differences in the number of occurrences between original and translated Spanish. Consequently, *ella* is not an anchor phenomenon and has been excluded from further analysis.

Next, to extract a representative sample of concordances of each pronoun for the analysis, the following formula has been employed: $n=N/(N-1)E^2+1$. The element n is the final sample we will analyse, N is the whole sample of occurrences, and E is the estimated error, in this case 0.05 for a 95% confidence margin. This formula is used to reduce the sample for analysis to a manageable size, but always maintaining representativeness of the whole sample. For example, the fiction section of CREA contains 4,635 instances of *yo*, but using the formula we reduce the sample to 368 cases, which are sufficiently representative of the whole. The study focuses on the results obtained from the representative samples (n). Table 3 below shows the final number of concordances selected in each case.

Table 3. Total populations and representative samples of the pronouns analysed

Pronoun	Population in CREA (N)	Samples in CREA (n)	Population in P-ACTRES (N)	Samples in P-ACTRES (n)
yo	4,635	368	1,160	298
tú	1,691	324	177	177
él	2,927	352	625	244
usted	1,675	323	198	133

Once all the instances have been extracted, we have carried out the qualitative analysis of all the cases of each pronoun.

4.2 First person singular pronoun: *yo*

The 1st person singular pronoun in native 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. When it does occur in discourse, it may convey a variety of different functions. The preliminary analysis (Figure 1 and Table 2) has shown that there is a statistically significant overuse of the pronoun *yo* in Spanish translations from English when compared with original Spanish fiction. Table 4 below shows the quantitative distribution of the functions of the pronoun *yo* in original and translated Spanish fiction, together with information on the statistical significance of the difference in use found in each function:

Table 4. Quantitative data for *yo*

YO	CREA		P-ACTRES		Statistical significance	
	Raw cases	Percentage	Raw cases	Percentage	z-score	p-value
Contrastive	181	49.18%	50	16.77%	8.74	0.00E
Optional emphasis	79	21.46%	82	27.51%	-1.81	0.0698
Neutral	37	10.05%	62	20.80%	-3.88	0.001
Narrative discourse marker	36	9.78%	90	30.20%	-6.69	0.0025
Formulaic	20	5.43%	4	1.34%	2.82	0.0048
Non-optional emphasis	15	4.07%	10	3.35%	0.49	0.6267
TOTAL	368	100%	298	100%		

It can be noted that the distribution of the various functions identified differs considerably between original and translated Spanish. The main use of the 1st person pronoun in original texts, with nearly 50% of occurrences, is clearly the contrastive function, which emphasizes the difference between the 1st person speaker and some other person (2nd or 3rd): ex. (11) *usted dijo horror y yo terror* (you said horror and I terror); ex. (12) *él lo leía y yo le miraba de reojo* (he read it and I looked at him out of the corner of my eye). Curiously enough, this particular function was only found in 16% of cases in Spanish translations from English, as in ex. (13) *I'll fill it for you, and you take the flowers. / Yo te lo lleno y tú coges las flores*. This difference is statistically significant (z-score is over 1.96 and p-value is under 0.05) and indicates that other uses are favoured by translators.

The second most common function of *yo* in original Spanish is optional emphasis, with 21% of cases. From a translational perspective, this function poses no problem, since the frequency of occurrence in translations is very similar (27%) and the difference is not statistically significant (z-score is over -1.96 and p-value is over 0.05).

The neutral category refers to cases where the use of the personal pronoun is obligatory for grammaticality, as in coordinated constructions: *tú y yo*. This function occurs twice more often in translations than in originals, and this difference is statistically significant. We have noticed that this often occurs when the source item in English *we* is split up into *me and you* in Spanish: ex. (14) *We haven't a hope. / Tú y yo no tenemos futuro.*

The use of the 1st person singular pronoun *yo* with a narrative function occurs in only 10% of cases in original Spanish. 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:

ex. (15) *Puri: Y la atamos.* (And we tied her.)

Toña: Y yo le di una hostia, aunque no me arrepiento. (And I slapped her, though I do not regret it.)

This particular 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. This difference is statistically significant too. An example is: (16) *It was in the day or two before the lineup and so it was all I was thinking about. / Era un par de días antes de la rueda de identificación y yo no podía pensar en otra cosa.*

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 which include the pronoun with a clearly pragmatic meaning. This function occurs more often in original texts (5%) than in translations (1%), which is an expected result, since pragmatic uses are more difficult to convey in translations. Examples from CREA are (17) *yo qué sé* (I have no idea), (18) *un qué sé yo* (a what-do-I-know), (19) *ya lo dije yo* (I said so), etc. An example from P-ACTRES is (20) *I was trying to get better and it ruined my life, with the shakes and panics and what not. / Yo estaba intentando recuperarme y me destrozó la vida a base de temblores y ataques de pánico y yo qué sé qué más.*

The function termed non-optional emphasis occurs slightly more often in originals than in translations, but this difference is not statistically significant.

4.3 Second person singular pronoun: *tú*

As shown in Figure 1 and Table 2 above, the unmarked form of the 2nd person singular pronoun *tú* is four times less common in original Spanish texts than the 1st or the 3rd person pronouns, as it mainly occurs in dialogue. This preliminary analysis has shown that there is a statistically significant underuse of the pronoun *tú* in Spanish translations. All the instances of *tú* have been analysed and classified into various functions. The results are shown in Table 5 below:

Table 5. Quantitative data for *tú*

TÚ	CREA		P-ACTRES		Statistical significance	
	Raw cases	Percentage	Raw cases	Percentage	z-score	p-value
Contrastive	108	33.33%	66	37.28%	-0.89	0.3742
Optional emphasis	103	31.79%	57	32.20%	-0.09	0.9244
Neutral	45	13.88%	35	19.77%	-1.72	0.856
Formulaic	29	8.95%	3	1.69%	3.17	0.0015
Non-optional emphasis	25	7.71%	11	6.21%	0.62	0.5339
Non-native speakers	6	1.85%	0	0%	1.82	0.0685
Narrative discourse marker	5	1.54%	4	2.25%	-0.58	0.5637
Generic role	3	0.92%	0	0%	1.28	0.1991
Metalinguistic use	0	0%	1	0.56%	-1.35	0.1756
TOTAL	324	100%	177	100%		

A first overview of Table 5 shows that the functions of *tú* follow a similar trend in the two subcorpora, original and translated Spanish, with few divergences. The only statistically significant difference identified between original and translated uses of *tú* is

the one corresponding to the formulaic use with pragmatic functions, which occurs in nearly 10% of cases in originals and only in 1% of cases in translations. The formulaic expression with *tú* is particularly versatile in original Spanish, with several examples occurring more than once and a total of 29 cases in CREA: ex. (21) *tú tranquilo* (don't you worry); ex. (22) *no te lo crees ni tú* (you are kidding), etc. The smaller frequency of occurrence of this function in translations relates to its pragmatic nature and its specificity in the Spanish language. This phenomenon could be referred to as 'unique-function hypothesis,' as it is closely related to the translation universal known as 'unique-item hypothesis,' which states that peculiar lexical items in a particular language tend to be absent or nearly absent from translations in that same language (Tirkkonen-Condit 2004, 183). An example from P-ACTRES is (23) *No, fuck you. / No, al cuerno te vas tú.*

None of the remaining functions present statistically significant differences between original and translated texts. This pronoun shows three peculiar uses in CREA, which are not present in any of the other pronouns: six examples correspond to a non-native speaker making grammatical mistakes, as in ex. (24) *Voy a contar de manera que tú entiendes seguro* (I going to tell so you sure understand); three examples correspond to the generic role of *tú*, as in ex. (25) *Pienso que lo bueno de la ciudad es que tú colocas un búho en mitad de la calle de Huertas...* (I think that the good thing about the city is that you put an owl in the middle of the Huertas street ...); and one more case is a metalinguistic use of the pronoun: ex. (26) *En alemán hay varias formas de decir tú* (In German there are several ways of saying you).

4.4 Third person singular pronoun masculine: *él*

The preliminary analysis in Figure 1 and Table 2 above has shown that, as in the case of *yo*, the pronoun *él* is significantly overused of in Spanish translations from English when compared with original Spanish texts. Table 6 below presents the distribution of functions of *él* in original and translated fiction, together with the statistical significance of the differences found:

Table 6. Quantitative data for *él*

ÉL	CREA	P-ACTRES	Statistical
----	------	----------	-------------

Functions					significance	
	Raw cases	Percentage	Raw cases	Percentage	z-score	p-value
Narrative discourse marker	153	43.46%	99	40.57%	0.70	0.4822
Contrastive	97	27.56%	97	39.75%	-3.12	0.0018
Non-optional emphasis	43	12.22%	17	6.97%	2.09	0.0363
Neutral	42	11.93%	23	9.43%	0.96	0.3346
Optional emphasis	13	3.69%	7	2.87%	0.55	0.5827
Formulaic	4	1.14%	1	0.41%	0.96	0.3390
TOTAL	352	100%	244	100%		

The analysis reveals that *él* is mostly used as a narrative discourse marker both in non-translated and translated Spanish (43.46% and 40.57%, respectively). The difference in this function is not statistically significant. In contrast, translated usage shows significant variation with respect to native usage in the contrastive and non-optional emphasis functions.

In the contrastive function, *él* is opposed to another personal pronoun in the same sentence, as in the following example from CREA: (27) *luego, llamaba a mi esposo, él lo leía y yo le miraba de reojo* (then, I would call my husband, he would read it and I would look at him out of the corner of my eye). The contrastive function is overused in Spanish translations of original texts in English, with nearly 40% of cases, whereas the frequency of occurrence in the native corpus was of only 27%. An example from the translation corpus is (28) *Arthur gave him the list and he read through it. / Arturo le entregó la lista y él la leyó*. This overuse may be due to an attempt to explicitate information in the target texts on the part of the translator, illustrating the explicitation hypothesis, which has long been recognized as a translation universal.

On the other hand, the function non-optional emphasis is underused in the translations. This function occurs when the pronoun is required to maintain discursive coherence: ex. (29) *¿No fue él quien le pidió que me recibiera en su nombre?* (Wasn't

he the one who asked you to receive me in his name?) Underuse of this function indicates a divergence between original and translated texts, mainly in the field of cohesion: ex. (30) “*He’s the sneak, isn’t he?*” / *Porque es él quien le ha venido con el cuento, ¿me equivoco?*

The remaining functions identified for *él* did not show statistically significant differences between original and translated texts.

4.5 Second person singular formal pronoun: *usted*

As shown in Figure 1 and Table 2 above, the formality marker *usted* seems to follow the same pattern as the 2nd person singular non-formal pronoun *tú*: overall, it presents statistically significant underuse in Spanish translations. This may be partly due to the fact that, since source texts are always in English, there is no functional equivalent (no 2nd person singular pronoun for marking formality) in that language to prompt the use of *usted* in Spanish translations. Table 7 shows the deictic, pragmatic and textual functions identified for the pronoun *usted* in our corpora of original and translated Spanish.

Table 7. Quantitative data for *usted*

USTED	CREA		P-ACTRES		Statistical significance	
	Raw cases	Percentage	Raw cases	Percentage	z-score	p-value
Optional emphasis (Marker of formality)	217	67.18%	74	55.63%	2.33	0.0197
Contrastive	44	13.62%	19	14.28%	-0.19	0.8520
Narrative discourse marker	28	8.66%	7	5.26%	1.24	0.2143
Formulaic	18	5.57%	8	6.01%	-0.19	0.8531
Non-optional emphasis	8	2.47%	17	12.78%	-4.39	0.0000

Neutral	7	2.16%	8	6.01%	-2.09	0.0363
TOTAL	323	100%	133	100%		

It is significant that original Spanish makes a heavy use of *usted* as a marker of formality (67.18%), as in ex. (31) *He pensado que usted podría ayudarme* (I thought you could help me), or as a way of clarifying the reference through the contrastive function (13.62%), as in ex. (32) *A todos los efectos, usted y yo no nos conocemos* (For all intents, you and I, we don't know each other). The remaining cases are less frequent.

From a translational perspective, only three functions showed statistically significant differences in the translation corpus. In Spanish translations, *usted* is underused as a marker of formality: ex. (33) *You know me; the woman called. / Usted sabe quien soy, me conoce - dijo la mujer*. In contrast, *usted* is overused when meaning non-optional emphasis, as in ex. (34) *And who are you, anyway? / ¿Quién es usted?*, and with neutral reference, as in ex. (35) *You didn't have kids. / Usted no tenía hijos*.

5. Discussion

5.1 Spanish personal pronouns as anchor phenomena

What is most noteworthy about the differences identified in the anchor phenomenon of singular personal pronouns in non-translated and translated Spanish is not the size of the significant results, but their effects on language use (Rabadán 2008). Divergences from original Spanish with respect to pragmatic and textual functions may not always lead to unintelligibility of the target text, but they may certainly have consequences on the general perception of a particular text on the part of the readership. Table 8 shows the most useful values, in the form of anchor phenomena, to describe divergences between original and translated Spanish:

Table 8. Anchor uses for *yo*, *tú*, *él* and *usted*

ANCHOR USES	YO	TÚ	ÉL	USTED
Neutral	overuse			overuse
Optional emphasis				underuse
Non-optional emphasis			underuse	overuse

Contrastive	underuse		overuse	
Formulaic	underuse	underuse		
Narrative reference marker	overuse			

The 1st person singular pronoun *yo* seems to be the one with more divergences in use, with overuse of the narrative and neutral functions and underuse of the contrastive and formulaic functions in translations. In contrast, the 2nd person singular non-formal pronoun *tú* presents uses that are virtually identical in original and translated texts. The only significant difference lies in the underuse of the formulaic function of this pronoun, which is 10 times more frequent in original texts than in translations. The 3rd person singular masculine pronoun *él* presents significant differences in two values: contrastive, which is overused in the translations, and non-optional emphasis, which shows underuse in translated Spanish. Finally, significant differences have been verified for three values of the formal marker *usted*: optional emphasis is significantly underused in translation, whereas non-optional emphasis and neutral reference are overused in the translations. While all these divergences between original and translated Spanish do not, in principle, interfere with comprehension, they do make themselves noticeable in the target texts by reducing their idiomaticity.

5.2 The translation assessment and post-editing tool PETRA

The list of problematic usages which constitute the anchor phenomenon of personal pronouns proves to be very useful for identifying differences between original and translated texts. The full descriptive procedure outlined above is, therefore, adequate for identifying those linguistic areas that will benefit from in-depth revision, providing reliable guide throughout the post-editing process. Because some of these differences may make translations less readable, these results on personal pronouns – together with other anchor phenomena (demonstratives, modal verbs, verbal tenses, -ly adverbs, negation, adjective-noun combinations, etc.) – have been implemented into a computerized tool for assessing translation quality and for the post-editing of Spanish translations of original texts in English (PETRA 1.0©, LE-222-13) (Rabadán et al. 2014). This tool enables us to review Spanish translations without reference to the source texts, focusing our attention on those areas where it has been proven that there are statistically significant divergences between original and translated texts.

Most post-editing tools available today in the market have been designed specifically for post-editing machine translation output, not human translations. The majority of these toolkits are part of machine translation systems or translation memory systems (Trados, Déjà Vu, Systran, Wordfast, Google Translator Toolkit, Caitra, Lingotek, OmegaT, etc.). These tools all enable translators to review translations from and into a large variety of languages using spell-checkers and quality-assurance devices. These quality-assurance devices highlight terminology inconsistencies and problems related to untranslated segments or spacing (Vieira and Specia 2011). None of these toolkits highlights any particular grammatical aspects the reviewer might change to improve the quality of the translation, which is precisely what PETRA does for the language pair English-Spanish.

The interface of the translation assessment and post-editing tool PETRA is in Spanish and English, but the intended users need not have a particularly good command of the source language English to use this tool. A native or native-like command of the Spanish language, however, is required as the target language to be post-edited is Spanish. PETRA assesses the grammatical quality of Spanish translations by highlighting those linguistic aspects that could be improved in an eventual post-editing process.

PETRA offers two different types of analysis: first, a basic quantitative analysis detects statistically significant differences in the frequency of occurrence of certain anchor phenomena in the Spanish translations. This analysis is carried out automatically and requires no human intervention at all. The information provided by the tool will indicate that there is a clear overuse or underuse of certain items and the reviewer may then start a post-editing process to improve the translation by following these suggestions. Second, a more advanced qualitative analysis requires the user to clarify the contextual usage of some other anchors, so that these data may also be considered in the post-editing procedure, if needed. The Spanish personal pronouns belong to this qualitative set of anchors which requires human intervention.

In the case of this particular anchor phenomenon of personal pronouns, the procedure followed in using the tool is as follows: in the advanced analysis the application highlights all the instances of *yo*, *tú*, *él* and *usted* in the text to be revised. An automatic message pops up next to each pronoun with a yes/no question and a short and clear example of one of the statistically significant functions listed in 2.1. The reviewer has to decide on whether the use of the pronoun in the translation is similar or

not to the one in the pop up window. No extensive metalinguistic knowledge on the part of the user is required to be able to answer these questions and this is why no technical terms are used. In fact, as the intended users are mainly post-editors or translation reviewers, sufficient linguistic expertise is assumed. Depending on the answer given, the tool will automatically assign that particular example to one category or another. For example, in the case of *tú*, the question is: *¿Es una frase hecha?* (is it an idiomatic expression?), and one of the examples is: *Vete tú a saber. SÍ / NO*. If the user considers that the instance highlighted in the text is similar to the one given in the example, he/she will click on the *SÍ* button and answer yes, and this example will be counted as a formulaic use of the pronoun *tú*. If the answer is *NO*, it will be ignored, as only the formulaic use is statistically relevant from a translational perspective in this particular pronoun. Figure 2 shows the interface of PETRA while analysing the second person singular Spanish pronoun *tú*, together with the corresponding pop up message for the qualitative analysis the reviewer must carry out.



Figure 2. Screenshot of PETRA analysing the second person singular pronoun *tú*

The same procedure will be followed with all the other personal pronouns, and with all the other anchor phenomena, until the application has gathered all the details

concerning the divergences, if any, between the text being assessed and original Spanish.

After all the anchor phenomena have been analysed, a final grading is provided on a scale from 0 to 5, where 0 identifies a translation which differs greatly from native usage of Spanish, and 5 a translation which is grammatically very similar to native usage of Spanish. PETRA also provides a short review of the translation, derived from the score achieved, as well as the list of grammatical anchor phenomena which showed deviations from native usage, and would consequently benefit from a careful revision and post-editing process. The final report can then be downloaded in pdf format. A video demo in English describes in detail how this computerized tool works: http://actres.unileon.es/?page_id=48&lang=es. In 2013 PETRA won the first prize in the 4th Prototype Contest organized by the University of León, Spain. The tool has been registered for intellectual property, but has not been made available commercially yet. A new version of PETRA is currently being developed, including more anchor phenomena.

The procedure described here shows how a careful analysis of empirical translation data can provide useful information for improving post-editing techniques, and how these data may be implemented in an applied computerized tool for the actual practice of translation assessment and post-editing.

6. Conclusions

For anyone who has ever practised translation in a systematic way, the post-editing stage is probably the most tedious and less rewarding part of the whole translation process. However, it cannot be ignored that the quality of the final output greatly depends on the thoroughness of this final stage. Existing post-editing guidelines are mostly based on personal experiences of professional translators, but are not generally grounded on larger corpora, and existing post-editing tools are too general and not language-pair based. In this paper the contrastive analysis of empirical data extracted from a monolingual reference corpus and a translation corpus has demonstrated that a number of descriptive anchor values, adequately conceptualized, can provide a much more reliable and systematic tool for translation assessment and post-editing.

The use of corpus data and relevant statistics inhibit the tendency to search for interpretations that confirm unverified views on the basis of raw quantitative data. The

results put forward in this paper constitute a transition from quantitative data to relevant information usable in post-editing (Rabadán 2010). Focusing on grammatical uses that have been empirically proven to cause distortion in translations of fiction from English into Spanish can be said to be a real help for the improvement of the linguistic quality of translated texts in post-editing procedures.

The results of our description corroborate the working hypothesis we put forward at the beginning: personal pronouns are indeed reliable indicators of divergences between original and translated Spanish. Empirical data have been extracted from a large reference corpus of Spanish (CREA) and an English-Spanish parallel corpus (P-ACTRES) to identify statistically significant differences in the distribution of the various functions of personal pronouns in original and translated Spanish. Since the analysis has been based exclusively on texts from the fiction subcorpora, the results are applicable to the post-editing process of translated fiction in particular. This procedure has demonstrated that the anchors we are proposing are truly useful for the pair English-Spanish. Our results indicate varying degrees of textual and linguistic deviations, mainly a defective management of the emphatic, contrastive, and formulaic uses of the pronouns, which affects text progression, has pragmatic implications, and may even hamper intelligibility in the translations. Even though these slight changes do not normally imply ungrammaticality, they do illustrate that minor nuances of pragmatic and textual functions are conveyed differently in the target language. The consequences are that the perception of the authors and their original work in English by their Spanish-language readers can be negatively affected, leading thus to low commercial and market figures.

However, we must always take into account that the style of the author of the source text may present certain peculiarities which may trigger particular deviations from the standard of acceptability in the translation. It is therefore important to remember that the procedure used in this paper does not necessarily evaluate a particular translation as good or bad, but rather determines up to what extent the translation looks more or less like a text written originally in the target language, in this case Spanish.

This study has focused on one single anchor phenomenon, namely, personal pronouns. However, the post-editing procedure benefits from a larger number of quantitative and qualitative anchor phenomena, which cover other grammatical divergences which may occur in this particular language pair. The ACTRES project has

proposed a number of guidelines in various areas, all of them developed following a similar procedure to the one described here for pronouns, including modal verbs (Rabadán 2006), adverbs (Ramón and Labrador 2008), verbal tenses (Rabadán 2009), demonstratives (Labrador 2011), negation (Rabadán and Izquierdo 2013), adjective-noun combinations (Ramón 2015), etc. All these data have been used to build PETRA, a computerized tool for the assessment and post-editing of Spanish translations of English originals.

Acknowledgements

Research for this article has been undertaken as part of the ACTRES (English-Spanish Contrastive Analysis and Translation in its Spanish acronym) and TRACE (Translation and Censorship) programs, both currently in progress at the University of León, Spain. The research has been partly funded by the Spanish Ministry of Economy and Competitiveness (FFI2016-75672-R, FFI2013-42994-R, FFI2012-39012-C04-03) and the regional government Junta de Castilla y León (LE227U13).

References

- Alarcos, Emilio. 1980. *Estudios de gramática funcional*. Madrid: Gredos.
- Baker, Mona. 1993. "Corpus Linguistics and Translation Studies: Implications and Applications." In *Text and Technology. In Honour of John Sinclair*, edited by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 233-250. Amsterdam: John Benjamins.
- Chesterman, Andrew. 1998. *Contrastive Functional Analysis*. Amsterdam: John Benjamins.
- Eldredge, Dee L., and Liesder Mayea Rodríguez. 2015. *Introduction to Spanish/English Linguistics*. Bloomington, IN: Xlibris.
- Enríquez, Emilia Victoria. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: CSIC.

- Fernández Soriano, Olga M. 1989. *Rección y ligamiento en español: aspectos del parámetro del sujeto nulo*. PhD diss. Universidad Autónoma de Madrid.
- Halliday, M.A.K., and Ruqaiya Hassan. 1976. *Cohesion in English*. London: Longman.
- House, Juliane. 2001. "Translation Quality Assessment. Linguistic Description versus Social Evaluation." *Meta* 46 (2): 243-257.
- Huddleston, Rodney, and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Izquierdo, Marlén, Knut Hofland, and Øystein Reigem. 2008. "The ACTRES Parallel Corpus: An English-Spanish Translation Corpus." *Corpora* 3: 31-41.
- Krings, Hans P. 2001. *Repairing Texts: Empirical Investigations of Machine-Translation Post-Editing Processes*. Kent, OH: Kent State University Press.
- Krzyszowski, Tomasz P. 1990. *Contrasting Languages. The Scope of Contrastive Linguistics*. Berlin: Mouton de Gruyter.
- Labrador, Belén. 2011. "A Corpus-Based Study of the Use of Spanish Demonstratives as Translation Equivalents of English Demonstratives." *Perspectives. Studies in Translatology* 19 (1): 71-87.
- Laviosa, Sara. 1998. "Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose." *Meta* 43 (4): 557-570.
- López Guix, Juan Gabriel, and Jacqueline Minett Wilkinson. 2009. *Manual de traducción inglés-castellano: teoría y práctica*. Barcelona: Gedisa.
- Luján, Marta. 1999. "Expresión y omisión del pronombre personal." In *Gramática descriptiva de la lengua española*, edited by Ignacio Bosque and Violeta Demonte, 1275-1315. Madrid: Espasa.
- Maia, Belinda. 1998. "Word Order and the First Person Singular in Portuguese and English." *Meta* 43 (4): 589-601.
- Marcos Marín, Francisco. 1978. *Estudios sobre el pronombre*. Madrid: Gredos.
- Martínez Vázquez, Montserrat. 1996. *Gramática contrastiva inglés-español*. Huelva: Universidad de Huelva.
- O'Brien, Sharon, Laura Winther Balling, Michael Carl, Michel Simard, and Lucia Specia, eds. 2014. *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Rabadán, Rosa. 2005. "The Applicability of Description. Empirical Research and Translation Tools." In *Contemporary Problematics of Translation Studies*,

- edited by Carmen Toledano, special issue of *Revista Canaria de Estudios Ingleses* 51: 51-70.
- Rabadán, Rosa. 2006. "Modality and Modal Verbs in Contrast. Mapping out a Translation(ally) Relevant Approach English-Spanish." *Languages in Contrast* 6 (2): 261-306.
- Rabadán, Rosa. 2008. "Refining the Idea of 'Applied Extension'." In *Beyond Descriptive Translation Studies*, edited by Anthony Pym, Miriam Shlesinger, and Daniel Simeoni, 103-117. Amsterdam: John Benjamins.
- Rabadán, Rosa. 2009. "The Present Tenses in English and in Spanish: A Corpus-Based Approach to Cross-linguistic Meaning and Grammatical Transfer." In *Cognitive Approaches to Language and Linguistic Data*, edited by Wiesław Oleksy and Piotr Stalmaszczyk, 461-478. Frankfurt am Main: Peter Lang.
- Rabadán, Rosa. 2010. "Applied Translation Studies." In *Handbook of Translation Studies*. Volume 1, edited by Yves Gambier, and Luc van Doorslaer, 7–11. Amsterdam: John Benjamins.
- Rabadán, Rosa, Belén Labrador, and Noelia Ramón. 2009. "Corpus-Based Contrastive Analysis and Translation Universals: A Tool for Translation Quality Assessment English→Spanish?" *Babel* 55 (4): 303-328.
- Rabadán, Rosa, and Marlén Izquierdo. 2013. "A Corpus-Based Study of English Approximate Negators and their Translations into Spanish." *BeLLs* 3 (1): 43-59.
- Rabadán, Rosa, Camino Gutiérrez Lanza, Ana Belén García Gallego, and Hugo Sanjurjo González. 2014. "PETRA 1.0©: A Semi-automated Aid for the Evaluation of the Grammatical Quality of English-to-Spanish Translations." Unpublished paper. *Mediterranean Editors and Translators Meeting 2014. Innovation and Tradition: Mining the Human Resource*. Euroforum, San Lorenzo de El Escorial. 30th October – 1st November 2014.
- Ramón, Noelia. 2015. "Comparing Original and Translated Spanish: A Corpus-Based Analysis of Adjective Position." *Babel* 61 (4): 527-551.
- Ramón, Noelia, and Belén Labrador. 2008. "Translations of '-ly' Adverbs of Degree in an English-Spanish Parallel Corpus." *Target* 20 (2): 275-296.
- Somers, Harold. 1997. "A Practical Approach to Using Machine Translation Software: 'Post-editing the Source Text'." *The Translator* 3: 193-212.

- 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?*, edited by Anna Mauranen and Pekka Kujamäki, 65-82. Amsterdam: John Benjamins.
- Toury, Gideon. (1995) 2012. *Descriptive Translation Studies and Beyond*. Revised edition. Amsterdam: John Benjamins.
- Toury, Gideon. 2004. "Probabilistic Explanations in Translation Studies: Universals – or a Challenge to the Very Concept?" In *Claims, Changes and Challenges in Translation Studies*, edited by Gyde Hansen, Kirsten Malmkjær, and Daniel Gile, 15-26. Amsterdam: John Benjamins.
- Vázquez Ayora, Gerardo. 1977. *Introducción a la traductología*. Washington D.C.: Georgetown University Press.
- Vieira, Lucas Nunes, and Lucia Specia. 2011. "A Review of Translation Tools from a Post-editing Perspective." In *Proceedings of the Third Joint EM+/CNGL Workshop Bringing MT to the User: Research Meets Translators (JEC 2011)*, 33-42. Luxemburg.
- Whitley, M. Stanley. 2002. *Spanish/English Contrasts. A Course in Spanish Linguistics*. Washington D.C.: Georgetown University Press.

Authors' addresses

Noelia Ramón
 Universidad de León
 Facultad de Filosofía y Letras
 Campus de Vegazana, s/n
 24007 LEÓN
 España

noelia.ramon@unileon.es

Camino Gutiérrez-Lanza
 Universidad de León
 Facultad de Filosofía y Letras
 Campus de Vegazana, s/n
 24007 LEÓN
 España

camino.gutierrez.lanza@unileon.es

