The Portrait of Dorian Gray: A Corpus-Based Analysis of Translated Verb + Noun (object) Collocations in Peninsular and Colombian Spanish¹

M.^a Victoria Valencia Giraldo¹ and Gloria Corpas Pastor²

¹University of Salamanca, Salamanca, Spain

victoriavalencia@usal.es

²University of Malaga, Spain

University of Wolverhampton, United Kingdom

gcorpas@uma.es

Abstract. Corpus-based Translation Studies have promoted research on the features of translated language, by focusing on the process and product of translation, from a descriptive perspective. Some of these features have been proposed by Toury (1995/2012) under the term of *laws of translation*, namely the law of growing standardisation and the law of interference. The law of standardisation appears to be particularly at play in diatopy, and more specifically in the case of transnational languages (e.g. English, Spanish, French, German). In fact, some studies have revealed the tendency to standardise the diatopic varieties of Spanish in translated language (Corpas Pastor, 2015a, 2015b, 2017, 2018). This paper focuses on verb + noun (object) collocations of Spanish translations of *The Portrait of Dorian Gray* by Oscar Wilde. Two different varieties have been chosen (Peninsular and Colombian Spanish). Our main aim is to establish whether the Colombian Spanish translation actually matches the variety spoken in Colombia or it is closer to general or standard Spanish. For this purpose, the techniques used to translate this type of collocations in both Spanish translations will be analysed. Furthermore, the diatopic distribution of these collocations will be studied by means of large corpora.

Keywords: Corpus-based Translation Studies, diatopic variation, collocations, standardisation.

1 Introduction

Corpus-based Translation Studies have promoted research on the features of translated language, by focusing on the process and product of translation, from a descriptive perspective. Some of these fea-

¹ The final publication is available at <http://link.springer.com>.

tures or regularities have been proposed by Toury (1995/2012) under the term of laws of translation, namely the law of growing standardisation and the law of interference. According to Toury's law of growing standardisation (2012), "in translation, source-text textemes tend to be converted into target-language (or target-culture) repertoremes". In other words, the complicated textual relations established in the source text are replaced in the target text by more common ones in the target discursive-textual community. Thus, translated texts tend to be more conventional and standardised than non-translated texts. Another related notion was introduced by Baker (1996) under the umbrella term 'translation universals' in order to encompass several common features of translated texts (simplification, explicitation, normalisation, convergence and transference). Based on Toury's law, some research has been carried out on the tendency to standardise the diatopic varieties of Spanish in translated language by means of collocations and other phraseological units (Corpas Pastor, 2015a, 2015b, 2017, 2018). In general, it can be inferred from those studies that the richness of the different Spanish national varieties is not reflected in translated language, nor in the existing bilingual dictionaries. However, to the best of our knowledge, no study has already determined whether the translation of collocations in a literary work reflects a specific Spanish variety or whether it is closer to general or standard Spanish².

In this paper, we argue that the translation of verb + noun (object) collocations found in the first two chapters of *The Portrait of Dorian Gray* by Oscar Wilde in a translation into Colombian Spanish appears closer to general or standard Spanish than to the Spanish variety of Colombia. In order to demonstrate this, the techniques used to translate this type of collocations in a translation into Peninsular Spanish³ (Wilde 1890/1999) and in another one into Colombian Spanish (Wilde 1890/1994) will be analysed. Moreover, the diatopic distribution of these Spanish collocations will be studied by means of large corpora. As a convenient theoretical background, we endorse Toury's laws, as well as previous studies on standardisation of diatopic varieties of Spanish in translation.

2 Translation of collocations and diatopic varieties of Spanish

The term collocation is understood as the property of language by which speakers tend to produce certain combinations of words among a large number of possible combinations (Cf. Haensch et al., 1982: 251). From a linguistic point of view, collocations are phraseological units which are free syn-

² General or standard Spanish is understood here as 'global' or 'unified' Spanish spoken/written all over the Spanish-speaking world (see Paffey, 2012).

³ The term Peninsular Spanish will be used to refer to Spanish from Spain, including the islands, Canary and Balearic Islands, Ceuta and Melilla.

tagms generated from rules, although, at the same time, they show different ranges of combinatorial restriction conditioned by use (Corpas Pastor, 1996: 53). A full discussion on the notion of collocation is beyond the scope of this work. We will present the main features we took into consideration to conduct this study.

Collocations are considered composite units and conventionally restricted structures, in which both collocates have a different semantic status (Hausmann, 1989). For example, in *to make a decision* and *tomar una decisión*, nouns constitute the autosemantic bases, while *to make* and *tomar* are collocates, i.e., the meanings of these depend entirely on their respective bases.

As far as verb + noun (object) collocations are concerned, verbs (as collocates) in these combinations exhibit a variety of ranges of collocability: from practically unlimited to practically fixed with intermediate categories (Corpas Pastor, 1996: 68-69). In the first type, we find collocations which share collocates and a base of the same semantic field, for instance *desempeñar un cargo*, *una función* or *un papel*, in Spanish. On the opposite extreme of the spectrum, we find practically fixed or very restricted collocations, such as *conciliar el sueño*, *acariciar una idea*, *grind one's teeth*, or *shrug one's shoulders*, whose collocates exhibit a figurative meaning or have a narrow or specific meaning (Aisenstadt 1979 in Howarth 1996: 39). Among the intermediate cases (*asumir una responsabilidad*, for example), a homogeneous group is formed by a delexical verb (or, rather, a verb used in a delexical sense, Howarth 1996), almost grammaticalised, and a noun, generally deverbal, which provides the fundamental meaning. Delexical verbs are highly polysemous verbs, such as *dar*, *tomar*, *hacer*, or *poner*; and *take* and *make*, etc., in English.

Collocations express typical relations between components, as can be observed in noun + verb and verb + noun collocations (Koike, 2001). *Tocar la guitarra* and *rasguear la guitarra* are collocations, due to the typical relation between its components (ibid.). On the contrary, *limpiar la guitarra* and *guardar la guitarra* are not considered collocations, because the noun *guitarra* can only express a typical relation as a musical instrument (ibid.).

By repeating a combination of words, speakers recognise it as familiar and use it as if it were a pre-fabricated fragment (Corpas Pastor, 2001). One way to determine whether a certain sequence is sufficiently frequent or salient to be considered conventionalised, habitual or typical is to count its occurrences in a corpus, as suggested by Manning and Schütze (1999). A large number of occurrences is enough to consider that a collocation is frequent, and therefore entrenched in the mind of speakers. Although, a significant collocation could be considered typical and cognitively salient (entrenched), even if it might not strike as particularly frequent (Corpas Pastor, 2015b: 6).

As Koike (2001) points out, lexical units that form a collocation do not necessarily have to be followed one after the other in a text. In some cases, one or more words appear in the middle of them.

Some authors (Sinclair 1966; Jones and Sinclair 1974) have proposed the concept of *collocational span* of four words before and after the keyword or node, but there are collocations in which constituent elements are separated from each other by more than four words (Koike 2001).

Collocations tend to be semantically transparent, but not always predictable (Corpas Pastor, 2015b, 2017). Unlike idiomatic phraseological units (idioms), collocations generally do not pose comprehension problems, but mainly in production, especially in second language learning, languages for academic/specific purposes, translation, and interpretation (Corpas Pastor, 2017, 2016).

Collocations are not only peculiar to a particular language, but they also reveal variation in their collocational patterns when it comes to a particular text or literary style (see Hardy, 2007), an author style (see Hoover, 2003), domain-specific genres and registers (Biber and Conrad, 1999; Gledhill, 2000; Williams, 2002, Corpas Pastor, 2015a, etc.). They are also useful to characterise linguistic change over time (see Hilper, 2006), and diatopic variation (see also Corpas Pastor, 2015b).

The translation of collocations into the target language is often characterised by the anisomorphism of languages when it comes to choosing collocates (Corpas Pastor, 2015b, 2017). While bases are usually translated literally, collocates vary enormously. For example, *to pay homage* cannot be translated into Spanish as **pagar homenaje*, but, generally, as *rendir homenaje*. In general, the translator should try to make that the translation of collocations sounds as natural as possible in the target language, except, of course, when the function of a collocation is precisely generating surprise or draw the attention of the recipients, as frequently occurs in poetry, advertising and humor (Baker, 2011).

In the case of transnational languages, such as English or Spanish, the rendering becomes even more complicated, since it is necessary to choose the language variety into which to translate (Corpas Pastor, 2015b). Let us illustrate this again with the collocation *to pay homage*. Spanish offers a wide range of collocates on a scale of formality and grammaticalisation, the most frequent of which are: *dar/hacer un homenaje* (informal), *rendir/tributar/brindar un homenaje* (formal). However, this list is further restricted if diatopic variation is taken into consideration. For instance, *dar* y *brindar* are not plausible verbal collocates for *homenaje* in Nicaraguan Spanish; in Mexican and Argentinean Spanish *dar* is not used, but *hacer* as delexical verb with *homenaje*. Dominican Spanish only allows collocations with *rendir* and *tributar*; in Honduran Spanish the only option is *rendir*, etc., (ibid).

Translating diatopy is not a small or trivial task, and bilingual dictionaries are not of much help either, as they tend to maintain a simplified and conventional approach to the translation of collocations. In fact, there is currently no Spanish-English bilingual collocations dictionary (Corpas Pastor, 2016, 2017), and in those where equivalents of collocations are included, e.g. Collins, Larousse and Oxford, *rendir homenaje* is offered as the only translation equivalent of *to pay homage* (Corpas Pastor, 2015b), which does not at all reflect the richness of the Spanish language. However, according to

Corpas Pastor (2015b), the varieties of transnational languages, like Spanish, should be taken into consideration in order to increase the fluency and naturalness of translated texts.

The tension between standard Spanish and transnational Spanish (pluricentric) has consequences for the type of Spanish (transnational variety) used in translation. For commercial reasons, the translation industry has adopted the presumed neutral Spanish, as opposed to the Spanish located for a specific market (diatopic variety) (Corpas Pastor, 2018). But this variety, far from being neutral, is loaded with a positive sociocultural value and prestige, in opposition to other diatopically marked varieties.

In the following section, the corpora used in this study case are described. Then, the analysis of the translation of collocations and their diatopic distribution in corpora are presented.

3 Choice of corpora

There are some diachronic corpora of Spanish documents on the web nowadays, such as *Corpus del Español* (CE)⁴, *Corpus de Documentos Españoles Anteriores a 1800* (CODEA)⁵, *Corpus Biblia Medieval* (BM)⁶, *Corpus Hispánico y Americano en la Red: Textos Antiguos* (CHARTA)⁷, *Corpus Diacrónico del Español* (CORDE) and *Corpus Diacrónico y Diatópico del Español Americano* (CORDIAM). For this study, the following diachronic corpora have been selected: the *Corpus Diacrónico del Español* (CORDE)⁸, (Diachronic Corpus of Spanish), which contains American Spanish varieties and Peninsular Spanish (available through the Spanish Royal Academy); and the *Corpus Diacrónico y Diatópico del Español de America* (CORDIAM)⁹ (Diachronic and Diatopic Corpus of American Spanish), which is accesible through the Mexican Academy of the Language.

CORDE was created by the Royal Spanish Academy in the late 1990s. It was the first historical corpus of Spanish. It is a corpus of texts from the all places where Spanish is spoken (23 countries): Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, USA, Spain, Philippines, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Portugal, Puerto Rico, Dominican Republic, Uruguay, Venezuela. Due to its historical perspective, CORDE allocates a total of 74% to Peninsular Spanish and 26% to the remaining 22 varieties. Additionally, there are records of Spanish from all periods, from its beginnings to 1974. CORDE currently contains 250 million words (approximately) from texts pertaining to various subjects (fiction, didactics, science and tech-

⁴ Available online through: <http://www.corpusdelespanol.org/>

⁵ Available online through: <http://corpuscodea.es/>

⁶ Available online through: <http://corpus.bibliamedieval.es/>

⁷ Available online through: <http://www.corpuscharta.es/>

⁸ *Corpus Diacrónico del Español* (CORDE): <http://corpus.rae.es/cordenet.html>

⁹ *Corpus Diacrónico y Diatópico del Español de America* (CORDIAM): <http://www.cordiam.org/>

nology, press, religion, history, law and legal). The CORDE interface allows narrowing searches to specific places, periods of time, and domains.

The Corpus Diacrónico y Diatópico del Español de América (CORDIAM) is divided into three subcorpora: CORDIAM-Documents, which contains chronicle, administrative, legal texts and letters; CORDIAM-Literature, where narrative, poetry, prose (satire, dialogues, first manifestations of modern essay, etc.) and theatre are found; and CORDIAM-Press, which is constituted by documents written exclusively in American Spanish from the 19 Latin American countries, besides Southern and Western United States, Jamaica, Haiti and Guyana (former territories of the Spanish crown). CORDIAM-Press contains only texts from the eighteenth and nineteenth centuries because the press exists in America from the eighteenth century. CORDIAM covers texts from four centuries, the oldest dating from 1494 and the most recent from 1905. It is a relatively small corpus containing 4,475,759 words in 6,058 texts (4,178 documents, 15 literature and 1,865 newspapers).

In this study, the searches in CORDIAM were not restricted to any particular type of text or genre, because the portion of literary texts in CORDIAM is too small. However, they were narrowed to the nineteenth century (chronological criterion).

We have chosen to work with CORDE and CORDIAM for the following reasons: both corpora offer users the possibility of performing query searches by using geographical and chronological filters; both contain a wide variety of genres and a large number of words.

In this study, the searches for verb + noun (object) collocations in CORDE have been restricted geographically for obtaining individual results from Colombia, Spain and also from general Spanish. Due to the fact that *The Portrait of Dorian Gray* was published in 1890, the chronological restriction here has been established between 1850 and 1950, because we consider this to be a reasonable period of time to observe the use of the language of that time, which must be also reflected in the target texts.

4 Rendering of collocations

Collocations were extracted manually from the source text (ST), as well as their corresponding translation equivalents from the two selected target texts, one from Colombia (Wilde 1890/1994) and the other one from Spain (Wilde 1890/1999). The translation techniques used in the rendering of collocations into Spanish were identified and analysed in the selected texts, based on the classification of techniques proposed by Hurtado Albir (2001).

First, all verb + noun (object) combinations of the first two chapters of the ST and their corresponding translation equivalents in the target texts (TT) were extracted manually. Then some were eliminated until a list of 60 collocations was obtained. The criteria to discard some combinations of this type

were typicality and frequency of use, hence the selected collocations express typical relations between their components and present high frequency of use. Among the combinations that were discarded are to *cause excitement*, *preach the importance*, *return to the ideal*, etc. In cases where there was doubt, some dictionaries and resources were consulted, namely, the monolingual dictionary *Oxford Collocations* (2002), the corpus EnTenTen13¹⁰ (Word Sketch through *Sketch Engine*), which indicates collocations or word combinations for any lemma queried, the *Diccionario Combinatorio Práctico del Español Contemporáneo*, the EsTenTen11¹¹ and the diachronic corpora CORDE and CORDIAM. This was done with caution, since these resources do not take into account the diachrony of language, except for CORDE and CORDIAM.

4.1 Translation techniques

It can be clearly observed that in the Colombian Spanish target text (hereafter CO_SP_TT) (see Table 1) a great variety of techniques were applied in the translation of verb + noun (object) collocations, while in the Peninsular Spanish target text (hereafter PEN_SP_TT) fewer techniques were used (see Table 2). Established equivalence was the preferred option in both target texts.

Table 1. Translation techniques identified in the rendering of collocations in the CO_SP_TT

TECHNIQUE	CO_SP_TT	PERCENTAGE
Established equivalent	47	77%
Linguistic compression	6	10 %
Modulation	2	3 %
Transposition	2	3 %
Literal translation	1	2 %
Discursive creation	1	2 %
Linguistic amplification	1	2 %
Reduction	1	2 %

¹⁰ Available through Sketch Engine: <https://www.sketchengine.co.uk/>. (Subscription required).

¹¹ Available through Sketch Engine: <https://www.sketchengine.co.uk/>. (Subscription required).

77% of collocations were translated using the **established equivalence** technique in the CO_SP_TT. In this category, we group those collocations that were translated as other collocations recognised (by usage or by the dictionary) in the target language (TL), and which, therefore, constitute the established equivalents.

In the CO_SP_TT, for instance, the translation of *to catch the gleam* as *ver el brillo*, or *to elevate one's eyebrows* as *arquear las cejas*; and the translation in both TT of *to close one's eyes* as *cerrar los ojos* and *to tell the truth* as *decir la verdad*. The collocation *to exhibit a picture* was translated using two established equivalents: *exponer un retrato* and *exhibir un retrato*.

In the vast majority of cases, the base was translated literally and the collocate did vary. In just one case the base was changed in the collocation, as in the translation of *to bear the burden* as *soportar el peso*.

Table 2. Translation techniques identified in the rendering of collocations in the PEN_SP_TT

TECHNIQUE	PEN_SP_TT	PERCENTAGE
Established equivalent	53	89 %
Linguistic compression	5	8 %
Literal translation	1	2 %
Reduction	1	2 %

Similarly, the **established equivalent** was the most frequently used technique in the translation of collocations in the PEN_SP_TT (see Table 2). For example, the collocations *to make a success* translated as *tener éxito* or *to play the piano* as *tocar el piano*. It would not make any sense to translate literally *to make a success* as *hacer un éxito*, hence it must be translated using the established equivalent in Spanish: *tener éxito*.

Only in one case, the **literal translation** technique was applied in both target texts. Many collocations in the English language are transparent, not only in terms of comprehension but also of production, which permits a word-for-word translation into Spanish, although this does not always imply that the resulting translation is a typical and frequent collocation in the TL, that is, the established equivalent. For instance, the collocation *to think (...) thoughts* translated as *pensar (...) ideas* in the PEN_SP_TT and *pensar con pensamientos* in the CO_SP_TT.

On the other hand, the **linguistic compression** technique was applied in the translation of collocations as single verbs in the TL, instead of as other collocations, resulting in a synthesis of linguistic

elements. Most of these cases occur with collocations formed by delexical, almost grammaticalised verbs, and a generally deverbal noun, which provides the main semantic load (Corpas Pastor, 1996: 69). Some collocations that illustrate this are: *to have no appreciation*, translated as *no apreciar*, in both CO_SP_TT and PEN_SP_TT; *to make no answer* translated as *sin responder* and *no contestar*, in the CO_SP_TT and in the PEN_SP_TT, respectively. Whereas in the PEN_SP_TT, we find: *to make a difference* translated as *distinguir*. Likewise, as an example of the use of this technique, we include collocations that were translated as more general verbs, like *to ring the bell* and *llamar* (in both TT).

Other techniques were used less frequently in the translations analysed. In the CO_SP_TT, for instance, a change of point of view (**modulation**) can be observed in the translation of *to keep one's promises* as *faltar a sus promesas* and *to break one's promises* as *no cumplir sus promesas*. Additionally, a change of grammatical category is noticed in the translation of *to cure the soul* as *remedio para el alma* (**transposition**). In this same TT, a linguistic amplification is employed in the translation of the collocations *to frame [a portrait]* as *poner en marco [el retrato]*; and an ephemeral equivalence (**discursive creation**) is established for the collocation *to convey the sense* translated as *plasmar la velocidad*, which is not a frequent collocation in Spanish. Also, the reduction technique is clearly used in the omission of the collocations *to have a romance* in the PEN_SP_TT and *to give a definition* in the CO_SP_TT.

Not all collocations from the source language were translated as other collocations in the target language. Some of them were translated as free combinations of verb + noun, ephemeral equivalents (**discursive creation**), as verbs that contain all the semantic load, or they were simply omitted. Most of these instances were observed in the CO_SP_TT and others, such as the discursive creation, were pinpointed exclusively in this TT.

The English collocations found in the ST were consulted in some popular bilingual online dictionaries too, namely the *Oxford Dictionary*¹², *Collins Dictionary*¹³, *Cambridge Dictionary*¹⁴, and the *Linguee Dictionary*¹⁵, with the aim of observing the repertoire of translation (English-Spanish) options they offer for collocations. Undoubtedly, the treatment of collocations in bilingual dictionaries is now extremely poor compared to that in monolingual dictionaries (Cf. Corpas Pastor, 2017), and their approach to the translation of collocations is usually very simplified. In general, the *Linguee dictionary* is the one that provides the highest number of possible equivalents for collocations, for example,

¹² *Oxford English-Spanish Dictionary* <<https://es.oxforddictionaries.com/>> (Accessed: 01-04-19).

¹³ *Collins English-Spanish Dictionary* <<https://www.collinsdictionary.com/>> (Accessed: 01-04-19).

¹⁴ *Cambridge English-Spanish Dictionary* <<http://dictionary.cambridge.org/es/diccionario/ingles-espanol/>> (last accessed: 01-04-19).

¹⁵ *Linguee English-Spanish Dictionary* <<http://www.linguee.es/>> (last accessed: 01-04-19).

to produce an effect is rendered into Spanish as *causar un efecto*, *tener un efecto*, *generar un efecto* and *producir un efecto*; the latter is the one that has more instances. In the *Collins Dictionary* this collocation is translated as *producir un efecto* and *surtir efecto*. Conversely, in the other two bilingual dictionaries analysed no translation equivalent is found for this same collocation. Another example is *to yield to (a) temptation*, whose only equivalent in the *Collins* and *Oxford dictionaries* is *ceder a (la) tentación*, whereas in *Linguee* the following are also retrieved: *rendirse a la tentación*, *caer en la tentación*, *sucumbir a la tentación*.

The English collocation *to exhibit the picture/portrait* is included in the *Collins* and *Cambridge dictionaries* with the verb *exponer* as the only translation equivalent, albeit in some Latin American countries the verb *exhibir* is a more common collocate for the nouns *cuadro/pintura*. Regarding the collocation *to pay attention*, its only translation equivalent found in these dictionaries is *prestar atención*, despite the fact that the verb *poner* is also a very frequent collocate used in American Spanish.

It is evident that bilingual dictionaries do not reflect the diatopic variation of Spanish, but rather favour general Spanish as regards the translation of verb + noun (object) collocations. These results support those obtained by Corpas Pastor (2015b, 2017). In the following section, the results of the search for collocations in CORDE and CORDIAM are presented.

4.2. Searching collocations in CORDE and CORDIAM

After identifying and analysing the techniques used in the translation of collocations in chapter 1 and 2 of *The Portrait of Dorian Gray*, it was necessary to discard those that are not considered collocations in the target language, in order to carry out the analysis in the Spanish corpora (CORDE and CORDIAM), whose results are presented in this section.

A total of 50 collocations were extracted from the PEN_SP_TT, out of which 23 (46%) were also found in the CO_SP_TT, which means that both target texts share a high percentage of collocations. With respect to the results yielded by the large corpora, 44 (88%) collocations out of those 50 also occurred in general Spanish (GEN_SP) from CORDE. 47 (94%) out of these collocations appeared in Peninsular Spanish (PEN_SP) from CORDE and 3 (6%) were not found in CORDE or CORDIAM. Whereas in CORDIAM, only 9 (18%) collocations out of the total 50 were found in the GEN_SP and only 3 (6%) collocations in the Colombian Spanish (CO_SP) from CORDIAM: *producir un efecto*, *decir la verdad* and *prestar atención*.

On the other hand, a total of 48 collocations were found in the CO_SP_TT, out of which 23 (47.91%) were also found in the PEN_SP_TT. As for large corpora, 46 (95.83%) collocations out of those 48 appeared in the GEN_SP from CORDE, and 29 (95.83%) out of these also occurred in the CO_SP from CORDE, 2 (4.16%) were not found in CORDE or CORDIAM, and all of these collocations occurred in the PEN_SP (100%) from CORDE. As we can observe, a high percentage of the collocations which occur in the Colombian Spanish target text are used in Peninsular Spanish and in many other diatopic varieties.

Regarding CORDIAM, only 11 (22.91%) collocations out of 48 were found in the GEN_SP. Only 2 (4.16%) collocations occurred in the CO_SP. This is due to the small size of CORDIAM and the fact that the queries were restricted to the 19th century, which further reduced the chances of finding occurrences in this corpus.

The results retrieved from CORDIAM were very poor, mainly those concerning Colombian Spanish. Furthermore, the larger proportion of Peninsular Spanish words contained in CORDE is justified by the fact that this is a historical corpus of Spanish from all the countries where Spanish has been spoken, thus the documents from Spain outnumber those from elsewhere.

The data reveals that the translation of collocations in the CO_SP_TT is closer to general Spanish than to Colombian Spanish, inasmuch as this TT and the PEN_SP_TT share a high percentage of collocations. Those collocations that were found only in the CO_SP_TT occurred in the corpora in other varieties too, especially in Peninsular Spanish, which indicates that none of these collocations are diatopically restricted to Colombian Spanish. For example, the collocations *arqurear las cejas*, *exhibir el retrato*, *sentir placer*, etc., despite appearing in the CO_SP_TT, were not found in the Colombian Spanish from CORDE or CORDIAM, albeit they were retrieved from other national varieties of CORDE.

However, no occurrences of the collocations *crear (un) efecto* and *ganar (un) empuje*, extracted from the CO_SP_TT, were observed in any of the corpora, suggesting a very low frequency of use of such collocations in Spanish between 1850-1950. In the case of the collocation *crear (un) efecto*, as in the vast majority, it was necessary to check each of the concordances and discard some of them for not complying with the syntactic function we expected to find in the corpora.

It is also noteworthy that there were no instances in CORDE or CORDIAM of four collocations extracted from the PEN_SP_TT: *captar (el) efecto*, *despilfarrar (el) oro*, *captar (el) significado y enmarcar (un) retrato*. This may be due to several factors: first, these collocations are neither very frequent nor very typical in the language; second, generally collocations allow the variation of at least one of their components without affecting the meaning of the others, for instance *despilfarrar el oro*, where the noun *oro* could be replaced by *dinero/fortuna/recursos*, etc. *Despilfarrar dinero* does have

occurrences in CORDE, although not in Colombian Spanish. As for the absence of the collocations *captar (el) significado* and *enmarcar (un) retrato* from both corpora, it could be attributed to a rare use of them in the period of time consulted.

The established equivalent was the most used technique in the translation of collocations in both target texts, which results in more conventional and standard translations, and less creative and experimental target language. Although a greater number of techniques were applied in the Colombian Spanish target text, besides the established equivalent, this TT reflects a general Spanish. In other words, it does not exhibit specific features of the Colombian Spanish variety, as regards collocations, but, on the contrary, it is more similar to general Spanish.

On the other hand, the bilingual dictionaries consulted are characterised by maintaining this tendency to standardise translated language, listing only one or two translation equivalents for certain collocations. These dictionaries are indeed far from reflecting the reality of the diatopic variety of a transnational language like Spanish. In most cases, the collocations that we looked up in these dictionaries were not even listed. This is evidence of standardisation in translated language and, at the same time, can be a contributing factor for maintaining standardisation in translated language since dictionaries are one of the translators' essential tools. Other important tools are translation memories and machine translation. According to Pym (2008: 323), the use of translation memories reinforces standardisation because they offer consistency in terminology and phraseology. By the same token, we consider that corpus-based machine translation tools (statistical machine translation and Neural machine translation) also reinforce standardisation because, broadly speaking, they rely on huge parallel corpora, which results in more usual and standard translated language and less creative and unusual one.

Additionally, Pym (2008: 325) argues that standardisation is a strategy used by translators to reduce communicative risk. Translators tend to select more frequent and typical collocations that are recognised and accepted by most of the Spanish-speaking community as the prestigious variety, rather than select a diatopically marked variety that might sound odd and incorrect to the majority. Furthermore, standardisation practices may also take place during the editing process carried out by publishing houses, as Peter Bush (1998/2001: 129) puts it: "In the English, Spanish and Portuguese-speaking worlds, for example, there will be issues of different dialects and editors who will only accept their variety of standard [...]. The editor's reading, however, need not simply be a threatening and standardizing project".

Today, research by means of diachronic and diatopic Spanish corpora face numerous limitations. In this study, we had to deal with the fact that CORDE and CORDIAM are neither lemmatised nor grammatically annotated, which implies a slower data extraction process. Therefore, it was necessary

to check each of the results yielded by the corpora to verify that these actually reflected the linguistic phenomenon we intended to analyse. To illustrate this, let us give the example of the query *mezcl* dist/5 color**, which yielded instances containing adjectives like *mezclado*, nouns such as *mezcla*, and verb + noun (subject) collocations, but these were ruled out as being beyond the scope of this study.

The total size of the subcorpus of Colombia, or that of any particular national variety, is not specified in either of the corpora. For this reason, it was not possible to establish the normalised frequency of each collocation. This, in turn, prevented us from comparing objectively the frequency of use of the collocations in different countries (subcorpora), in order to determine whether each collocation was more typical of and more frequent in one national variety or another – which is what was intended, generally speaking, in this research. Another disadvantage of CORDE is that in order to be able to check the concordances resulting from a given query, the maximum number of these cannot exceed 2,000. Consequently, it is necessary to formulate query searches as specific as possible so that all results are shown.

As far as CORDIAM is concerned, its size is much smaller than that of CORDE. Therefore, when query searches in CORDIAM are narrowed using geographical and chronological criteria, the results are very poor, or even zero. Furthermore, this corpus is composed mainly of legal, administrative, private documents and chronicles. Only a small percentage of the total corpus is made up of literary texts, which largely conditions the results of queries.

5 Conclusions

Generally, translated language tend to exhibit more features of standardisation than non-translated language. In the case of translated Spanish, we observe a tendency to avoid the use of diatopically marked phraseological units and, on the contrary, favour those that correspond to general or standard Spanish. In this paper, we argue that the rendering of verb + noun (object) collocations into Colombian Spanish of the first two chapters of *The Portrait of Dorian Gray* by Oscar Wilde is indeed closer to general Spanish than to the Colombian Spanish variety. The data reveal that the most frequently applied technique in the rendering of collocations in both target texts was the established equivalent. The Colombian Spanish target text shares 23 out of the 48 collocations (47.91%) with the peninsular Spanish target text. Secondly, the results of the analysis of the diatopic distribution of the collocations in CORDE show that 46 out of the 48 collocations (95.83%) found in the Colombian Spanish target text are also used in general Spanish and 100% of these in Peninsular Spanish. Moreover, no diatopically restricted collocations to Colombian Spanish variety were found.

The evidence presented in this study also reveals that bilingual dictionaries (English-Spanish) show a general and simplified approach to translation since they usually list only one or two translation equivalents, or even none, for certain collocations. This may indicate that the reality of diatopic varieties of Spanish, as a transnational language, is not taken into consideration. In conclusion, these findings support the starting hypothesis of this study, and they contribute to the research on standardisation of national varieties of Spanish in translated texts, insofar as they provide evidence on translated literature from English into Spanish. This could be explained by many factors, including, but not limited to, commercial interests of the translation industry, the use of translation memories, machine translation tools, and bilingual dictionaries (English-Spanish), risk-aversion translators' performance, editing process in both countries, among others. In any case, it will be necessary to delve deeper into other publishing-related, socio-cultural, and even cognitive variables involved in the tendency towards standardisation, to be able not only to describe translation behaviour, but also to explain it.

However, a note of caution is order at this point. The limitations of this study could have affected the distributional results obtained from CORDE and CORDIAM as to frequency and diatopy. Moreover, these conclusions are also limited by the fact that this study only deals with two translated texts. Further studies of this nature for Spanish and other types of phraseological units are badly needed in order to reveal 'true' regularities in translated language. A first step will be to revise and expand the existing diachronic and diatopic Spanish corpora, by size increase and annotation enrichment (particularly at the morphosyntactic level).

Acknowledgements

M.^a Victoria Valencia Giraldo is supported by Banco Santander–Universidad de Salamanca International PhD Scholarship.

References

1. Academia Mexicana de la Lengua, Corpus Diacrónico y Diatópico del Español de América (CORDIAM), last accessed 2019/04/20

2. Baker, M.: Corpus linguistics and translation studies: Implications and applications. Text and technology: In honor of John Sinclair. In: Baker, M., Francis, G., Tognini-Bonelli, E. (eds.), pp. 233–250. John Benjamins, Amsterdam/Philadelphia (1993).
3. Baker, M.: Corpus-based translation studies: the challenges that lie ahead. In: Somers, H. (Ed.), *Terminology, LSP and translation*, pp. 175–186. John Benjamins, Amsterdam/Philadelphia (1996).
4. Baker, M.: Investigating the language of translation: A Corpus-based approach. *Goal* 43 (4), pp. 480–485. (1998).
5. Baker, M.: *In other words: a coursebook on translation*. Routledge, London/New York (2011).
6. Biber, D., Conrad, S.: *Register, genre, and style*. (Cambridge textbooks in linguistics). Cambridge University Press, Cambridge (2009).
7. Bosque, I. (dir.): *Redes. Diccionario combinatorio del español contemporáneo*. Ediciones SM, Madrid (2004).
8. Bush, P.: Literary translation practices. In: Baker, M. (ed.), *Routledge encyclopedia of translation studies* (1st edition), pp. 127-130. Routledge, London, (1998/2001).
9. Chesterman, A.: Hypotheses about translation universals. In: Hansen, G., Malmkjaer, K., Gile, D. (eds), *Claims, changes and challenges in translation studies*, pp. 1-13. John Benjamins, Amsterdam, (2004).
10. Corpas Pastor, G.: Laughing one's head off in Spanish subtitles: a corpus-based study on diatopic variation and its consequences for translation. In: Mogorrón, P., Martines, V. (Eds.), *phraseology, diatopy and translation*, Collection IVITRA research in linguistics and literature, pp. 54-106. John Benjamins, Amsterdam (2018).
11. Corpas Pastor, G.: Collocations in e-bilingual dictionaries: from underlying theoretical assumptions to practical lexicography and translation issues. In Torner, S., Bernal, E. (eds.), *Collocations and other lexical combinations in Spanish. Theoretical and applied approaches*, pp. 139-160. Routledge, London (2017).
12. Corpas Pastor, G.: Collocations dictionaries for English and Spanish: the state of the art. In Orlandi, A., Giacomini, L. (eds.), *Defining collocations for lexicographic purposes: from linguistic theory to lexicographic practice*, Series 'Linguistic Insights', pp. 173-208. Peter Lang, Frankfurt (2016).
13. Corpas Pastor, G.: Register-specific collocational constructions in English and Spanish: A usage-based approach. *Journal of Social Sciences*, Eleven (3), pp. 139–151 (2015a).

14. Corpas Pastor, G.: Translating English verbal collocations into Spanish: On distribution and other relevant differences related to diatopic variation. *Linguisticæ Investigationes*, 38(2), pp. 229–262 (2015b).
15. Corpas Pastor, G.: Investigar con corpus en traducción: los retos de un nuevo paradigma. Peter Lang, Frankfurt (2008).
16. Corpas Pastor, G.: En torno al concepto de colocación. *Euskera*, 1, pp. 89–108 (2001).
17. Corpas Pastor, G.: Manual de fraseología española. Gredos, Madrid (1996).
18. Davies, M.: Creating useful historical corpora: A comparison of CORDE, the Corpus of Spanish, and the Corpus do Português. *Diachrony of Ibero-Romance languages: New contributions from corpus linguistics*, pp. 137-166 (2009).
19. Gledhill, C.: Collocations in science writing. Gunter Narr, Tübingen (2000).
20. Haensch, G., Wolf, L., Ettinger, S., Werner, R.: La lexicografía. De la lingüística teórica a la lexicografía práctica. Gredos, Madrid (1982).
21. Hardy, D. E.: The body in Flannery O'Connor's Fiction: Computational Technique and Linguistic Voice. University of South Carolina Press, Columbia (2007).
22. Hausmann, F. J.: Le dictionnaire de collocations. In Hausmann, F. J. Reichmann O., Wiegand H.E., Zgusta L. (eds.), *Wörterbücher, Dictionaries, Dictionnaires*, 1, pp. 1010-1019. De Gruyter, Berlin/New York (1989).
23. Hilper, M.: Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory*, 2(2), pp. 243-257 (2006).
24. Hoover, D.: Frequent collocations and authorial style. *Literary and Linguistic Computing*, 18 (3), pp. 261-286 (2003).
25. Howarth, P. A.: Phraseology in English academic writing. Some implications for language learning and dictionary making. Max Niemeyer Verlag, Tübingen (1996).
26. Hurtado Albir, A.: Traducción y traductología: Introducción a la traductología. Cátedra, Madrid (2001).
27. Jones, S., Sinclair, J. M.: English lexical collocations. A study in computational linguistics. *Cahiers de lexicology*, 24, pp. 15-61 (1974).
28. Koike, K.: Colocaciones léxicas en el español actual: estudio formal y léxico-semántico. Universidad de Alcalá / Takushoku University, Alcalá de Henares (2001).
29. Manning, C. D., Schütze, H.: Foundations of statistical natural language processing. The MIT Press, Cambridge/Massachusetts (1999).
30. Oxford collocations dictionary for students of English. Oxford University Press, Oxford (2002).

31. Paffey, D.: Language ideologies and the globalisation of 'standard' Spanish. Bloomsbury, London/New York (2012).
32. Pym, A.: Exploring translation theories. Routledge, London/New York (2010).
33. Pym, A.: On Toury's laws of how translators translate. In: Pym, A., Shlesinger, M., Simeoni, D. (eds), *Beyond descriptive translation Studies*, pp. 311–328. John Benjamins, Amsterdam/Philadelphia (2008).
34. Real Academia Española: Data Bank (CORDE) [online]. Corpus Diacrónico del Español. <http://www.rae.es>, last accessed 2019/04/20
35. Sinclair, J.: Beginning the study of lexis. In: Bazell, C., Catford, J., Halliday, M., Robins, R. (eds.), *In memory of J.R. Firth*, pp. 410-430. Longman, London (1966).
36. Toury, G.: *Descriptive translation studies and beyond*. 2nd edn. John Benjamins, Amsterdam/Philadelphia (1995/2012).
37. Wilde, O.: *The picture of Dorian Gray*, Penguin Classics, London (1985) (original work published 1890).
38. Wilde, O.: *El retrato de Dorian Gray*, Panamericana, Bogotá (1994).
39. Wilde, O.: *The portrait of Dorian Gray*, translated by López Muñoz, J., El Mundo, Madrid (1999).
40. Williams, G.: In search of representativity in specialised corpora – categorisation through collocation. *International Journal of Corpus Linguistics*, 7 (1), pp. 43-64 (2002).