USING CORPUS PATTERN ANALYSIS FOR THE STUDY OF AUDIOVISUAL TRANSLATION: A CASE STUDY TO ILLUSTRATE ADVANTAGES AND LIMITATIONS

BLANCA ARIAS-BADIA
blanca.arias@upf.edu
Universitat Pompeu Fabra

Resumen

La identificación de normas, es decir, de patrones recurrentes de comportamiento, es un objetivo común de la Lingüística, los Estudios de Traducción y los Estudios de Televisión. Este artículo ofrece una revisión metodológica crítica de la aplicación del análisis de patrones de corpus (CPA), una técnica propuesta desde el campo del análisis léxico, al estudio de la traducción audiovisual. El artículo ilustra la aplicación de esta técnica de investigación mediante la presentación de un estudio de caso que busca la identificación de combinatoria léxica anómala en el diálogo en inglés y en los subtítulos en español de tres series de televisión por medio de CPA. Después de ilustrar la metodología, el énfasis de la evaluación recae en la reflexión acerca de las ventajas y limitaciones que el uso de CPA conlleva para el estudio de la traducción audiovisual.

Palabras clave: Análisis de patrones en corpus; Traducción audiovisual; Subtitulación; Combinatoria léxica anómala; Estudios Descriptivos de la Traducción.

Abstract

The identification of norms, i.e. of recurrent patterns of behaviour, is a common concern for Linguistics, Translation Studies, and Television Studies. This paper offers

a critical methodological review after applying Corpus Pattern Analysis (CPA), a technique proposed from the field of lexical analysis, to the study of audiovisual translation. The paper illustrates the application of this research technique by presenting a case study pursuing the identification of anomalous collocations in the English spoken dialogue and in the Spanish subtitles of three television series by means of CPA. After illustration, emphasis of the methodological assessment is placed on the discussion of advantages and limitations entailed in the use of CPA for the study of audiovisual translation.

Keywords: Corpus Pattern Analysis; Audiovisual translation; Subtitling; Anomalous collocates; Descriptive Translation Studies.

1. Introduction[1]

The identification of norms, i.e. of recurrent patterns of behaviour, is a common concern for Linguistics and Translation Studies. Corpus linguistics has been proved to serve as a powerful methodological tool or research methodology (Olohan 2004; Walsh, Morton & O'Keeffe 2011) to account for linguistic and translational norms. Corpus-driven approaches in Translation Studies propose a bottom-up analysis of lexicogrammatical and phraseological patterns in source texts and translations by keeping preconceptions about the texts in hand to a minimum (Corpas 2008: 53).

Television Studies are also interested in norms. The discipline has been nourished by Literary Theory on the notion of genre and reflects on shared conventions governing television text production (Fiske 1987, Feuer 1992, Neale and Turner 2001). It is well known that genre entails recognition (Wolf 1984: 190): audiences of crime TV shows expect to see weapons, uniforms and handcuffs on screen as much as they expect to hear sirens, doors howling or shouts in every episode. Genre and format norms provide the necessary framework for recognition in the television medium.

This paper assesses advantages and limitations of applying a methodology proposed from the field of Linguistics and, more specifically, from

---

lexical analysis, to the study of audiovisual translation. Namely, Corpus Pattern Analysis (CPA) (Hanks 2004, 2013) is the methodological framework proposed in order to determine when screenwriters or subtitlers opt either for normal usage, i.e. "the prototypical syntagmatic patterns with which words in use are associated" (Hanks 2004: 87), or for creativity in television dialogue and its translation.

The following sections are structured as follows: Section 2 starts by framing CPA; Section 3 reports on a case study conducted using CPA; Section 4 summarises the conclusions drawn from applying this technique to a corpus of audiovisual translation; and Section 5 provides closing remarks.

2. Corpus Pattern Analysis (CPA)

The case study reported in Section 3 employs notions from the Theory of Norms and Exploitations (TNE). This approach to the study of lexicon is based on the idea that, when "making meanings", the users of a language either conform to norms, i.e. "pattern[s] of ordinary usage in everyday language with which a particular meaning or implicature is associated" (Hanks 2013: 92), or exploit such norms creatively by means of "a deliberate departure from an established pattern of normal word use, either in order to talk about new or unusual things or in order to say old things in a new, interesting, or unusual way" (*ibid*: 212). Since they are instances of unpredictable uses of language, exploitations may be the result of various linguistic processes. Hanks (*ibid*: 215-250) provides a description of some kinds of exploitations and their link to classical tropes. The present paper focuses on one specific creative device, i.e. anomalous collocates, which will be defined in Section 3.

As a methodological application of TNE, Corpus Pattern Analysis (CPA) aims to build an inventory of normal patterns of word use. Its "objective is to identify, in relation to a given target word, the overt textual clues that activate one or more components of its meaning potential" (Hanks 2004: 6). The analytic procedure to apply CPA is the following: first, concordances for each target word are scanned; then a random sample is scrutinised and annotated by the lexicographer. This methodology is ultimately based on Sinclair's (2003) proposal of a procedure for the analysis of corpus concordances. In his

seven-step guidelines, the author notes that interpretation (i.e. formulation of hypotheses after concordance reading and before consolidation by means of consultation in other sources) is an inextricable part of corpus analysis. On the one hand, CPA annotation focuses on the identification and classification of normal patterns of behaviour; on the other, it accounts for exploitations of each pattern found in the random sample under analysis (*ibid*: 9).

The *Pattern Dictionary of English Verbs* (PDEV) is currently being developed following this methodology. Verbs have been given priority in applying CPA because they are the pivot of the clause and "we make conversation by using clauses" (Hanks n.d.: 12). Each entry in the PDEV shows the patterns of the verb found on a sample of at least 250 corpus occurrences —the more frequent the verb, the larger the sample— and the implicature of each pattern. The pattern displays normal semantic types in combination with the verb and the implicature glosses the meaning of the verb for each pattern. Consider the pattern and implicature of the verb *listen* below:

> **Pattern:** Human | Institution *listen* to Human 2 | Proposition | Sound | Performance
> **Implicature:** [[Human | Institution]] concentrates on hearing or paying attention (to [[Human 2]] | [[Proposition]] | [[Sound]] | [[Performance]])

Thus, a normal use of the verb would be found in a sentence such as *Dolores* [[Human]] *is listening to the New Year's Concert* [[Performance]]. But this pattern may be exploited for pragmatic purposes —consider the imperative sentence *Listen to the chocolate* on the website of a chocolate boutique in Australia.[2] The noun *chocolate* does not fall into any of the lexical type categories described above. The object semantic type has been replaced by the type [[Food]], thus causing pattern exploitation. Even the authors of this creative sentence realize its meaning may not be clear for the reader and specify: "No, this isn't listening to the chocolate saying 'Eat me, Eat me!'. Break your chocolate and listen for a clear 'snap', indicating the chocolate is in a good-tempered state, and cocoa butter is stable". The meaning becomes clearer now. And this is precisely what exploitations are about: they are instances of creative language, but not so creative as to hinder effective

---

2. <https://josophans.com.au/pages/faqs> [30.05.2020].

communication. Curiously enough, such is also the aim of screenwriters, who teach screenwriting trainees to "make […] words seem lifelike, but not so much that they impede the flow of meaning" (Wolff & Cox 1988: 56).

3. The case study

This section offers a case study to illustrate the application of CPA to the study of audiovisual translations. The study was part of my doctoral dissertation, which pursued a holistic description of the language used in contemporary television series of the crime genre and their translation. The dissertation employed a combination of corpus-based and corpus-driven methodologies to report on more than twenty lexical and morphosyntactic features. The results of the whole study can be found in Arias-Badia (2020), and the initial methodological proposal was presented in Arias-Badia (2015). In this paper, only a synthesis of the approach to anomalous collocation (one of the studied features) is provided to clarify the way in which CPA may be used for the purposes of audiovisual translation research.

3.1. Framing the study

This case study is concerned with anomalous collocation in television dialogue and the way in which it is rendered in interlingual subtitling. To date, the specialized literature has not agreed on a single definition of the notion of collocation. Emphasis of the notion is placed on the idea that words show restrictions when it comes to combining them with one another. Criteria to establish which lexical combinations involve collocation range from educational proposals arguing that combinations prove useful for the language learner (Sinclair 1991, Lewis 1993) to purely statistical measures accounting for lemma cooccurrence in general usage corpora (Kilgarriff *et al.* 2004) (Torner and Arias-Badia, in press). Using the CPA framework, for the purposes of this paper the term will be understood as word combinations that fit the patterns of normal use. In the present study, by contrast, I focus on instances of collocation that are *not* predictable considering the normal patterns of each of the collocates. As defined by Leech (1990/1974: 17), they are "improbable combinations". According to Hanks (2013: 217), anomalous collocates are one type of language exploitation that can be defined

as "noncanonical members of a lexical set", which means that they do not follow the rules of semantic selection.

Within audiovisual translation, subtitling is defined as a translation modality in which a written text is added to the source audiovisual content, to account for the linguistic elements of the source text by keeping synchrony with the moving picture (adapted from Matamala 2019: 127). Interlingual subtitles are hybrid in nature in that they a) transfer semantic content from one language to another; b) involve medium conversion, i.e. spoken to written; and c) are constrained by temporal and spatial parameters well spread in the practice of subtitling —although parameters vary among different clients and companies, subtitles must be typically displayed for 1-6 seconds on screen, have one or two lines, and make use of up to 36-42 characters per line—, thus being well-known for producing condensed translations (Arias-Badia, 2020, p. 27). Those factors have been said to affect lexical selection, since a pragmatic-oriented translation is favoured in which "the speech act is always in focus; intentions and effects are more important than isolated lexical items" (Gottlieb 1998, p. 247). The literature converges on the idea that the use of conventional or standardised lexicon, to the detriment of creative language or language deviating from the norm, is frequently the result of such an approach in subtitling (Zaro 2001; Díaz-Cintas 2003; Bartoll 2012).

Not *all* translation solutions in subtitling, however, are based on a reduction or condensation principle, since sometimes the subtitles are longer than the creative proposals of the ST. Consider Example (1) from our corpus:

(1) DM   ST: Pretty fucking please with cheese on top. [D01, 00:09:08][3]
          TT: Un jodido favor de los gordos|con queso encima.
          BT: A fucking favour of the fat ones|with cheese on top.

In (1a), Dexter's sister Debra produces a sentence in which she assigns the semantic type [[Surface]] to the interjection *please*, thus attributing the

capacity to hold "cheese on top" to the interjection. This creative phrase is to some extent preserved in the TT, although the translator resorts to the noun *favor* ('favour') instead of using the Spanish interjection *por favor* ('please'). The subtitle lasts 2,5 seconds, which means that, ideally, its length should not exceed 30 characters (Bartoll 2012: 136). However, in this case, the subtitler seems to give priority to the creativity of the character over conventional professional guidelines and offers a 46-character-long two-liner.

Thus, the aim of this case study is twofold: on the one hand, it intends to establish the degree of occurrence of this form of lexical creativity in the source and target texts, namely the first two episodes of the first season of the television series *Castle* (2009), *Dexter* (2006) and *The Mentalist* (2008) and their subtitles published on DVD format —the corpus extension is of 36,995 words (ST) and 34,019 words (TT). On the other, it sets out to describe the types of translation solutions used to render anomalous collocation in subtitling. Unusual collocations have been paid little attention in audiovisual translation. An exception is the work of Teixeira (2015), who looks at the English-Portuguese subtitling of idiomatic collocational patterns. Díaz-Cintas & Remael (2007: 177), for their part, provide advice on collocates segmentation in subtitling. One of the few discussions on anomalous collocation in audiovisual translation can be found in Chaume (2004), when he describes the lexical combination in one of his examples as "odd".

3.2. Methodology: Adapting CPA for the study of lexical exploitation in audiovisual translation

Section 2 has described the analytic procedure of CPA for lexicographic purposes. Beyond the *Pattern Dictionary of English Verbs*, such a procedure has been used in lexicographic practice in other languages, like Spanish, and has been encouraged as a resource for language learning (Renau 2012). The present study proposes to use the CPA language description methodology for a systematic study of translations and, specifically, of audiovisual translation. However, adaptations are needed to pursue this aim.

When conducting corpus-driven research on a specific translation, the researcher does not usually have access to over 250 concordances of each lemma. This hinders a strict application of CPA, which, as stated above,

involves annotation of a relatively large random sample of occurrences of the lemma under analysis to account for its patterns of use.

Therefore, the proposed adaption of CPA involves the following two-step process: a) manually identifying potential creative uses of language in the ST and the TT; and b) checking against general usage corpora and lexicographic tools in order to establish which instances are exploitations. In the present study, the annotation process has been repeated at three different stages of the research in order to foster intra-annotator agreement and minimise human error.

The first step closely follows the first concordance scanning proposed by CPA (see above). In this adaptation, first the transcript and subtitles are read, and potential query terms are identified.

As regards the second step, the corpora and dictionaries employed for the study of the ST in American English have been the *Pattern Dictionary of English Verbs* (PDEV), the *Corpus of Contemporary American English* (COCA), the *Macmillan English Dictionary* (MED), the *Oxford English Dictionary* (OED) and the *Merriam Webster Online Dictionary* (MWD).

For the analysis of the TT in Castilian Spanish, the *Diccionario de la lengua española* (DLE), *Redes: Diccionario combinatorio del español contemporáneo*, *Diccionario combinatorio práctico del español contemporáneo*, Davies' *Corpus del Español* (CDE), and the *Corpus del español del siglo XXI* (CORPES XXI) and have been used to check the exploitation candidates after initial TT scanning. These sources have been complemented with browser queries when needed, although these queries have been kept to a minimum.

To illustrate the annotation process and how the line between conventionalised or true creativity is drawn following this methodology, consider examples (2-4) taken from the ST corpus. The content of these dialogue lines could be understood to fall into one shared conceptual metaphor: LIFE IS A DRAMATIC PERFORMANCE.

(2) DM   ST: My sister **puts up a front** so the world won't see how vulnerable she is. [D02, 01:34:41]

(3) PJ   ST: It doesn't **play** nearly as well. [M01, 00:24:02]

(4) DM   ST: Miami is a great place for me to **play**. [D01, 00:17:43]

In (2), to *put up/on a front* is a lexicalised expression in English —the MED includes it under the following definition of *front*: "behaviour that is not sincere because you want to hide your real feelings". Example (3) may be paraphrased as 'It is not as effective, right?' —which, actually, is the literal subtitling that has been given to this sentence from *The Mentalist* in Spanish. It is difficult to find a dictionary definition of *play* that exactly matches this sense of effectiveness —however, the conceptual metaphor makes it really easy for a native or proficient user of the language to understand the meaning of the sentence, and it can be found in larger corpora, so it would be difficult to classify it as an exploitation: this is a conventionalised metaphor. Finally, in (4), Dexter means that Miami is a great place for him to 'kill people without being noticed or chased by the police.' This sense of *play* is fully understandable in context, like the previous one, but constitutes a clear deviation from the standard sense of *play* —apart from using a linguistic metaphor, Dexter is being sarcastic. Therefore, cases like (4) have been tagged as exploitations in the corpus.

3.3. *Summary of main findings*

As stated above, a thorough discussion of the case study can be found in Arias-Badia (2020). This section provides a synthesis of the main findings of the study to show the kind of results that may be obtained after application of the proposed methodology.

Anomalous collocations are the most frequent type of lexical exploitation in the studied series, closely followed by creative metaphor. They are traceable in the six studied episodes, unlike other types of exploitation which seem to be favoured by specific writers, but are not a constant to the genre of crime. A total of 43 instances have been annotated as anomalous collocations in the ST.

In TV series, anomalous collocations adopt the two major roles of this kind of exploitation as described by Hanks (2013: 217). On the one hand, they are used for referring to "an abnormal situation" when speaking in a normal way. This is the case of Example (5), in which Dexter, a blood spatter analyst who needs specific phrases to describe the blood stains on a wall

("abnormal situation"), comes up with expressions like *nice, clean sprays of blood*.

(5) DM   ST: nice, clean sprays of blood [D01, 00:18:34]
         TT: una lluvia de sangre|limpia y ligera
         BT: a rain of blood|clean and light

On the other hand, anomalous collocates are used in the studied series as a pure "rhetorical device" (Hanks 2013: 217). This use is illustrated in Example (6), where Castle, a best-selling author characterized by his wit and expressivity, is entitled to use supposedly improvised "rhetorical devices" such as *My lifeless remains cannot sue the city*, where the prototypical semantic type functioning as a subject of the verb *sue* —i.e. [[Human]] or [[Institution]] according to the PDEV— is replaced by the semantic type [[Body Part = Lifeless]].

(6) RC   ST: My lifeless remains cannot sue the city. [C02, 00:00:34]
         TT: Mis restos no pueden|demandar a la ciudad.
         BT: My remains cannot | sue the city.

In crime TV shows, the common denominator to these uses, and one which applies to the 43 anomalous collocates found in the corpus, is that they serve as markers of fictional orality, i.e. a device whereby screenwriters aim to artificially provide their texts with greater authenticity and vivacity (Brumme and Espunya 2012). This is expected if we assume that lexical exploitation is part of naturally occurring conversation and, as stated above, screenwriters strive for making television dialogue lifelike. What we find is a more or less elaborate fictional orality in accordance with each character's profile.

The corpus study following the adapted CPA methodology has revealed the occurrence of domain-specific anomalous collocations, that is, non-canonical collocations directly connected to the prototypical domains of specialisation in the genre of crime (law, police procedure, forensic medicine). An example of these anomalous collocates is *close-case doughnuts* [M01, 00:49:31], translated as *donuts de caso resuelto* ('doughnuts of solved case'). Domain-specific collocations amount to more than a half of the anomalous collocations annotated. Some of them are likely to become norms in the

future, if they become a tendency in the genre. *Gruesome souvenir* [D02, 00:05:51], which occurs once in *Dexter*, stands out in this sense —while there is only one occurrence of this collocation in general usage corpora such as the COCA, it is traceable in some crime novels and war memoir books. The collocation is also present in the press, albeit only rarely —consider headings such as "Cat's head may be 'gruesome souvenir'" from *The Border Mail* (October 9, 2013), referring to the case of a woman who found her cat beheaded, or "A very gruesome 'souvenir'" from *News&Record Greenboro* (July 16, 1990), reporting on the murder of a 15-year-old. In *Dexter*, the collocation has been translated as *un truculento regalo* ('gruesome present'), which is indeed anomalous in Spanish.

This example shows how intertextuality plays a crucial role in the interplay of conventionalised and totally novel uses of language. If a particular exploitation can be found repeatedly in different products belonging to the same genre, for example, that paves the way for it to become a norm (i.e. lexicalised expression) in the future. Likewise, the authors' quoting may be unconscious: they may regard their production as novel, when it actually exists already, or may even have become conventional for a reduced group of people. In the same way, audiences may not share the author's world knowledge and not engage in intertextuality. An interesting example from the corpus in this sense is the expression *There she blows* (M01, 00:23:14), which reproduces a quote from *Moby Dick* referring to a whale, but is used in the series to refer to the face painted by a killer with the victim's blood. The use of the pronoun *she* to refer to a painting strikes as surprising, and it has been labelled as an exploitation in the context of the corpus, but intertextuality has been vital in word selection in this case. In Spanish, the sentence has been translated as *Ahí está* ('There it is'), thus not conveying the indirect literary reference nor making use of an exploitation in Spanish.

Following the ideas posited by the specialized literature regarding subtitling, translation solutions in Spanish have favored standardized renderings of the anomalous collocates found in English in the ST (24 vs. 19 instances). Importantly, no instances of creativity have been traced in the subtitles in segments that were not annotated as creative in the ST. Let us present an example of each major type of solution.

In Example (7), Castle exploits the first pattern of the verb *drink* described in the PDEV: [[Human]] *drink* [[Beverage]] by using an object with the semantic type [[Food]] instead of [[Beverage]], to mean that 'he drank a lot of alcohol' during a specific week. The translation preserves the idea of 'drinking a lot of alcohol' by means of a verbal periphrasis expressing continuity over time (*pasar(se) bebiendo*, 'spend drinking'), which is a standard structure in Spanish.

    (7) RC   ST: I drank every meal for a week. [C02, 01:14:07]
          TT: Me pasé toda la semana bebiendo.
          BT: I spent all the week drinking.

In Example (8), the negatively-loaded noun *monster*, which Dexter calls himself because he is a murderer, collocates with the positive adjective *neat*. This is understandable in the context of the series plot because Dexter presents himself as cautious about the remains of his victims. The TT opts for a word-for-word translation of each unit and offers *monstruo pulcro* ('monster neat'). Neither of the options in English or Spanish has matching records in corpora. Instead, in the COCA, *monster* collocates with epithets dealing with physical appearance, such as *green-eyed*, *big* or *two-headed*, as well as with negative adjectives such as *scary* or *evil*. *Neat*, by contrast, collocates with substantives related to the organization or distribution of physical objects, such as *piles*, *lines*, or *rows*. The same kind of collocates are found in Spanish reference corpora.

    (8) DM   ST: I am a very neat monster. [D01, 00:12:08]
           TT: Soy un monstruo muy pulcro.
           BT: I am a monster very neat.

The present case study has demonstrated the presence of anomalous collocates in television dialogue. It has also shown how American crime TV series make use of salient, domain-specific anomalous collocates which bear intertextuality with literary products of the crime genre, or the report of crime cases in the press. Although the total occurrence of anomalous collocates is relatively low (43 instances in the whole corpus), the fact that these elements convey information about the characters' personality which should

not be overlooked poses a challenge for audiovisual translators. Therefore, professional subtitlers and trainees could benefit from studies raising awareness of such lexical exploitations present in audiovisual material. TNE and CPA have proved their usefulness as both a theorical and methodological framework for such studies.

A tendency towards normal use has been observed in the subtitling of these creative units. The comparison of these findings with the existing literature on the translation of other instances of creative language, such as metaphor, may lead to the conclusion that using normal, non-exploited language is a preferred solution by professionals when dealing with lexical exploitation. Future studies on the treatment of other kinds of exploitations could result in the identification of a translational norm in translation practice (Toury 1995).

4. Methodological review

This section is devoted to reflecting on the shortcomings of the proposed methodology, as well as on means to tackle them —when possible— and on the advantages entailed in adopting this approach for the study of audiovisual translation. Each disadvantage or advantage is expressed in a sentence and further specified below.

4.1. Limitations

a) CPA and its adaptation entail introspection on the part of the researcher.

Studies dealing with style tend to rely on a first manual, intuitive approach to the text in hand. Leech and Short (1981: 4) pose the question in the following terms:

> [S]tylistics, as the study of the relation between linguistic form and literary function, cannot be reduced to mechanical objectivity. In both the literary and the linguistic spheres much rests on the intuition and personal judgement of the reader, for which a system, however good, is an aid rather than a substitute. There will always remain, as Dylan Thomas says, "the mystery of having been moved by words".

Indeed, it is frequent for translation scholars to undertake corpus studies individually, and the manual annotation proposed for this study does not escape introspection and an inherent bias. As specified by Hanks (2012: 54), however, "introspection of this kind can be confirmed and extended by examination of patterns in corpora". Therefore, corpora and dictionaries are proposed as the "aid", following the term used in the citation above, to minimise the impact of individual bias. Inter-annotator agreement tests undertaken with other researchers in the field are also encouraged when possible. If not possible, as in the case study described, intra-annotator agreement tests are deemed necessary, again, to obtain robust results.

b) The availability of corpora and lexicographic resources varies greatly across different languages.

This study focuses on the English-Spanish language pair, for which extensive resources are available to be used as elements of contrast. Indeed, it would be difficult to conduct this type of research on underrepresented language pairs, since the contrast of personal intuitions with lexicographic sources of information is mandatory to obtain relevant results. Although they are not desirable because they typically produce noise, browser queries could be useful to tackle this limitation.

c) Multimodality must not be neglected in the analysis of audiovisual translations.

When annotating general usage corpora, CPA lexicographers have access to text segments which allow them to assign a specific pattern to each occurrence of the word under analysis. In the case of audiovisual translation, both television dialogue and subtitling take place within the broader audiovisual text. To escape a lexical analysis deprived of contextual information, thus, the audiovisual translation researcher must take the image into consideration in the annotation process. Indeed, "corpus investigations focusing exclusively on the verbal component are at risk of overlooking the importance of the other semiotic codes to the meaning making process in audiovisual products" (Díaz-Cintas 2008: 3). To date, however, audiovisual translation research has approached multimodality in a rather qualitative manner,

indeed constrained by the individual bias of researchers. This is surely the case in the study reported. In this sense, it is hoped that recent research proposing systematic ways to address multimodality in other audiovisual translation modalities, such as Reviers' (2018) account of audio description by means of the integration of a multimodal concordancing system, will enrich the methodology hereby proposed in the near future.

d) Manual annotation is a difficult task: there is a vast "grey area" to consider in the study of the lexicon.

In the corpus annotation process, it is often difficult to draw the line between conventionalised, seemingly creative expressions, and truly creative exploitations. This is because there is no such dividing line, only a "fuzzy grey area" (Hanks 2010, 2013), as noted by lexicographers who have previously faced the CPA annotating task (Renau 2012: 141). In this sense, the above mentioned "aids" become vital for the researcher to provide solid arguments for annotation. However, the difficulty entailed in the manual annotation methodology *per se* must be considered a clear shortcoming of the proposal. The accuracy rates of automatic semantic taggers reported to date, however, are low (Rees 2018: 211) for this kind of task. So, for now there is no potential alternative to manual annotation.

e) Results of this kind of study become obsolete over time: the lexicon changes continuously.

The fact that the lexicon is in constant change is undeniable. Therefore, researchers adopting the approach proposed here must bear in mind that the results obtained will only describe what was (or was not) creative at the time of the research. Corpora and dictionaries reporting on the language used in different time periods may be employed to address the language of a specific television product. In fact, such lexicographic resources are also known for needing constant updating, so the arguments provided for analysis at some point in time may vary if we repeat the annotation process with the same texts after updates in the lexicographic resources have been implemented or general usage corpora have been extended.

It is encouraged that studies using CPA for the study of translation reflect on likely future changes in the results gathered, as done in the case study reported with the phrase *gruesome souvenir*, which can now be labelled as unusual or anomalous, but may become a norm of the genre in the future, since it is already traceable in a number of crime fiction-related products.

4.2. Contributions

a) Adapting a methodology that considers "normal" use of language is a step towards cohesive interdisciplinary research.

As posited above, the identification of "normal" patterns is a focus of interest common to Television Studies, Linguistics, and Translation Studies. Accordingly, visual marks or structural invariants (i.e., invariable patterns) are key to the identification of audiovisual genres and types of product from the standpoint of Television Studies; these patterns provide the "frame of reference" in order for audiences to understand the shows they are exposed to, and for scholars to study them (Richardson 2010: 84). Great efforts are devoted in linguistic research to studying normal patterns of language use. TNE is a clear example of the interest raised by norms in the discipline. In a similar vein, within Descriptive Translation Studies, Toury (1995) postulates the Theory of Norms in Translation, a research framework the object of which is to trace recurrent patterns in translational behaviour. Thus, by establishing a link in the core normative concept in these disciplines, the proposed methodology strives to take the first step towards cohesive research.

b) Adopting a methodology that has proved useful in another field is convenient for Translation Studies.

Just like Television Studies rely on Literary Theory to account for genre norms, i.e. the younger discipline resorts to the older one, it seems natural to suggest that replicating a solidly grounded theory within lexical analysis for the study of translations is a sound practice for the discipline. CPA is fully in accordance with Descriptive Translation Studies principles and provides a systematic approach to the study of translation creativity based on data beyond the researcher's intuition.

c) CPA allows the researcher to escape the hegemony of the source text in their studies.

The independent annotation (and analysis) of the whole source and target texts allows researchers to decide where to place emphasis when conducting their study. Although this has not been the case in the study reported above, the results of this process may reveal creativity in parts of the target text that are not creative in the source text, thus challenging the assumption that translations are usually less creative. CPA allows a true bottom-up, corpus-driven approach to translations as creative texts, deserving as much attention as source texts, and deals with them from a *tabula rasa* perspective.

d) CPA allows a statistical analysis of the results.

As noted by Corpas (2008: 53), corpus-driven studies allow statistical analyses. Hanks (2013: 415) explains that "being able to make predictions is much more useful than speculating about the boundaries of possibility." Only studies undertaken with proved systematic, replicable annotation may lead to the eventual identification of rules, norms and tendencies in Translation Studies. According to Toury (1995), such is the aim and scope of researchers in Descriptive Translation Studies. As put by Laviosa (2002: 79):

> This type of analysis is performed not to evaluate the quality of a given translation, but to understand the decision-making process underlying the product of translation and to infer from it the translational norms adopted by the translator.

Of course, it is difficult to draw the line between tendencies and norms, which typically depend on corpora representativeness (Tognini-Bonelli 2001; Martínez Sierra 2011; Corpas & Seghiri 2016). This aspect, i.e. the degree to which the research intends to extrapolate the results obtained, must always be considered in the light of the specific research questions raised in a study adopting the methodology proposed here.

5. Closing remarks

This paper provides a critical evaluation of a methodological proposal made in my own previous work. It has reported on the methodology used for a

case study aiming to describe the occurrence of anomalous collocations as a creative device in television dialogue and subtitling. Such a methodology entails the adaptation of Corpus Pattern Analysis, which had been hitherto used to study lexicon in corpora for lexicographic and language learning purposes.

The main goal of the paper has been to provide an overview of the limitations and contributions of the proposed methodology for the study of audiovisual translation. Toury (1995: 69) aptly noted that

> […] achievements of actual studies can themselves supply us with clues as to necessary and possible methodological improvements. Besides, if we hold up research until the most systematic methods have been found, we might never get any research done.

Indeed, the proposed adaptation of CPA is lacking in some respects, among which I would highlight the need to further account for multimodality in the case of audiovisual translation. It is, however, a sound methodology for the study of lexicon in translation in the sense that it is based on a well-grounded methodology employed in neighbouring research fields. It has already brought interesting results for language pairs which allow a systematic checking of creative uses of language in extensive corpora and dictionaries.

References

ARIAS-BADIA, Blanca. (2015) "Towards a Methodology for the Analysis of Neutralisation in Spanish Subtitling." In: Corpas, Gloria; Seghiri, Miriam; Gutiérrez, Rut; Urbano, Miriam (eds). *New Horizons in Translation and Interpreting Studies. Proceedings of the 7th AIETI Conference*. Geneva: Tradulex, pp. 513-526.

ARIAS-BADIA, Blanca. (2020) *Subtitling Television Series: A Corpus-Driven Study of Police Procedurals*. Oxford: Peter Lang.

BARTOLL, Eduard. (2012) *La subtitulació: aspectes teòrics i pràctics*. Vic: Eumo.

BRUMME, Jenny & Anna Espunya (eds.) (2012) *The Translation of Fictive Dialogue*. Amsterdam/New York: Rodopi.

CHAUME, Frederic. (2004) "Discourse markers in audiovisual translating". *Meta* 49:4, pp. 843-855.

CORPAS, Gloria. (2008) *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Frankfurt am Main: Peter Lang.

CORPAS, Gloria & S Miriam Seghiri. (2016) *Corpus-based Approaches to Translation and Interpreting.* Frankfurt am Main: Peter Lang.

DÍAZ-CINTAS, Jorge. (2003) *Teoría y práctica de la subtitulación inglés-español.* Barcelona: Ariel.

DÍAZ-CINTAS, Jorge. (2008) "Introduction." In: Díaz-Cintas, Jorge (ed.) *The Didactics of Audiovisual Translation.* Amsterdam/Philadelphia: John Benjamins, pp. 1-18.

DÍAZ-CINTAS, Jorge & Aline Remael. (2007) *Audiovisual translation: Subtitling.* Manchester: St. Jerome Pub.

FEUER, Jane. (1992) "Genre study and television." In: Allen, Robert Clyde (ed.) *Channels of discourse, reassembled.* Chapel Hill: The University of North Carolina Press, pp. 138-160.

FISKE, John. (1987) *Television Culture.* London: Routledge.

GOTTLIEB, Henrik. (1998) "Subtitling." In: Baker, Mona & Kirsten Malmkjær (eds.) *Routledge encyclopedia of translation studies.* London/New York: Routledge, pp. 244-249.

HANKS, Patrick. (n.d.) "Corpus Pattern Analysis: How people use words to make meanings". Electronic version: <http://www.patrickhanks.com/powerpoint.html>.

HANKS, Patrick. (2004) "Corpus Pattern Analysis." In: Williams, Geoffrey & Sandra Vessier (eds.) *Proceedings of the 11th Euralex International Congress*, Lorient: Euralex, pp. 87-97.

HANKS, Patrick. (2012) "How people use words to make meanings: Semantic types meet valencies." In: Thomas, James & Alex Boulton (eds.) *Input, Process and Product: Developments in Teaching and Language Corpora.* Masaryk: Masaryk University Press, pp. 52-67.

HANKS, Patrick. (2013) *Lexical Analysis: Norms and Exploitations.* London: The MIT Press.

KILGARRIFF, Adam; Pavel Rychly; Pavel Smrz & David Tugwell. (2004). "The Sketch Engine." In: Williams, Geoffrey & Sandra Vessier (eds.) *Proceedings of the 11th Euralex International Congress*, Lorient: Euralex, pp. 105-116.

LAVIOSA, Sara. (2002) *Corpus-based Translation Studies: Theory, Findings, Applications.* Amsterdam: Rodopi.

LEECH, Geoffrey. (1990/1974). *Semantics: The Study of Meaning.* London: Penguin.

LEECH, Geoffrey & Mick Short. (1981) *Style in fiction: A linguistic introduction to English fictional prose.* London: Longman.

LEWIS, Michael. (1993) *The Lexical Approach*. Hove: Language Teaching Publications.

MATAMALA, Anna. (2019) *Accessibilitat i traducció audiovisual*. Vic: Eumo.

MARTÍNEZ SIERRA, Juan José. (2011) "De normas, tendencias y otras regularidades en traducción audiovisual." *Estudios de Traducción* 1, pp. 151-170.

NEALE, Steve & Graeme Turner. (2001) "Introduction: What is genre?" In: Creeber, Glen (ed.) *The television genre book*. London: British Film Institute, pp. 1-7.

OLOHAN, Maeve. (2004) *Introducing corpora in translation studies*. New York: Routledge.

REES, Geraint Paul. (2018) *A Phraseological Multi-Discipline Approach to Vocabulary Selection for English for Academic Purposes*. Universitat Pompeu Fabra (Barcelona). Doctoral dissertation.

RENAU, Irene. (2012) *Gramática y diccionario: Las construcciones con se en las entradas verbales del diccionario de español como lengua extranjera*. Universitat Pompeu Fabra (Barcelona). Doctoral dissertation.

REVIERS, Nina. (2018) *Audio Description in Dutch. A corpus-based study into the linguistic features of a new, multimodal text type*. University of Antwerp (Antwerp). Doctoral dissertation.

RICHARDSON, Kay. (2010) *Television Dramatic Dialogue: A Sociolinguistic Study*. Oxford: Oxford University Press.

SINCLAIR, John. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

SINCLAIR, John. (2003) *Reading Concordances*. London: Longman.

TEIXEIRA, Isadora. (2015) *Subtitling of collocational patterns in children's animated movies: A corpus-based study*. Universidade Federal de Santa Catarina (Florianópolis). MA thesis.

TOGNINI-BONELLI, Elena. (2001) *Corpus Linguistics at Work*. Amsterdam/ Philadelphia: John Benjamins.

TORNER, Sergi & Blanca Arias-Badia. (in press) "Los diccionarios de colocaciones." In: Battaner, Paz; Renau, Irene; Torner, Sergi (eds.) *Routledge handbook of Spanish lexicography*. London: Routledge.

TOURY, Gideon. (1995) *Descriptive Translation Studies and Beyond*. Amsterdam/ Philadelphia: John Benjamins, pp. 53-69.

Walsh, Steve, Morton, Tom & Anne O'Keeffe. (2011) "Analysing university spoken interaction. A CL/CA approach." *International Journal of Corpus Linguistics* 16:3, pp. 325-344.

Wolf, Mauro. (1984) "Géneros y televisión." *Anàlisi. Quaderns de Comunicació i Cultura* 9, pp. 189-198.

Wolff, Jurgen & Kerry Cox. (1988) *Successful scriptwriting.* Cincinnati: Writer's Digest Books.

Zaro, Juan Jesús. (2001) "Conceptos traductológicos para el análisis del doblaje y la subtitulación." In: Duro, Miguel (ed.) *La traducción para el doblaje y la subtitulación.* Madrid: Cátedra, pp. 47-64.

NOTA BIOGRÁFICA / BIONOTE

Blanca Arias-Badia is a tenure-track lecturer at the Universitat Pompeu Fabra in Barcelona, where she teaches translation for general purposes and audiovisual translation. She holds a PhD in Translation and Language Sciences (UPF) and did a postdoc in Audiovisual Translation and Media Accessibility at the Universitat Autònoma de Barcelona, awarded by the Spanish Research Agency under the Juan de la Cierva postdoctoral scheme. She is a member of the research group InfoLex (UPF) and the research coordinator of the Catalan Association for the Promotion of Accessibility (ACPA). She is the author of *Subtitling Television Series*: *A Corpus-Driven Study of Police Procedurals* (Peter Lang, 2020).

Blanca Arias-Badia es profesora en vías de permanencia en la Universitat Pompeu Fabra de Barcelona, donde imparte asignaturas sobre traducción general y traducción audiovisual. Es doctora en Traducción y Ciencias del Lenguaje (UPF) y realizó un posdoc Juan de la Cierva centrado en la traducción audiovisual y la accesibilidad en la Universitat Autònoma de Barcelona, financiado por la Agencia Estatal de Investigación. Es miembro del grupo InfoLex (UPF) y coordinadora de investigación de la Asociación Catalana para la Promoción de la Accesibilidad (ACPA). Es autora del libro *Subtitling Television Series: A Corpus-Driven Study of Police Procedurals* (Peter Lang, 2020).