Phraseology and Style in Margaret Juli Costa's English Translation of Marías' Corazón Tan Blanco (1992)¹

MIGUEL FUSTER-MÁRQUEZ

Universitat de València miguel.fuster@uv.es https://orcid.org/0000-0003-3813-1770

PHRASEOLOGY AND STYLE IN MARGARET JULL COSTA'S ENGLISH TRANSLATION OF MARÍAS' CORAZÓN TAN BLAN-CO (1992)

ABSTRACT: Many studies have paid attention to Javier Marías' novel Corazón tan blanco (1992), within narratology and literary criticism approaches. An outstanding feature of it is phraseological repetition. Although repetition has been treated by specialists in Marías' work, no studies have quantified this feature systematically in translation studies to date. This contribution is based on stylistics and literary translation studies which make use of a corpus methodology to focus on the analysis of linguistic features and follows a text-oriented perspective. It analyses quantitatively and qualitatively the complexity of phraseological repetition and, particularly, sequences of five or more words as foregrounding features which host important narrative functions. The results show that Jull Costa's English translation reproduces on a roughly equal percentage this phraseology and divergent renderings. The qualitative analysis of the English translation also attempts to provide reasons for some of these departures by the translator.

FRASEOLOGÍA Y ESTILO EN LA TRADUCCIÓN INGLESA DE CORAZÓN TAN BLANCO (1992) DE JAVIER MARÍAS POR MAR-GARET JULL COSTA

RESUMEN: Muchos estudios han prestado atención a la novela de Javier Marías Corazón tan blanco (1992) dentro de los enfoques de narratología y critica literaria. Una característica destacable es la repetición fraseológica. Aunque ésta ha sido tratada por especialistas en la obra de Marías, no hay estudios que lo hayan cuantificado sistemáticamente en estudios traductológicos. Esta contribución se basa en la estilística y los estudios de traducción literaria que hacen uso de métodos de corpus en el análisis de las características lingüísticas desde una perspectiva orientada al texto. Analiza particularmente la complejidad de las secuencias de cinco o más palabras, características de 'foregrounding', con importantes funciones narrativas. Los resultados muestran que la traducción al inglés de Juli Costa reproduce esta fraseología y presenta divergencias en igual medida. El análisis cualitativo de la traducción al inglés también intenta ahondar en las razones de algunas de estas desviaciones por parte del traductor.

PHRASÉOLOGIE ET STYLE DANS LA TRADUCTION AN-GLAISE DE CORAZÓN TAN BLAN-CO (1992) DE JAVIER MARÍAS PAR MARGARET JULL COSTA

RÉSUMÉ: De nombreuses études ont porté sur le roman Corazón tan blanco (1992) de Javier Marias, dans le cadre d'approches de narratologie et de critique littéraire. Une caractéristique remarquable de celui-ci est la répétition phraséologique. Bien que la répétition a été traitée par des spécialistes dans les travaux de Marías, à ce jour, aucune étude n'a quantifié systématiquement cette caractéristique dans les études de traduction. Cette contribution est basée sur la stylistique et les approches des études de traduction littéraire qui utilisent une méthodologie de corpus pour se concentrer sur l'analyse des caractéristiques linguistiques et suivent une perspective axée sur le texte. On analyse quantitativement et qualitativement la complexité de la répétition phraséologique et, en particulier, des séquences de cinq mots ou plus en tant que caractéristiques de 'foregrounding' qui hébergent des fonctions narratives importantes. Les résultats montrent que la traduction anglaise de Jull Costa reproduit sur un pourcentage à peu près égal cette phraséologie et des rendus divergents. L'analyse qualitative de la traduction anglaise tente également de fournir les raisons de certains de ces départs par le traducteur.

 Fecha de recepción:
 14/12/2021

 Fecha de revisión:
 25/02/2022

 Fecha de aceptación:
 18/05/2022

 Fecha de publicación:
 01/12/2022

¹ This research paper is part of the project News Values and Ideology: The Discursive, Cross-cultural Construction of Gender and Social Inequalities in (Digital) Press through Online, Real Time Corpus Linguistic Tools. The Gender Gap Tracker and Kaleidographic (Ref. PID-2019-110863GB-100) financed by MCIN.

KEYWORDS: literary translation; corpus stylistics; clusters; variation; Javier Marías; *Corazón tan blanco*; *A Heart So White.*

SUMMARY: 1. Introduction. 2. Javier Marias, Corazón tan blanco and the English Translator. 3. Stylistics and literary translation studies. 4. Repetition in Corazón tan blanco. 5. Corpus methodologies in stylistics and literary translation studies. 6. Data processing and methodology. 7. Analysis of the ST. 8. Treatment of the ST clusters in the TT. 9. Conclusions. 10. References.

PALABRAS CLAVE: traducción literaria; estilística de corpus; agrupamientos; variación; Javier Marias; Corazón tan blanco; A Heart So White.

SUMARIO: 1. Introducción. 2. Javier Marías, Corazón tan blanco y el traductor de inglés. 3. Estudios de estilística y traducción literaria. 4. Repetición en Corazón tan blanco. 5. Metodologías de corpus en estudios de estilística y traducción literaria. 6. Tratamiento y metodología de los datos. 7. Análisis del TO. 8. Tratamiento de los agrupamientos en TO y TM. 9. Conclusiones. 10. Referencias.

MOTS-CLÉS: traduction littéraire; stylistique de corpus; clusters; variation; Javier Marías; *Corazón tan blanco*; *A Heart So White*.

SOMMAIRE: 1. Introduction. 2. Javier Marías, Corazón tan blanco et le traducteur anglais. 3. Études de stylistique et de traduction littéraire. 4. Répétition dans Corazón tan blanco. 5. Méthodologies de corpus dans les études de stylistique et de traduction littéraire. 6. Traitement des données et méthodologie. 7. Analyse de la ST. 8. Traitement des clusters dans TS et TC. 9. Conclusions. 10. Références.

1. Introduction

Published in 1992, *Corazón tan blanco* (henceforth *CTB*) is a well-known novel by the unfortunately recently deceased Spanish writer Javier Marías. Research in narratology and literary criticism has focused on various themes and stylistic features in this novel. Arguably, one of its most conspicuous traits is the use of various forms of repetition. The important literary functions performed by repetition have been treated by experts on Marías' narrative production (see Grohmann, 2002; Basso, 2002; Florenchie, 2003; Steen, 2011). More recently, Fuster-Márquez (2020) explored repeated long phrases in *CTB* by adopting a corpus methodology.

Repetition is also a central aspect in stylistics and in literary translation studies. Undeniably, repeated phraseology may pose a formidable challenge to translators since they need to be aware of the effects it achieves in *CTB*. The aim of this paper is to explore how Margaret Jull Costa, a renowned British translator of Contemporary Spanish and Portuguese fiction, approached phraseological repetition in her translation *A Heart So White* (1995) (henceforth *AHSW*).

Two research questions we hope to answer here are:

- RQ1 To what extent has Jull Costa been able to reproduce this repeated long phraseology in her translated English text?
- RQ2 Which obstacles or decisions could prevent the translator from faithfully reproducing this kind of phraseology?

The first question requires a quantitative analysis, whereas the second requires a more qualitative approach. Since repeated long clusters are a crucial literary feature in the source text (henceforth ST), our hypothesis is that the English translator will attempt to reproduce it in her target text (henceforth TT).

The approach and methodology adopted here are inspired by corpus stylistics and literary translation studies, where the linguistic treatment is prioritized in the analysis of literary works and their translations using methods developed by corpus linguists. These can be found in contributions by Stubbs (2005), Mahlberg (see, for example, 2007a, 2007b, 2013a, 2013b, 2014), or Mahlberg and McIntyre (2011); but most particularly in the work of literary translation researchers who have applied corpus methods (see Čermáková, 2015, 2018; Johnson, 2016; Winters, 2018).

This paper attempts to analyse the challenging complexity of repeated long clusters both in the ST, as they serve important textual functions in Marías' novel, and in the English translation produced by Jull Costa. In addition, the internal complexity of phraseological variation, without which it would be hard to grasp the relevance of phraseology in CTB, is analysed. This research is concerned with the style of a single work, and text-oriented since it attempts to measure the impact of a stylistic feature present in the ST on its TT. A different stylistic approach has been followed by Saldanha (2011a, 2011b), whose interest has been to discuss per se the style and Jull Costa as an English translator. Nevertheless, the distinction between text-oriented and target-oriented approaches in the analysis of literary translation is not clear-cut (Saldanha, 2014; Winters, 2018). As argued by Saldanha (2021), scholars adopting a text-oriented perspective also need to be aware of translational choices related to target audiences. In fact, as will be shown, some strategies adopted by Jull Costa bear witness to the need of having in mind the target readership. Nevertheless, one main assumption is that repeated long clusters are rather unusual linguistic manifestations in ordinary usage. Thus, they deserve attention in the exploration of literary works and their translations as cases of 'foregrounding' (cf. Miall and Kuiken, 1994; Čermáková, 2015).

2. Javier Marías, Corazón Tan Blanco and the English Translator

Javier Marías has been a well-known and prolific writer in contemporary Spanish literature. In his career, Marías has combined his profession as a writer with that of a translator of English literature, an autobiographical fact that somehow sneaks into his novels, including *CTB*, where Juan, the narrator, and his wife, Luisa, are characterised as interpreters in international organisms. In *CTB*, the intrigue of a 'tragic' event which predates the birth of Juan (the narrator and Ranz's son), the suicide of his aunt Teresa in front of the mirror mentioned in the first chapter, will lead us to the story of a previous crime, the murder of Ranz's first Cuban wife, kept secret until the penultimate chapter. *CTB* has been classified as suspense since, as Vila Sánchez (2015: 299) claims, "the clarification of a death is the trigger for the

narrative."2 Grohmann (2002: 190) views CTB as an 'anti-detective' novel since the secret of a crime which only the culprit knows, and seemingly the narrator (his son), does not wish to know about. Navarro Gil (2004) points out that Marías leads readers to the disturbing events of the past with a narrative which is sinuous and fragmentary. Readers reach the solution of the intrigue by zigzagging among minor stories that are intertwined with consequences leading to the main plot, as the narrator gradually incorporates memories and reflections throughout the novel. This is how readers can interpret the embedded stories of Miriam and Guillermo, the story of Berta and Bill, or even Juan's allusions to fragments of Shakespeare's Macbeth (indeed, the title of the Spanish novel, *Corazón tan blanco*, is inspired by Shakespeare's words "a heart so white" in *Macbeth*), while working as an interpreter with his wife.

By means of repetition, the narrator gradually contributes memories and reflections, thus underpinning "presentiments of disaster", to which he refers five times in *CTB*. Navarro Gil (2004) states that *CTB* is anchored on a speculative narrator who seeks coherence as he reconstructs past events in his family. Navarro Gil also highlights that digression, uncertainty and dilemma are intertwined in *CTB*, taking shape in the repetition of words and phrases, in a discursive context in which reflection and nuance are key aspects (cf. also Pérez-Carbonell, 2016).

As a translator, Jull Costa has translated into English the works of contemporary Portuguese and Spanish writers, amongst these the works of Marías or Saramago. In an interview for the *White Review* (2011),³ Jull Costa states that her translation strategy includes "fact-checking, research, queries to the author [...] and queries [...] to my favourite and faithful native speaker consultants." She confessed that María was helpful and understood what she would need from him. *A Heart so White* earned Jull Costa the International Dublin Literary Award (1997) which, according to her is "the only one of its kind that gives equal weight to novels originally written in English and novels translated from another language into English." (*White Review*, 2011)

3. Stylistics and Literary Translation Studies

For Jeffries and McIntyre (2010: 9), stylistics can be applied to literary and non-literary texts, although for practical and historical reasons it has been associated mostly with literary research. Borrowed from the works of Mukařovský, and coined in English as *foregrounding* (Garvin, 1964), it has become a central concept in stylistics. Similar terms are 'deautomatization', 'actualisation' or 'de-familiarization' (see

² Translations of Spanish quotations into English here are mine.

³ https://www.thewhitereview.org/feature/interview-with-margaret-jull-costa [last accessed 09/09/2021]

Miall and Kuiken, 1994; Gregoriou, 2014). Leech and Short (2007: 39) define foregrounding as 'artistically motivated deviation'. Put simply, foregrounding elements may occur in two ways in the literary work (Simpson 2004: 5): through elements that deviate from the norm (deviation), and through repetition. These foregrounded elements lead readers to search for the literary sense. Repetition can fulfil important textual functions in literary texts. For example, Halliday and Hasan (1976: 279) refer to lexical repetition as a factor of textual cohesion. Also, Aitchison (1994: 20) points out that the repetition of lexical units contributes to textual cohesion in discourse, and Szegedv-Maszák (1983: 43) finds repetition to be an essential element that confers unity to literary and musical works.

Čermáková (2015) claims that translators need to be aware of how mechanisms of deviation operate in the ST, bearing in mind that "[i]n a literary work, content and style work together and are inseparable." (Čermáková, 2015: 356). The present study will consider phraseological repetition of long sequences as a stylistic device employed by Marías, its identification in the ST, and its treatment in the English TT.

As mentioned, we have adopted a source-text orientation that relies on close comparisons between ST and TT see (Čermáková, 2015, 2018; Johnson, 2016; Winters, 2018) in order to see whether Jull Costa replicates the style in CTB. The role of repetition in the ST is too important to be dismissed since "[it] may provide the key to the reading, understanding or even decoding of the literary text," (Ben-Ari 1998: 69). Although there is a firmly established trend in translation studies to disregard equivalence between ST and TT, scholars within a text-oriented corpus stylistic tradition (see Abdulla, 2001; Boase-Beier, 2006; Winter, 2018) have brought it to the fore. Obviously, the identification of equivalence can only be achieved by means of a prior analysis of repetition in the ST.

Repetition in Corazón tan Blanco

Javier Marías' fiction is a place where readers will come across numerous instances of repetition. Florenchie (2003: 2) claims that repetition is in fact an essential aesthetic element of Marías' style. She cites a collaborative article published in El Semanal, "Máquinas del tiempo", where Marías comments on repetition in his works:

It so happens that in my latest novels there is what I call a system of echoes and resonances, and that phrases that at first seemed anecdotal and without special significance reappear many pages later, sometimes identical but in another context, sometimes with variations, 'motives' in the musical sense of the term, and thus they take on a deeper meaning, or else they turn out to be not so accidental and to be an essential part of the story. (Marías, 1997: 72)

Bouguen (2000: 112) notes that in some novels Marías reproduces phrases or even entire pages. For Grohmann (2002), repetition is particularly remarkable in CTB. In his study, Fuster-Márquez (2020) gives examples of different types of repetition in CTB, e.g., anaphora, alliteration, homoioptoton, polysyndeton, or mere repetition, as well as various overlapping rhetorical figures which are based on repetition. For Basso (2002:14), repetition is a constructive resource and theme which manifests itself with a multiplicity of forms in CTB: "Situations, objects, phrases, words, pass from one episode to another, proliferate, vary, move and recombine incessantly, in a truly musical treatment of motives." Narrative scenes or situations are iterated with different characters, making readers aware of parallels between them. For example, the scene on the balcony during the wedding trip of Luisa and Juan to Havana is found later when Juan is waiting under Berta's balcony in New York, or in the scene of the young Custardoy in front of the narrator's house.

Grohmann (2002) identifies six sets of effects that are accomplished by repetition in *CTB*, namely:

- 1. Echoing and foreshadowing;
- 2. Linking different parts and determining its circular structure;
- 3. Producing rhythmic associations of bound textual elements;
- 4. The temporal movement and its consequences;
- 5. Drawing readers' attention to the narrative form;
- 6. Configurations of the uncanny.

Although repetition manifests itself in various ways in the narrative, the most crucial aspect is that stylistic effects are achieved in a very important way, formally, by its repeated phraseology. The rich and complex narrative function of repetition in the ST cannot be overlooked by the translator if the relevant stylistic effects it achieves are to be preserved in the TT. Following an author-oriented approach, I will analyse the foregrounding effect of repeated long clusters in the ST and, subsequently, the stylistic differences which may emerge between the ST and the TT. Due to space restrictions, this study will be limited to the exploration of key long clusters. Some of the crucial narrative functions will be highlighted.

5. Corpus Methodologies in Stylisticsand Literary Translation Studies

Mahlberg (2007c: 115) claims that corpus analysis can provide information which cannot be reached by researchers' intuition. Semino and Short (2004) show that the analysis of literature with corpus methodologies has numerous applications. For example, in the exploration of a single work, or a relevant aspect of it, that is the case of Stubbs' study (2005) of Conrad's *The Heart of Darkness*, or Grego-

ri-Signes' study (2020) of characterization by means of naming patterns in *CTB*. The literary function of key clusters in the works of Charles Dickens has been thoroughly examined by Mahlberg in different contributions (see, for example, Mahlberg 2007a, 2007b, 2007c, 2013a, 2013b; Mahlberg, Conklin and Bissón, 2014). There are no limits to the applicability of corpus methods to the analysis of literature.

According to Saldanha (2021), current stylistic studies of literary translation frequently make use of corpus methods. For example, Johnson (2016) combines keyword and two-to-five-word key clusters as part of her methodology in translation analysis; Mastropierro and Mahlberg (2017) focus on keywords in translation. For Čermáková (2015: 356), corpus methods are useful for the identification of 'universals' in translation in the TT of different languages, which appear to operate "independently from source languages" (see Baker, 1993: 243; Hu, Xiao and Hardie, 2019). Čermáková (2015: 358) suggests that corpus techniques may "support translator's stylistic analysis and facilitate an identification of possibly more "difficult" places in the text."

Applying a corpus methodology entails the application of techniques developed by corpus linguists. One main advantage is to effectively deal with quantification, thereby helping to determine more accurately whether certain stylistic features are characteristic of literary works. Its usefulness becomes evident when the researcher wishes to tackle the analysis of a large number of words with the purpose of finding patterns which are distributed throughout one or more works and, consequently, are difficult to see with the naked eye.

Different corpus methods can be incorporated into stylistic analysis. McIntyre and Walker (2019: 11) highlight that the methodologies employed in stylistic analyses are eclectic. However, there is no agreement as to which specific techniques should be required. What techniques should be applied must be related to the researcher's aim. However, any corpus linguistic approach teaches us that qualitative and quantitative analyses are intimately associated and mutually support each other. This is where the strength of corpus linguistics lies. As suggested by Leech and Short (2007: 50), "just as the study of style cannot entirely rely on quantitative data, neither can it ultimately do without them." Given that our approach is source-text-oriented, in order to profile the style of the ST and the stylistic differences which may emerge between the ST and the TT (Winters 2018: 148), the implementation of corpus methods has been particularly helpful in the identification and quantification of long clusters in source and target texts.

6. Data Processing and Methodology

Fuster-Márquez's (2020) approach was initially adopted in this study by establishing four criteria which would enable us to quantify systematically phraseological features in the ST: (1) size of ngrams; (2) variation (flexibility/inflexibility of n-grams); (3) frequency, and (4) distribution/accumulation.

CTB and the translated text AHSW (ed. Penguin 2012)⁴ were turned into plain text. The ST has 97,918 tokens and 11,026 types, whereas the TT has 112,387 tokens and 7,321 types, reflecting syntactic and vocabulary differences between Spanish and English. Chapters were divided into separate files but assembled in two corpora as single ST and TT. Separating chapters was relevant for our analysis of the relationship of repeated sequences (accumulation and distribution) throughout the story by means of the concordance plot tool.

Most of the quantitative and qualitative analysis has been carried out by means of Antconc 3.5.9., a piece of corpus software created by Laurence Anthony. The tools used here are (1) the concordance, (2) the plot or concordance graph, (3) the ngram tool and (4) File View. Thanks to the quantification analysis provided by corpus techniques, we have been able to know quite precisely the frequency with which these repetitions occur in the ST. Also, thanks to the concordance plot, we may know not only which sequences are repeated but also where in the ST there is greater or lesser frequency of them. The n-gram technique is the method by means of which the extraction of clusters was performed. The n-gram tool performs an automatic search for sequences of a set length in a corpus with little assistance from the researcher, except for the need to fix thresholds for size, frequency and number of texts, or chapters in the case of a single text, where the ngram is present. Regarding size, we generated ngrams with a minimum of five words; the cut-off frequency was 2, and the threshold for range (number of files) was 1, to obtain sequences which were located either in one or more chapters.

After applying this fundamentally corpus-driven technique, we carried out a manual filtering since the n-gram tool produced too many overlapping sequences. The existence of very long sequences, whose precise length could not be immediately established due to the screen limitations of the concordance tool, required the aid of Ant-Conc's File View tool. The File View tool was also crucial to identify variation in clusters. The results, after filtering out overlapping clusters, and including cases of some internal variation, yielded no less than 202 repeated long clusters.

 $^{^4}$ Although the English translation of Javier Marías novel was originally published in 1995, this is the edition we have used.

This initial exploration of ngrams was complemented by means of a Key ngram analysis. To determine the peculiarities of a text or style, corpus linguists make comparisons between the study corpus they want to investigate and a reference corpus containing, often, a larger number of texts which is held to be representative of a language, a variety, a particular genre, or the total production of an author, so that it can be proved which clusters would show a significantly unusual frequency (Scott 1997: 236). To obtain a representative corpus to be compared with our study corpus, we accessed the interface of Versión Anotada del Corpus de Referencia del Español Actual (CREA)⁵, a corpus of 141,419,032 words, (111,265 documents), hosted on the institutional site of the Royal Spanish Academy (RAE) and reflects the use of Spanish in a wide spectrum of oral and written records from 1975 to 2000. This is the period when CTB was published, and thus represents the norm of its contemporary readers and writers. To obtain a list of key clusters, only the subcorpus of Spanish and Latin American fiction (32,613,011 tokens), was selected as our reference corpus.

I based the cluster analysis on the ST and focused on the instances which include at least two repeated tokens for each type. Two keyness measures were applied to the initial list of 202 clusters in Fuster-Márquez's study (2020) to obtain key clusters: Log Likelihood (LL) and Log Ratio for effect size. These calculations were automatically performed online by means of the Log Likelihood and Effect Size calculator. We found it convenient to combine LL, a widely employed statistical measure of significance in CL, with an effect size measure like Log Ratio, to take into consideration the size of the frequency difference between the study corpus and the reference corpus. We applied high critical values, 15,13 for LL (significant at the level of p < 0.0001) (see Rayson, Berridge and Francis, 2004), and for Log Ratio, 5,16. This meant that 22 out of the initial 202 clusters had to be discarded, thus, we were left with a list of 180 key clusters. All discarded items were five and six-word clusters.

A focus on long clusters is quite unusual in corpus stylistics studies but very suitable to analyse *CTB*. A threshold of five or more words allowed us to avoid common phraseology in Spanish fiction, usually linked to shorter clusters. Such sequences most often have grammatical functions but few meaningful associations, as observed in bigrams and trigrams that are very frequent in the ST: *lo que* (423 cases), *de la* (354 cases), *de lo que* (69 cases), etc. Four-word clusters show an intermediate behaviour. Many still retain this grammatical character, such as *a lo largo de* (9 cases), *de vez en cuando* (15 cases), *y a la vez* (11 cases). By contrast, sequences of five or more words point to

⁵ https://www.rae.es/banco-de-datos/crea [last accessed 01/08/2021]

⁶ http://ucrel.lancs.ac.uk/llwizard.html [last accessed 04/08/2021]

greater uniqueness and stylistic relevance in the ST, even when their frequency is as low as 2 cases. Interestingly, most of the key clusters identified in this study were found only in Marias' *CTB* in CREA. In fact, all clusters of 7 words and longer have the ST as the only source in this reference corpus.

One of the most challenging aspects of the selected clusters has been the determination of variation. Our decision has been to restrict this research to cases of low phraseological variation and complexity. Indeed, often variation affects single words. Less frequently, particularly in the case of very long clusters, variation can affect more words but, in most cases, only a particular feature, as in Example 1, where the differences between the two occurrences affect the underlined words, the singular is used in chapter 4, and plural forms the second time, in chapter 15:

(1) y en medio de la noche, al despertar <u>sobresaltado(s)</u> por una pesadilla o ser incapaz de conciliar el sueño al padecer una fiebre o creerse/creernos <u>solo(s)</u> y <u>abandonado(s)</u> a oscuras, no tiene/tenemos más que <u>darse/darnos</u> la vuelta y ver de frente, el rostro que le/nos protege, que se dejará besar lo que en el rostro es besable [...] (chapters 4 and 15)

Although variation may also involve changes of longer segments, most often these are the result of assembling different clusters collected in this study, thus forming longer patterns of partial repetition.

7. Analysis of the ST

The recognition of the quantitative features in the ST, as shown in Tables 1, 2 and 3 are crucial to understand the choices made by the translator, therefore we will look at the figures in detail. Table 1 shows that Marías often makes use of long clusters and favours the alternation between sequences both with and without variation.

Table 1 shows an even proportion in terms of size amongst the 180 key cluster types, according to the cluster divisions which were established beforehand. These 180 types correspond to 548 tokens in the text. Some 5-word clusters might look unsurprising and conventional, such as a mi me lo contaria (2 tokens), or el (mismo) dia de mi boda (6 tokens); however they are significant in this narrative. Even more significant are clusters such as I have done the deed (5 tokens), la verdad es que no, or la mujer de la calle (4 tokens), los actos se cometen solos, or escuchar es lo más peligroso (3 tokens each), etc., since they introduce narrative themes and motifs, such as murder, knowledge or ignorance of facts, etc. Most key clusters have six or more words as, for example, the 7-word cluster todo es cuestión de distancia y tiempo (3 cases); the 12-word cluster la lengua en la oreja es también el beso

que más convence (2 cases). Finally, Example 2 shows a 55-word cluster, classified as containing twenty or more words.

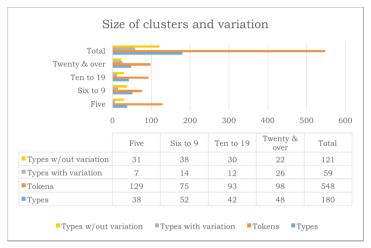


Table 1: Cluster size and variation in ST

(2) volcamos toda nuestra inteligencia y nuestros sentidos y nuestro afán en la tarea de discernir lo que será nivelado, o ya lo está, y por eso estamos llenos de arrepentimientos y de ocasiones perdidas, de confirmaciones y reafirmaciones y ocasiones aprovechadas, cuando lo cierto es que nada se afirma y todo se va perdiendo. (3 cases, chapters 3, 11, and 16 in ST)

This high percentage of long clusters, especially cases like Example 3, shows that Marías performed an incredible task of phraseological storage so that he would conveniently distribute this phraseology in different sections of the ST. As Table 1 shows, 121 cluster types have no internal variation, while 59 suffered alterations of different kinds according to our procedure. Table 1 also shows that variation is related to size, i.e., longer sequences show greater variation than shorter ones. In some cases, variation consists of omissions, additions, also change of inflection, for example, adapting verb conjugations. Example 3 is a 42-word cluster where the only kind of variation is the addition of the inflected verbal form *pensé* (E. thought), in square brackets. This sequence is found in chapters 5 and 15 in the ST:

(3) Escuchar es lo más peligroso, [pensé] es saber, es estar enterado y estar al tanto, los oídos carecen de párpados que puedan cerrarse instintivamente a lo pronunciado, no pueden guardarse de lo que se presiente que va a escucharse, siempre es demasiado tarde. (chapters 2 and 15 in ST)

An interesting feature that emerges in our findings is that a very long cluster, like Example 3, may host shorter clusters, in this case the 5-word cluster escuchar es lo más peligroso (TT. 'listening is the most dangerous thing'). This kind of embeddedness affects fourteen clusters within the groups of 5, and also 6 to 9 words. The reasons for including them separately in this study is that shorter cluster types show higher token frequencies than longer ones. Therefore, this is the source of another kind of variation in the ST, where, at times, the author reuses shorter clusters in more elaborate longer sequences which enjoy lower frequency rates. Table 2 shows the results of combining size and frequency of clusters in the ST.

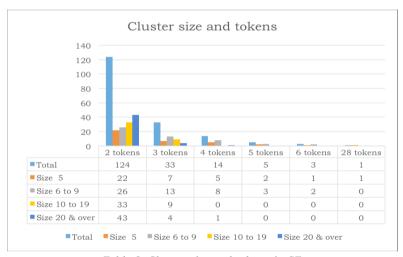


Table 2: Cluster size and tokens in ST

Note that most types – 69% – have only two tokens. This is especially true for longer clusters, where 43 out of a total of 48 are found twice. This occurs also in the group of 10-to-19-word clusters, where 33 out of a total of 42 have two tokens. Exceptionally, we found five or more tokens assigned to the shorter clusters (5-word clusters and 6-to-9-word clusters). So, for example, *en el cuarto de baño* (E. 'in the bathroom') 11 tokens, and *no me lo cuente si no quiere* (E. Don't tell me if you don't want to) 5 cases. These sequences are thematically more relevant in the ST, the former as a referential point to the place where Teresa committed suicide in chapter 1, and the latter as a reference to secrecy.

Finally, Table 3 shows how clusters have been introduced in the course of the narrative. In fact, this table consists of two parts; on the one hand, we have noted whether the repeated sequence occurs in one or more chapters and, on the other, how many clusters reach the last

two chapters, since these two contain Ranz's confession of his murder, the dénouement (chapter 15), and a reflective epilogue (chapter 16).

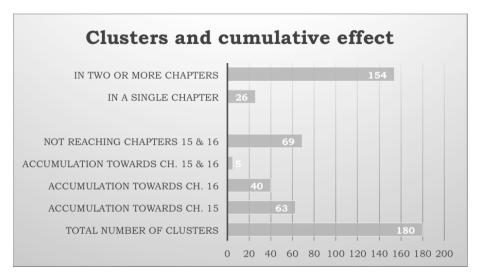


Tabla 3: Presence of clusters in chapters: cumulative effect

These results show that, regardless of size, readers find repeated cluster types in more than one chapter (154 out of 180). It seems to indicate that the author wanted his readers to establish mental associations as the context is gradually being transformed from one chapter to another, from an earlier situation to a newer one. A case in point is the Shakespearean sequence I have done the deed (5 tokens), used in chapters 5, 12, 13 and 15. Further, Table 3 shows that among those sequences that occur in more than one chapter, more than half of them cumulatively reach chapters 15 and 16. To be precise, 74 tokens reappear in chapter 15, the chapter that hosts Ranz's confession of murder to Luisa while the narrator is overhearing the conversation. Many of the longer sequences, which show lower frequencies also reach this penultimate chapter. Likewise, the analysis of sequences in separate chapters by means of the plot technique show strong connections, for example, between clusters found in chapters 5 and 9 with chapter 15. Chapter 5 deals with the narrator's reflections and translation attempts around Macbeth. Two important themes expressed by clusters in chapter 5 are murder and instigation. Chapter 9 deals mostly with conversations between the narrator and his wife, Luisa, when they are back in Madrid. On the other hand, chapters 2 (the narrator's honeymoon), 3 (the minor story of Miriam and Guillermo), and 6 (dealing with the narrator's wedding and marriage), are closely connected to chapter 16, the epilogue.

Some important issues emerge from the quantitative analysis of the ST in this study. Arguably, the most important one is that Marias made a significant use of an unusually large number of long clusters. Also, that the size of the ngrams and internal variation could pose a significant challenge to the translator. Variability, embeddedness of shorter clusters into longer ones, and the presence of clusters in different chapters clearly mean that repeated sequences are applied and adapted to suit different contexts as the narrative progresses. This repetition is used to emphasise the narrator's memories as echoes of past moments, and foreshadowing events yet to come in his reflections. Textual cohesion and text unity are two overall significant stylistic effects mainly achieved by the accumulation of clusters in different chapters. Particularly, the great accumulation of clusters in chapters 15 and 16, collected from earlier occurrences in *CTB*, stresses out the structural unity and circularity of the text.

8. Treatment of ST clusters in the TT

My aim has been to judge if or to what extent the 180 key clusters of the ST have been preserved verbatim in the TT. In this section we will examine the phraseological translating solutions provided by Jull Costa in her English translation AHSW. Like Čermáková (2015: 364), we hypothesise that these longer cases of repeated sequences "constitute an intended stylistic device that should not escape the translator's attention." We have quantified and scrutinised more closely the translator's treatment of long clusters. The results show that out of the 180 key clusters in the ST, in 107 cases (59%), the translator provides equivalent sequences in the TT. Leaving aside the intrinsic quality of the wording in the TT, my purpose has been to analyse whether the identified key clusters in the ST are also equally repeated in the TT. Therefore, we have here the answer to RQ1, the translator only partly reproduces these long clusters. That group includes clusters of various sizes, like tres veces es mucho azar (4 cases) ('three times is a lot of coincidences' in TT); or clusters with two occurrences each: Bill, puede que sea mi nombre definitivo (Bill, which might be my definite name' in TT); A los toxicómanos y a los delincuentes futuros (The drug addicts and the future delinquents', in TT); lo que se da es idéntico a lo que no se da, lo que descartamos o dejamos pasar idéntico a lo que tomamos y asimos, lo que experimentamos idéntico a lo que no probamos (What takes place is identical to what doesn't take place, what we dismiss or allow to slip by us is identical to what we accept and seize, what we experience identical to what we never try' in TT). In this group we would include clusters which show variation and which Jull Costa attempts to reproduce, as in Example 4 (two occurrences), where the words in brackets indicate this variation:

(4) [supiera/sabe] desde el primer momento quien quería ser abordado y con qué propósito, en un local o en una fiesta o incluso en la calle. TT '[as if he knew/ who knows] instantly who wanted to be approached and why, in a bar or at a party or even in the street.'

Often this variation consists in small additions, which, like sign-posts, announce the presence of the repeated sequence. This may be indicated by means of the reporting verb "decir" (to say) or the verb "pensar" (to think), which hints at the reflexive mood of the narrator, highlighted by critics of Marías' novel. This stylistic device is used to bring some kind of memorised event or thought known the reader has come across, as shown in example 5, where "me dijo" (in brackets) is introduced the second time the sequence is reproduced, and "pensé" in example 6:

- (5) Nuestra única posibilidad es que un día muriera ella [me dijo] y con eso no puede contarse.
 - TT 'Our one possibility is if she should die,' [she said to me], 'and we can't rely on that.
- (6) lo estuve mirando durante unos segundos, [pensé] como crecía y se iba ensanchando el círculo, una mancha a la vez negra y ardiente que se comía la sábana.
 - TT 'And I watched for a few seconds, [I thought], watched how the circle grew and widened, a stain that was at once black and fiery, consuming the sheet'.

In the remaining 73 clusters (41%), however, the TT shows various degrees of departure from the ST. Departures are found in sequences of different sizes with or without variation. In the three cases of *voy por ti o yo te mato* in the ST, the TT shows some differences: (1) "I'll get you or I'll kill you"; (2) "I'm gonna get you, I kill you"; (3) "I'll get you, I kill you". On one occasion, the 7-word cluster *la boca está (siempre) llena y es la abundancia* (3 tokens in ST), which shows variation through the addition of "siempre", is rendered as invariable in the TT, as "The mouth is always full, abundance itself", thus the TT is more repetitive than the ST.

Differences between the ST and the TT are shown by sequences of various sizes, as in the 7-word cluster *pese al olvido de las reglas sociales* (2 tokens) found in the TT as "despite the abandonment of all social niceties", or "despite the lack of social niceties", where 'all' is an addition in the TT; or in the 5-word cluster *hombre atezado con sombrero* (2 tokens in ST), found in the TT as "a man with a tanned face wearing a hat" and the second time as "the swarthy-complexioned man: he was wearing an old hat", where 'old' would also be an addition in the TT. A departure is also noticeable in the 10-word cluster *aún no ha claudicado ni se ha dado de baja*, translated once as 'she hasn't

given up on herself or written herself off entirely', and the second time as 'She hasn't lost her nerve or given up'. As reflected in some of the examples above, often what the TT shows is minor differences from the ST.

Particularly significant are those clusters that reach chapters 15 or 16. For example, escuchar es lo más peligroso is found on three occasions in chapters 5, 12 and 15. It revolves around the theme of getting to know what the narrator apparently does not wish to know, the fact that his father is a murderer, and has an obvious effect of foreshadowing. The TT offers two different renderings, twice as "Listening is the most dangerous thing of all", and once as "Listening is the most dangerous of acts". Similarly, la lengua en la oreja es el beso que más convence (chapters 5 and 15 in ST) has these two renderings in the TT: "The tongue in the ear is also the kiss that most easily persuades" and "The tongue in the ear is always the most persuasive of kisses".

An interesting departure, probably motivated by awareness of cultural distance between the Spanish context and the target English readership relates to the cluster *en el casino de Alcalá 15* (chapters 12 and 15). This is a highly specific reference, where "Alcalá" is well-known to Spaniards when referring to Madrid. The TT omits the street name and offers instead two different renderings: "in the Casino de Madrid" and "in the old Casino de Madrid", to make clear for the audience that the "Casino" is in the Spanish capital, note also the free addition of "old".

Finally, an interesting group of key clusters in the ST that deserves some attention has its origin in quotations from the play Macbeth (chapter 5). For example, the ST shows the 5-word cluster "I have done the deed", with a clear reference of the word "deed" to murder, and also one of the most important themes in CTB, where Macbeth's murder is associated with the murder of Ranz's first wife. This same cluster, worded in English, is also embedded in a much longer sequence, together with clusters where Marías gives the Spanish translation y ya está hecho el hecho. Understandably, Juli Costa reproduces the wording found in Shakespeare's play invariably as "I have done the deed". Another quote taken from Macbeth in the ST is the 10-word cluster "Los dormidos y los muertos no son sino como pinturas", with three tokens. Previously, Marías has made use on one occasion of the English sentence: The sleeping and the dead are but pictures. That is the reason why the TT shows four tokens of this same quotation, all of them worded in English.

Therefore, as shown above, the answer to RQ2 is that there is more than one reason for not reproducing partly these long clusters in the TT. Indeed, at times the translator could be even more repetitive than Marías. Therefore, there is no apparent desire to be less repetitive.

9. Conclusions

This study has explored the contribution of corpus linguistics to stylistics and literary translation studies. This approach has been author-based, and its aim to seek for equivalences between the ST and the TT. Priority has been given to the identification of features in the ST to seek equivalence (or lack of) in the TT. My aim has been to look at phraseological equivalence, most exclusively I focused on the analysis of long clusters, a prominent stylistic feature in the ST. While various other kinds of repetition deserve due attention, the results suggest that repeated clusters of five or more words is a remarkable stylistic element of CTB. Unarguably, Marias would have been aware of the effects of such clusters in his work. It is hard to imagine that Marías would not have been very fully conscious of their existence the second or subsequent times they were used (not a few are sequences longer than 20 words). The sheer number of clusters and their complexity prove the challenge for translators of this work to any language. A thorough knowledge of the phraseological choices in the ST is crucial if translators wish to reproduce equivalent stylistic effects in their own translations. In this regard, Margaret Jull Costa was also quite aware of the overall stylistic effects achieved by Marías' phraseological choices. Instead of avoiding repetitiveness, she attempted to maintain this phraseology in a TT which would not be stylistically distant from the ST. Long clusters were preserved in the translation in half of instances. Differences that arise between ST and TT are often minor, or else can be interpreted as culture-specific differences between the source and target languages which would have required a different strategy. Nevertheless, it is hard to find a clear motivation for lack of equivalence in some cases. Certain phraseological differences would reveal the translator's desire to offer some sort of stylistic variation, thus avoiding excessive repetition, a universal feature in translated works. However, not a few clusters may contain well over ten words, and many of the clusters in the ST show very few occurrences. In addition, the existence of phraseological variation adds greater complexity to the task of translating. Therefore, it could also be claimed, as an alternative explanation, that Jull Costa could not see all this phraseology in the ST. Finally, this study has tried to prove that corpus methods can be most useful not only in literary translation studies, as we have tried to do here, professional literary translators can also benefit from their use.

10. References

- ABDULLA, A. K. (2001): "Rhetorical repetition in literary translation", *Babel*, 47(4), pp. 289-303. https://doi.org/10.1075/babel.47.4.02abd
- AITCHISON, J. (1994): "Say, say it again Sam": The treatment of repetition in Linguistics", Jucker, A. (ed.), *Repetition*, Tübingen: Gunther Narr Verlag, pp. 15-34.
- ANTHONY, L. (2020): AntConc (Version 3.5.9): Waseda University. http://www.laurenceanthony.net/software
- BASSO, E. (2002): "La repetición en Corazón tan blanco de Javier Marías", *Fragmentos*, 23, pp. 11-40.
- BAKER, M. (1993): "Corpus linguistics and translation studies: Implications and applications", Baker, M., G. Francis, G., and Tognini-Bonelli, E. (eds.), *Text and technology: In honour of John Sinclair*, Philadelphia and Amsterdam: John Benjamins, pp. 233-250.
- BAKER, M. (2000): "Towards a Methodology for Investigating the Style of a Literary Translator", *Target*, 12(2), pp. 241-66. https://hathanhhai.files. wordpress.com/2012/11/towards-a-methodology-for-investigating-the-style-of-a-literary-translator.pdf
- BEN-ARI, N. (1998): "The ambivalent case of repetitions in literary translation. Avoiding repetitions: A 'universal' of translation", *Meta*, 43(1), pp. 68-78. https://doi.org/10.7202/002054ar
- BOASE-BEIER, J. (2006): Stylistic Approaches to Translation, Manchester: St. Jerome.
- BOUGUEN, C. (2000): Secret et vérité dans l'oeuvre de Javier Marías, Unpublished doctoral dissertation: UTM, Toulouse.

- ČERMÁKOVÁ, A. (2015): "Repetition in John Irving's novel A Widow for One Year A corpus stylistics approach to literary translation", International Journal of Corpus Linguistics, 20(3), pp. 355-377. https://doi.org/10.1075/ij-cl.20.3.04cer
- ČERMÁKOVÁ, A. (2018): "Translating children's literature: some insights from corpus stylistics", *Ilha do Desterro*, 71, pp. 117-133. https://doi.org/10.5007/2175-8026. 2018v71n1p117
- FLORENCHIE, A. (2003): La répétition dans l'oeuvre de Javier Marías, Unpublished doctoral dissertation. Littératures. Université Michel de Montaigne. Bordeaux III. https://tel.archives-ouvertes.fr/tel-00842119/document.
- FUSTER-MÁRQUEZ, M. (2020): ""Con tan enfermo cerebro": Fraseología recurrente en *Corazón tan blanco* de Javier Marías", *Círculo de Lingüística Aplicada a la Comunicación*, 83, pp. 51-64. https://doi.org/10.5209/clac.70563
- GARVIN, P. L. (1964): A Prague school reader on aesthetics, literary structure, and style, Georgetown: Georgetown University Press.
- GREGORI-SIGNES, C. (2020): "Caracterización mediante formas denominativas: una aproximación estilística de corpus a Corazón tan blanco (1992) de Javier Marías", Circulo de Lingüística Aplicada a la Comunicación, 82, pp. 137-150. https://doi.org/10.5209/clac.68969
- GREGORIOU, C. (2014): "The linguistic levels of foregrounding in stylistics", Burke, M. (ed.), *The Routledge Handbook of Stylistics*, Abingdon and New York: Routledge, pp. 87-100.

- GROHMANN, A. (2002): Coming into one's own: The Novelistic Development of Javier Marías, Amsterdam: Rodopi.
- HU, X., XIAO, R. & HARDIE, A. (2019): "How do English translations differ from non-translated English writings? A multi-feature statistical model for linguistic variation analysis", Corpus Linguistics and Linguistic Theory, 15(2), pp. 347-382. https://doi.org/10.1515/cllt-2014-0047
- HALLIDAY, M. A. K. and HASAN, R. (1976): Cohesion in English, London: Longman.
- JEFFRIES, L. and MCINTYRE D. (2010): Stylistics, Cambridge, C.U.P.
- JOHNSON, J. H. (2016): "A comparable comparison? A corpus stylistic analysis of the Italian translation of Julian Barnes' Il Senso di una Fine and the original text The Sense of an Ending", Language and Literature, 25(1), pp. 38-53. https://doi.org/10.1177/0963947015623360
- JULL COSTA, M. (2012): Javier Marias. A Heart So White, London: Penguin Books.
- LEECH, G. N. & SHORT, M. (2007, 2nd ed.): Style in fiction: A linguistic introduction to English Fictional Prose, London: Pearson Education.
- MAHLBERG, M. (2007a): A corpus stylistic perspective on Dickens' *Great Expectations*, Lambrou, M. and Stockwell, P. (eds.), *Contemporary Stylistics*, London: Continuum, pp. 19-31.
- MAHLBERG, M. (2007b): "Clusters, key clusters and local textual functions in Dickens", *Corpora*, 2(1), pp. 1-31. https://doi.org/10.3366/cor.2007.2.1.1
- MAHLBERG, M. (2007c): "Corpora and translation studies: Textual function of lexis in *Bleak House* and in a translation of the novel into German", Intonti,

- V., G. Todisco, G. and Gatto, M. (eds.), *La Traduzione. Lo Stato dell'Arte. Translation. The state of the Art*, Ravenna: Longo, pp. 115-135.
- MAHLBERG, M. (2013a): Corpus stylistics and Dickens's fiction, London: Routledge.
- MAHLBERG, M. (2013b): "Textual building blocks of fictional worlds", Mahlberg, M. (ed.), Corpus stylistics and Dicken's Fiction, London: Routledge, pp. 27-41.
- MAHLBERG, M. (2014): "Corpus Stylistics", Burke, M. (ed.), *The Routledge Handbook of Stylistics*, London: Routledge, pp. 378-392.
- MAHLBERG, M. & MCINTYRE, D. (2011): "A Case for Corpus Linguistics: Ian Fleming's Casino Royale", English Text Construction, 4(2), pp. 204–227. https://doi.org/10.1075/etc.4.2.03mah
- MAHLBERG, M., CONKLIN, K. & BIS-SÓN, M. (2014): "Reading Dickens' Characters: Employing psycholinguistic methods to investigate the cognitive reality of patterns in texts", *Language and Literature*, 23(4), pp. 369-388. https://doi.org/10.1177/0963947014543887
- MARÍAS, J. (1996): *Corazón tan blanco*, London: Penguin Random House.
- MASTROPIERRO, L. & MAHLBERG, M. (2017): "Key words and translated cohesion in Lovecraft's *At the Mountains of Madness* and one of its Italian translations", *English Text Construction*, 10(1), pp. 78–105. https://doi.org/10.1075/etc.10.1.05mas
- MCINTYRE, D. & WALKER, B. (eds.) (2019): Corpus Stylistics: Theory and Practice, Edinburgh: Edinburgh University Press.
- MIALL, D. S. & KUIKEN, D. (1994): "Foregrounding, defamiliarization, and affect: Response to literary theories", *Poetics*, 22, pp. 389-407.

- NAVARRO GIL, S. (2004): "Una aproximación al estilo literario de Javier Marías", *Revista de Filología*, 22, pp. 187-193. http://riull.ull.es/xmlui/hand-le/915/19536
- PÉREZ-CARBONELL, M. (2016): The fictional world of Javier Marias: Language and Uncertainty, Leiden: Rodopi.
- RAYSON P., BERRIDGE, D. & FRANCIS, B. (2004): "Extending the Cochran rule for the comparison of word frequencies between corpora", Purnelle, G., Fairon, C. and Dister, A. (eds.), Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data (JADT 2004), Louvain: Presses universitaires de Louvain, pp. 926-936.
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. Corpus de referencia del español actual. https://www.rae.es/ banco-de-datos/crea
- SALDANHA, G. (2011a): "Translator Style: Methodological Considerations", *The Translator* 17(1), pp. 25–50. https://doi.org/10.1075/tis.19067.sal
- SALDANHA, G. (2011b): "Style of translation: The use of foreign words in translations by Margaret Jull Costa and Peter Bush", Kruger, A., Wallmach, K. and Munday, J. (eds.), Corpus-Based Translation Studies: Research and Application, London: Continuum, pp. 237–258.
- SALDANHA, G. (2014): "Style in, and of, translation", Bermann, S. and Porter, C. (eds.), A Companion in Translation Studies: Research and Applications, Chichester: Wiley-Blackwell, pp. 95-106. https://doi.org/10.1002/9781118613504. ch7
- SALDANHA, G (2021): "The translator: literary or performance artist?", *Translation and Interpreting Studies*, 16 (1), pp. 61-

- 79. https://doi.org/10.1075/tis.19067.sal
- SEMINO, E. & SHORT, M. (2004): Corpus Stylistics. Speech, Writing and Thought Presentation in a Corpus of English Writing, London: Routledge.
- SCOTT, M. (1997): "PC analysis of key words –and key key words", System, 25(2), pp. 233-45. https://doi.org/10.1016/S0346-251X(97)00011-0
- SIMPSON, P. (2004): Stylistics: A resource book for students, London: Routledge.
- STEEN, M. S. (2011): "La repetición en Corazón tan blanco ¿Forma de instalar una ética?", España contemporánea: Revista de literatura y cultura, 24(1), pp. 21-32.
- SZEGEDY-MASZÁK, M. (1983): "Forms of repetition in music and literature", *Yearbook of Comparative and General Literature*, 32, pp. 39-49.
- STUBBS, M. (2005): "Conrad in the computer: examples of quantitative stylistics methods", *Language and Literature* 14(1), pp. 5-24.
- VILA SÁNCHEZ, J. A. (2015): El estilo sin sosiego. La génesis de la poética integradora de Javier Marías: 1970-1986, Unpublished doctoral dissertation, Barcelona: Universitat Pompeu Fabra.
- WINTERS, M. (2018): "The case of Natasha Wodin's Autobiographical Novels: A Corpus Stylistics Approach", Boase-Beier, J., Fisher, L. & Furukawa, H. (eds.), The Palgrave Handbook of Literary translation, Cham: Palgrave, pp. 145-166.