

Detection approaches to plagiarism in literary translation

M. Teresa Turell
Universitat Pompeu Fabra
teresa.turell@upf.edu

1. Introduction

A twofold difficulty is involved in the protection of translated literary texts under Intellectual Property Acts. On the one hand, the fact that all translations will tend to reflect the form and content of the original piece of work; on the other hand, the circumstance that the more faithful to the original text these translations are, the more difficult it will be to establish their originality. The translation of a literary work is considered to be another literary work, and literary works share several traits with didactic and scientific works. In fact, these three types of work are considered under the same legal and juridical system, both in the Spanish Intellectual Property Act (*Ley de Propiedad Intelectual* (LPI) - 1996) and in comparative law. For the purposes of this article, however, the important point to be considered is that the Spanish LPI does not protect the data, but rather their original selection and grading (Article 12).

In the case of translations of a literary work it is possible to grant intellectual property to that part of the translation that presupposes an original contribution by the translator. According to Rodríguez Tapia (1995), once the translation's originality has been granted, this originality falls within the scope of protection established by Articles 10 and 11 of the Spanish LPI, although this scope is more reduced since neither its title, nor the original plot, characters, and proper names (patronymics and toponymics) are protected. As is reasonable, these constituents are liable to the original author's intellectual property rights. In Rodríguez Tapia's own words, 'the translator's property will be on the translation, its structure, the syntax, and

several common nouns that he/she might have selected as alternative terms to previous translations' (Rodríguez Tapia 1995: 774).

Given the complex nature of translation, the approaches to detect plagiarism of translated texts - in particular literary texts - must necessarily be varied. Both qualitative and quantitative approaches are used to detect plagiarism of translated literary texts. In fact, it is argued that these types of methods may have to complement each other and that the degree of complementation may be related to the nature of the specific cases of plagiarism between translations under analysis.

This article makes a comparative evaluation of the different approaches used in the detection of the kind of plagiarism which occurs between different translations of the same literary work, by considering a) the difference/similarity of the linguistic units and categories found, b) the ease/difficulty of the approaches themselves, and c) the advantages/ disadvantages of taking one approach or another.

2. The plagiarised text as a forensic document

In a plagiarism context, a text becomes a forensic document when either the copying of ideas¹ or/and linguistic plagiarism can be detected.

2.1. The copying of ideas

The copying of ideas occurs when the content or ideas, expressed in some literary (original or translated), artistic or scientific text, or any other type of text, are used in another piece of work as if original. It is obviously much more difficult for linguists to establish when the

¹ Although the term "plagiarism" usually accounts for both the borrowing of somebody's ideas and/or the same message employed to express them, for the purposes of this article I will use the terms "copying of ideas" and "linguistic plagiarism", respectively.

copying of ideas takes place than when linguistic plagiarism occurs, but it must be remembered that ideas are expressed through language, and this means that if there is linguistic plagiarism, then the copying of ideas also occurs, although copying of ideas without linguistic plagiarism is also possible.

The copying of ideas may take place under several circumstances:

a. In literary works, when exactly the same structural elements (plot, characters, place, time, stream of consciousness, and others) which attribute unity to a specific literary piece of work are borrowed and used in another piece of work as if original.

In Spain, for example, *Fermoso v. Cela* is a clear-cut case of copying of ideas, although the case didn't get very further in court. Both Fermoso's *Carmen*, *Carmiña*, *Carmela* (CCC) –submitted to the Planeta Prize in due time–, and *La Cruz de San Andrés* (LCSA) –submitted to the same prize after the deadline was due– narrate the story and relationship of three women from the same family. As to time, in CCC the action is developed between 1931 and 1994, although the core of the plot takes place during the sixties and seventies, the same as in LCSA. Furthermore, both in CCC and LCSA the action develops in the city of A Coruña, but many other places are shared by both novels: La Habana, Buenos Aires, Morocco and also several specific places in A Coruña are mentioned in both works: Torre de Hércules, Instituto Eusebio da Guarda (the secondary school where Carme Fermoso studied), El Carballo de San Pedro de Nos, Ordés, Betanzos, and also one street in Madrid: Fuencarral. The three main characters in CCC, Carmen, Carmela and Carmiña, find their correlates in LCSA: Matilde Verdú and her daughters Matty y Betty Boop, but not only do these two novels share the number and gender of their protagonists, they also share similitude in physical appearance, feelings, anecdotes and experiences. Finally, the novels also share narrative topics such as: Magic, Powers, Sorcery, Rituals, Solitude, Aging, Death.

b. In scientific contexts, when the same topics are used in the description of a historical period or in a contribution to a field of specialisation as if original.

The copying of ideas in the production of scientific works is found in the use of the same topics in the description of a historical period or in a contribution to a field of specialisation. *B. v. RBA*² provides an illustration of this type of copying of ideas. An analysis of the disputed article, written for NG in 2000, showed that P. partially used the same source as B.' PhD dissertation, written in 1991.

c. In scientific text books, when a creative methodology devised to teach a particular discipline is reproduced in another text book.

Thus, the copying of ideas may also have to do with the reproduction of a creative methodology devised to teach a particular discipline. *G. v. Editorial Vicens Vives* illustrates an example of a "supposed" copying of ideas in which the content of what was plagiarised has to do with methodology, in this case, the methodology used in teaching mathematics to secondary students.

2.2. Linguistic plagiarism

Following Menasche (1977) and Roig (2006),³ linguistic plagiarism takes place when the following circumstances occur:

- "When exactly the same words and/or sentences are used in order to write about one's own or other people's ideas.
- When there exists paraphrase, that is, when someone uses other people's ideas with his or her own words but makes use of the main bulk of the original words, phrases and sentences.

² RBA is the publisher in Spain of the Spanish version of the journal *National Geographic*.

³ See also <http://www.sscnet.ucla.edu/history/bartchy/classes/194a/98F/plagiarism.html> and <http://facpub.stjohns.edu/~roigm/plagiarism/Index.html>, respectively.

- When one uses several words and sentences without quotations but changes others.
- When the original syntax is maintained and only words are replaced by synonyms.
- When there is acknowledgment of the original author, but the changes only involve one or two words, word order (WO), voice (active v. passive) and/or the verbal tense and aspect of the sentences or the whole text" (Turell 2008: 281).⁴

As mentioned above, and due to the intrinsic nature of the linguistic sign, if linguistic plagiarism occurs, the copying of ideas also takes place. Thus, the notion of linguistic plagiarism is usually extended to a) self-plagiarism, that is, when a writer reuses his/her own material or data used in a previous article, without letting the audience know that this text material has been published in another piece of work, and b) the segmentation or fragmentation of data and research results in one or two publications, with the additional implied problem of distortion.

According to Coulthard, when forensic linguists are "called in to help a court" (2005: 249) decide whether or not plagiarism has occurred, the detection of linguistic plagiarism usually implies trying to answer the following question: *Who is the author of a particular text?*⁵ Taking the notion of *idiolectal style* (Turell 2010), according to which the phrases and sentences occurring in an individual's language production and use are unique, the possibility that two writers or two translators produce the same phrases and sentences by chance is low; thus, when a substantial proportion of these linguistic units in two texts under comparison are the same, one should suspect that either those texts have been produced by the same author, or that plagiarism has occurred between one text and the other and they both share a common source.

⁴ For a complete overview of plagiarism in education and the directions given to avoid plagiarism, see Roig (2006); for an interesting analysis of plagiarism in the Internet, see DeVoss and Rosati (2002).

⁵ The other question usually posed is "*What does a given text 'say'?*" (Coulthard 2005: 249).

3. The nature of plagiarism in translated literary texts

Literary works, in particular poetic texts, involve more textual difficulty than other types of texts, that is, more complex lines, more connotations and more stylistic resources are being used by original writers. It is also an indisputable fact that in a translation context this textual difficulty should entail a less favourable context for coincidence in the recreation of the decoded source text into the target language. In other words, the difficulty in interpreting poetic, sometimes cryptic, passages because of the wide range of linguistic, stylistic and rhetorical tokens increases the final disparity in translating in similar or identical terms what both translators understood in the original piece of work.

The mere fact of wishing to do a new translation presupposes that the translator is willing to contribute something new, that is, something that does not exist in the other translations of the same original literary work. Therefore, it is presupposed that a new translation of a literary work should be performed independently from previous translations. However, the fact is that many existing translations of literary works exhibit a suspicious degree of linguistic coincidence beyond the accepted premise that there will be a common substrate in a new translation and the previous ones, and beyond justifications such as intertextuality and *imitatio*.

The two contexts that are described in this section⁶ illustrate the multi-dimensionality of plagiarism between translations of the same literary text, where both the copying of ideas and linguistic plagiarism concur.

d. The use by one translator of all or almost all original rhetorical figures from the author of a previous translation of the same original literary work.

In *Julius Caesar*, for example, Shakespeare makes use of six puns, which are brilliantly translated into Spanish by Pujante (1987) and

⁶ Others contexts will be highlighted throughout the qualitative analysis that will be considered in the following section.

which Vázquez Montalbán (1988)⁷ copies without making a single change.

e. The copying of a translated version, if the translated version itself makes an explicit contribution, by changing this version from prose to verse, by dehistoricising a classical work, or by historicising a contemporary work.

If a future translation of Shakespeare's *Julius Caesar* made use of Vázquez Montalbán's dehistoricised version, to make it timeless, the author of this new translation could be accused of plagiarism because a translation's version is something protected by law. According to Rodríguez Tapia (1995: 774), the Spanish LIP protects the translation, that is, its version, its structure, its grammar and alternative vocabulary, *vis-à-vis* previous translations referred to common entities. Vázquez Montalbán achieved this dehistoricisation by erasing all references to Rome in such a way that *Rome* becomes *city*; *Italy* becomes *this country*; *Tiber* becomes *the river*; the *Romans* is replaced by *citizens*, and finally, *Spain* becomes *campaign*.

4. Qualitative approaches to plagiarism in literary translation

Qualitative approaches to plagiarism between translations of literary works involve the observation of a) the literary text as a whole, searching a global characterisation of the translations under analysis, b) the discrete components and slots where linguistic plagiarism is bound to occur, that is, the specific units which are analyzed according to syntactic, semantic-pragmatic and rhetorical criteria and c) the

⁷ From now on, this case will be referred to as *Pujante v. Vázquez Montalbán*. The corpus used in the analysis of this case has involved the consideration of the non-disputed text (Pujante 1987 (P.)), the disputed text (Vázquez Montalbán 1988 (VM.)) and two other translations of *Julius Caesar* by Astrana Marín (1962 (A.)) and Valverde (1968 (V.)).

plagiarism which is produced by copying the most original translation strategies used by translators and the language used to express them.

4.1. A global characterisation of the translated texts

The approach to the literary text as a whole informs us of the nature of the discrete units and slots where plagiarism occurs and facilitates a further search for the specific strategies used by the original translator and the plagiarist. In *Pujante v. Vázquez Montalbán* it was useful to have an overall look at the four existing translated texts of *Julius Caesar*. This preliminary observation facilitated a global characterisation of the texts under analysis and of the version used in each of them. Following Turell (2004): “Astrana’s is a more rhetorical and flowery translation”; “Valverde’s is the most literal of all, a word-for-word translation”; Pujante’s non-disputed text consists of “a praiseworthy recreation of the original text, achieved by means of his sensitivity to literary discourse” and to “the stylistic and figurative resources used by Shakespeare”, and finally, Vázquez Montalbán’s disputed translation is mostly a faithful paraphrase of Pujante’s version, although as mentioned above Vázquez Montalbán de-historicises the text, whose version would be protected from future plagiarism practice.

4.2. Plagiarized units and slots

4.2.1. Single and compound lexical structures

In *Pujante v. Vázquez Montalbán*, the prosecution expert’s report (Zozaya 1990, cited in Turell 2004) involved the analysis of single verses in the four translated texts under comparison and the proposal of several lexico-syntactic taxonomies (*one-word calques, multi-word calques, whole-verse calques, calques of a series of verses*). Examples (1) to (4)⁸ illustrate the type of units identified by the prosecution expert in her qualitative analysis of the four translations considered.

⁸ Where A. stands for Astrana Marín, V. stands for Valverde, P. stands for Pujante, and VM. Stands for Vázquez Montalbán.

One-word calques

(1) GOOD MORROW, BRUTUS (II, i, 87)

A. Buenos días, Bruto.

V. Buenos días, Bruto.

P. Buenas noches, Bruto.

VM. Buenas noches, Bruto.

In (1) Pujante translates ‘morrow’ as *noche* and Vázquez Montalbán replicates the same translation. The expert reports at least 11 instances of this type of matching where the semantic content of a key word adjusts the line to the desired effect of the whole phrase or sentence.

Multi-word calques

(2) INDEED IT IS A STRANGE-DISPOSED TIME (I, iii, 33)

A. Es ésta una época bastante extraña.

V. Desde luego, es un tiempo de extraña disposición.

P. En efecto. Vivimos unos tiempos singulares.

VM. Bien es cierto. Vivimos unos tiempos singulares.

The expert’s case also reports a frequent use of multi-word calques, including whole Sentences, Noun Phrases, Verb Phrases, Adjective Phrases and Adverbial Phrases.

Whole-verse calques

(3) YOU PULLED ME BY THE CLOAK (I, ii, 215)

A. Me habéis tirado del manto.

V. Me habéis tirado del manto.

P. Me has tirado de la toga.

VM. Me has tirado de la toga.

In this example, Pujante detaches himself from the more literal solutions by Astrana Marín and Valverde and introduces two new elements – the second person singular form of the verb: *Me has tirado*, instead of *Me habéis tirado* used by the first two translators, and the more appropriate, although less exact, translation of ‘cloak’ by *toga* – and Vázquez Montalbán copies the exact words and morphological options.

Calques of a series of verses

(4) AND GROANING UNDERNEATH THIS AGE'S YOKE (I, ii, 61–2)

A. Y gimiendo por la opresión de la época,

V. gimiendo bajo el yugo de esta época

P. que gimen bajo el yugo de los tiempos

VM. que gimen bajo el yugo de los tiempos

HAVE WISHED THAT NOBLE BRUTUS HAS HIS EYES.

A. suspirar por que el noble Bruto abriese los ojos.

B. han deseado que el noble Bruto tuviera sus ojos.

P. y quisieran que Bruto no fuese tan ciego.

VM. y quisieran que Bruto no fuera tan ciego.

In relation to this example, the expert's observation is that it is precisely Pujante's ability to recreate the original text that makes Vázquez Montalbán's matching so much more obvious.

Another instance of lexical calquing appears in the qualitative analysis of supposed plagiarism between the existing translations into Catalan of Kafka's *Metamorphosis*. Llovet's (Ll.), published in 1985, Sòria Parra and van Lawick's (S/L), which appeared in 1989 and Gala's (G) translation, published in 2004, a case which has not reached the court either. Examples 5 to 7 illustrate this type of calquing.

(5) UNRUHIGEN TRÄUME

Ll. somnis **neguitosos**

S/L. somnis **agitats**

G. somnis **agitats**

(6) BOGENFÖRMIGEN VERSTEIFUNGEN

Ll. **estreps** arquejats

S/L. **durícies** arquejades

G. **durícies** arquejades

(7) MIT EINEM PELZHUT UND EINER PELZBOA VERSEHEN

Ll. **guarnida** amb un barret i un boà de pells

S/L. **abillada** amb un barret de pells i un boà de pells

G. **guarnida** amb barret i boà de pells

4.2.2. Syntactic, semantic and pragmatic features

Many plagiarised units involve the combination of syntactic, semantic and pragmatic features which have to do with the diverse use of the informational packaging options (Vallduví 1992) that different languages offer such as end-focus, left-dislocation, right dislocation, grammatical metaphor (Downing 1989, 1992) and also of the different translators' preferences for the choices available in one same language.

In *Pujante v. Vázquez Montalbán*, the expert highlights an example of a syntactic translation change divergent from Shakespeare's original text. In (8) it can be observed that Pujante inverts in Spanish ((**vosotros**) No volveríais a mirarme a la cara) the original Subject-Object functions in English (I'll never look you in th' face again) and Vázquez Montalbán copies him in making use of this syntactic inversion.

- (8) I'LL NEVER LOOK YOU I' TH' FACE AGAIN
A. Pues que no os mire más a la cara
V. no os volvería a mirar más a la cara
P. no volveríais a mirarme a la cara
VM. no volveríais a mirarme a la cara

Finally, by making use of total recreation, as in (9), Pujante makes substantial use of Spanish syntactic, morphological and semantic structures that are very divergent from those used by Shakespeare in the original text. *Pujante v. Vázquez Montalbán*'s case illustrates a pragmatic deviation by Pujante from Shakespeare's use of a grammatical metaphor (in shame of cowardice), which is literally reproduced by Astrana Marín and Valverde; Pujante, on the other hand, prefers the congruent syntactic option in terms of using the structure Verb + Object (censuran al cobarde), which is partially copied by Vázquez Montalbán, who in turn substitutes "censuran" (to censor) by "castigan" (to punish).

- (9) THE GODS DO THIS IN SHAME OF COWARDICE (II, ii, 41)
A. ¡Eso lo hacen los dioses para vergüenza de la cobardía!
V. Los dioses hacen eso para avergonzar la cobardía.
P. Así es como los dioses censuran al cobarde.
VM. Así es como los dioses castigan al cobarde.

4.2.3. *The rhetorical nature of plagiarised translated literary texts*

Literary texts, in particular poetic ones, are linguistically constructed through the use of hyperboles, similes, metaphors, images, rhyme, syncope, ellipsis, and other figures of speech, as well as puns and famous quotations. All of these constitute very useful linguistic markers to evaluate a good translator's quality, creativity and, above all, originality.

The use of puns in Elizabethan theatre, and the difficulty involved when translating them, has been extensively considered (Delabastita 1997, von Flotow 1997). In *Julius Caesar*, Shakespeare makes use of several puns, some of which are brilliantly translated into Spanish by Pujante while Vázquez Montalbán copies Pujante's translations without making a single change. For instance, example (10) reproduces one pun, which appears at the beginning of *Julius Cesar*, during the conversation between Flavius and the cobbler, and constitutes one of the most difficult textual passages for the translator in the whole work.

(10) A TRADE, SIR, THAT I HOPE I MAY USE WITH A SAFE
CONSCIENCE; (I, i, 13-14)

A. Un oficio, señor, que espero podré ejercer con la
conciencia tranquila,

V. De un oficio, señor, que espero se me agradezca con
buen consuelo:

P. Señor, un oficio que siempre hace el bien:

VM. Señor, un oficio que siempre hace el bien:

WHICH IS INDEED, SIR, A MENDER OF BAD SOLES

A. pues, en verdad, es el de reparador de malas suelas.

V. con malas suelas trabajo, señor.

P. a quien mal anda, lo con-suela.

VM. a quien mal anda, lo con-suela.

The cobbler begins by playing with the meaning of 'soles' used as the under-surface of a shoe and as the non-material part of a human body, a wordplay that cannot be achieved in Spanish; Pujante solves the wordplay introduced by Shakespeare in 'soles' by playing with the Spanish prepositional phrase 'con-suela', which means 'with soles', and the Spanish verb 'consuela' (consolar) which involves to 'give

comfort'. Vázquez Montalbán calques this translation solution without even changing a comma.

However not all puns are translated as such in order to keep the original author's ultimate objective or produce the original author's effect on the audience. For example, none of the translators of Shakespeare's *Julius Caesar* into Spanish chooses to translate as such the series of sexual puns found in this tragedy. This series begins in Scene II.i when the conspirators' main concern is "Whether Caesar will come ... to-day", a double meaning of 'come' repeated six times in the first fifteen lines of interaction between Decius and Caesar. This is followed in Scene II.ii by Caesar's qualification of his sexual refusal with a further sequence of sexual puns in which he explains that the cause of his refusal is in his "will" (penis) and that this reason should "satisfy" (sexually gratify) the Senate. Example (11) illustrates one fragment of these series of untranslated sexual puns in the four Spanish translations of *Julius Caesar*.

- (11) CAESAR: THE CAUSE IS IN MY WILL: I WILL NOT COME;
THAT IS ENOUGH TO SATISFY THE SENATE.
- A. ¡La causa es mi voluntad! ¡Que no iré!
Esto es bastante para satisfacer al Senado;
- V. La causa está en mi voluntad: no quiero ir,
y eso es bastante para satisfacer al Senado.
- P. **La razón está en mi voluntad. No pienso ir.
Que baste para satisfacer al Senado.**
- VM. **La razón está en mi voluntad. No quiero ir.
Y basta. El Senado se dará por satisfecho.**

4.3. Translation strategies

Qualitative approaches to plagiarism of translated literary texts allow the researcher to observe the nature of those translation strategies which are more bound to be plagiarised. Prior to this analysis, it is necessary to make a distinction between a) expected translation solutions due to the clarity, straightforwardness and preciseness of the original text and which cause parallelism between translated texts and b) flagrant

examples of linguistic plagiarism. Example (12), from *Pujante v. Vázquez Montalbán*, illustrates this double circumstance.

(12) MY HEART IS THIRSTY FOR THAT NOBLE PLEDGE.

A. !Mi corazón está sediento de este noble brindis!

V. Mi corazón está sediento de este noble brindis.

P. Mi corazón está sediento de ese brindis.

VM. Mi corazón está sediento de este brindis.

FILL, LUCIUS, TILL THE WINE O'ERSWELL THE CUP.

A. !Llena, Lucio, llena de vino mi copa hasta que se derrame!

V. Llena, Lucio, hasta que el vino rebose de la copa:

P. Lucio, llena la copa hasta que se desborde.

VM. Lucio, llena la copa hasta los bordes.

I CANNOT DRINK TOO MUCH OF BRUTUS' LOVE.

A. Jamás beberé lo bastante por el afecto de Bruto.

V. no puedo beber suficiente del cariño de Bruto.

P. Jamás podré apurar la amistad de Bruto.

VM. Jamás podré apurar la amistad de Bruto.

The first two lines in the four translations under analysis are linguistically parallel. In the third line, however, it is possible to observe several translating strategies used by the translators involved: on the one hand, Astrana Marín's translation involves an error of interpretation since it establishes an inexistent causality and Valverde's text reflects a word-for-word translation; on the other hand, Pujante reproduces the epigrammatic nuance of the original text by the three initial voiced spondees, keeps the line's versification and introduces the innovating translation of "apurar" for "drink too much" and "amistad", rather than "afecto" or "cariño", for "love" (much more appropriate in Cassius' words, since friendship is a key issue in *Julius Caesar*) and Vázquez Montalbán translates this line in exactly the same terms, which produce exactly the same poetic effects.

Other translation strategies observed through the qualitative approach to the disputed (Vázquez Montalbán's translation (1988)) and

the non-disputed texts (Pujante's translation (1987)) include *addition* and *omission* calques, as illustrated in (13) and (14).

(13) O YES, AND SOUNDLESS TOO! (V, i, 31 – 38)

A. ¡Oh, sí! ¡Y también su ruido,

V. Ah sí, y también su ruido:

P. Sin aguijón y sin voz;

VM. Sin aguijón y sin voz;

(14) YON CASSIUS HAS A LEAN AND HUNGRY LOOK (I, ii, 194)

A. He allí a Casio con su figura extenuada y hambrienta.

V. Ese Casio tiene aire macilento y hambriento.

P. Ese Casio tiene un aspecto famélico.

VM. Este Casio tiene un aspecto famélico.

5. Quantitative approaches

5.1. Indexes of textual similarity

Quantitative approaches to plagiarism, using concordance and statistical techniques and tools, allow the researcher to establish in a much more reliable way what base-line one needs to have for degree of similarity between translations and the point at which this similarity becomes suspicious.

The tool CopyCatch⁹ calculates several measures that indicate textual similarity between translations and one of them is *overlapping vocabulary*.¹⁰ Empirical evidence for comparisons of texts written in

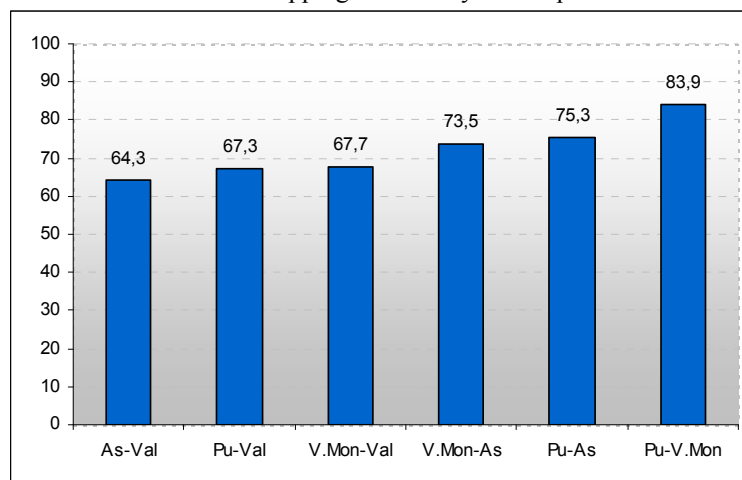
⁹ This tool (created by David Woolls fro CFL Development) contains several programmes that were not exactly designed to detect plagiarism, but that, as Johnson (1997: 220) points out, 'knowing the kind of statistical output they could produce, [...] might shed light on the plagiarized texts in a way that qualitative and manual text analysis cannot adequately do'.

¹⁰ Other measures of textual similarity obtained through CopyCatch are: *shared once-only words* (that occur only once in each file (of the pairs compared) but occur in both members of the pair; the higher the number, both as a raw score and as a proportion of the shared vocabulary, the greater the similarity between the

the same language suggests that up to 35 per cent similarity is normal and up to 50 per cent is not unusual, but the further above 50 per cent the more likely it is to indicate a borrowing relationship between the texts. In cases of plagiarism between translations - still with little empirical evidence available - it seems reasonable to suggest a higher threshold –70 per cent– of *overlapping vocabulary* to show a strong relation between texts.

Figure 1 indicates *inter-translation overlapping vocabulary*¹¹ shared by the four Spanish translations of *Julius Caesar* and shows that the non-disputed and disputed pairs of translation exhibit the highest percentage (83,9%) of shared vocabulary.

Figure 1. Inter-translation overlapping vocabulary Shakespeare's *Julius Caesar* (%)



Source: Turell 2004.

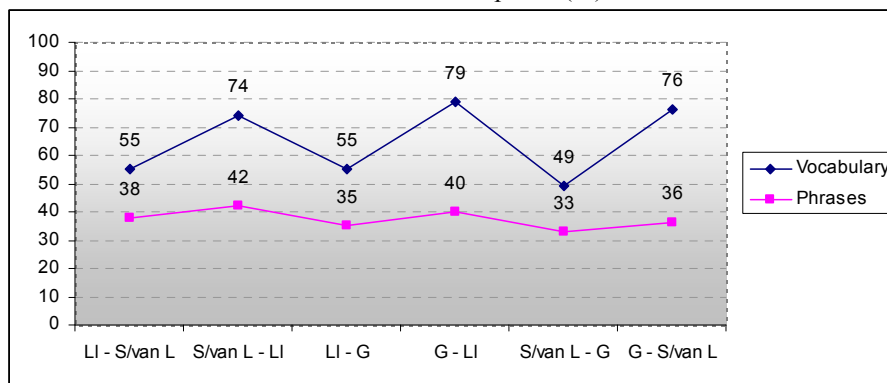
two texts); *unique vocabulary* (words which are unique to each translation; if two translations have been produced independently, they will include a higher proportion of unique lexical items, than those which have not been produced independently; *shared once-only phrases* (the higher the number, the lower the probability that the two texts in the file comparison have been independently produced).

¹¹ The percentage of overlapping vocabulary is calculated with regards the total number of lexical words used in both files. The distinction between lexical (content) and closed (function) words that CopyCatch discriminates is important here. In order to be able to control for certain lexical words, namely patronymics and toponymics, which might be recurrent in all the translations under analysis, it was decided to treat these units as closed words and include them in the stop list, which is a file containing all closed (function) words.

In view of this evidence, it could be argued that in fact apart from the disputed/non-disputed pair (Pujante v. Vázquez Montalbán), two other pairs: Astrana v. Vázquez Montalbán and Astrana v. Pujante, show a high percentage of *overlapping vocabulary*, something which in any case was expected because of the nature of translation itself, whose activity, it should be remembered, involves a tendency on the translator’s side to reflect the author’s original form and content so that its originality becomes something more difficult to detect. This means that other pieces of evidence have to be found in order to be able to raise ‘reasonable doubt’.

Figure 2 shows *inter-translation overlapping vocabulary and phrases* between the three translations of Kafka’s *Metamorphosis* into Catalan (also calculated by using CopyCatch).

Figure 2. Inter-translation overlapping vocabulary and phrases Kafka’s *Metamorphose* (%)



Source: Turell 2007.

This figure allows the investigator of plagiarism between translations of literary works to draw the following conclusions:

- a. Sòria/van Lawick’s translation seems to exhibit a higher percentage of overlapping vocabulary (74%) and phrases (42%) vis-à-vis Llovet’s translation than the percentage of overlapping vocabulary (38%) –not the one referring to phrases (55%)– of Llovet’s translation with regards Sòria/van Lawick’s version. These results would confirm that Sòria/van Lawick’s translation contains much more text from Llovet’s original translation than

the other way round, and that there is linguistic plagiarism, whose directionality in any case is in turn confirmed by publication date (Llovet 1985; Sòria/van Lawick 1989).

- b. Gala's translation seems to exhibit a higher percentage of overlapping vocabulary (79%) and phrases (40%) vis-à-vis Llovet's translation than the percentage of overlapping vocabulary (35%) –not the one referring to phrases (55%)– of Llovet's translation with regards Sòria/van Lawick's version. These results would confirm that Gala's translation contains much more text from Llovet's original translation than the other way round, and that there is linguistic plagiarism, whose directionality in this case is in turn confirmed by publication date (Llovet 1985; Gala 2004).
- c. Gala's translation seems to exhibit a higher percentage of overlapping vocabulary (76%) and phrases (36%) vis-à-vis Sòria/van Lawick's translation than the percentage of overlapping vocabulary (33%) and of phrases (49%) of Sòria/van Lawick's translation with regards Gala's version. Once again, these results would confirm that Gala's translation contains much more text from Sòria/van Lawick's translation than the other way round, and that linguistic plagiarism has taken place.

A more qualitative exploitation of Copycatch allows the analyst to examine identical and very similar lexical strings by using the *Marked Up-Sentences Only* option. In Table 1 (reproduced from Turell 2004) we can see the first 16 lines of Act I, Scene 1 in Pujante's and Vázquez Montalbán's translation of *Julius Caesar*, where the identical strings are marked in boldface.

Table 1. Phrase sharing (Pujante’s and Vázquez Montalbán’s translations) – Turell 2004¹²

| <i>Julius Caesar</i> , Act I, Scene 1 | | Characters: Marullus, Flavius and a carpenter |
|---|---|---|
| PUJANTE | VÁZQUEZ MONTALBÁN | |
| [P3 S6] ¿No sabéis los artesanos que no debéis andar en día de trabajo sin el distintivo de vuestra ocupación? {P5 S5} | {P5 S5} ¿Acaso no sabéis los artesanos que en días laborables debéis andar por las calles con el distintivo de vuestro oficio? [P3 S6] | |
| [P3 S8] Señor, soy carpintero. {P6 S1} | {P6 S1} Pues, señor, soy carpintero. [P3 S8] | |
| [P4 S1] ¿Dónde está tu mandil? {P7 S1} | {P7 S1} ¿Dónde está tu mandil de cuero? [P4 S1] | |
| [P4 S2] ¿Y tu escuadra? {P7 S2} | {P7 S2} ¿Y tu escuadra? [P4 S2] | |
| [P5 S1] Sinceramente, señor , al lado de un artesano de verdad, yo sólo soy lo que pudiéramos llamar un chapucero. {P8 S1} | {P8 S1} Para ser sincero, señor , si me comparo con un artesano de verdad sólo soy lo que podríamos llamar un chapucero. [P5 S1] | |
| [P6 S1] Pero , ¿qué oficio tienes? {P9 S1} | {P9 S1} Pero , ¿cuál es tu oficio? [P6 S1] | |
| [P7 S1] Señor, un oficio que siempre hace el bien: a quien mal anda, lo con-suela. {P10 S2} | {P10 S2} A quien mal anda lo con-suela. [P7 S1] | |
| [P10 S1] ¿Qué quieres decir? {P13 S1} | {P13 S1} ¿Qué quieres decir? [P10 S1] | |
| [P11 S1] Claro, señor : componerte los zapatos. {P14 S1} | {P14 S1} Claro, señor , componerle los zapatos. [P11 S1] | |
| [P12 S1] Así que eres zapatero remendón. {P15 S1} | {P15 S1} Acabáramos; así que eres un zapatero remendón. [P12 S1] | |
| [P13 S2] Vivo del punzón. {P16 S2} | {P16 S2} Vivo de la herramienta. [P13 S2] | |
| [P13 S3] Yo no me meto en asuntos de artesanos, ni en asuntos de | {P16 S3} Y no me meto en asuntos de trabajo, ni de mujeres si no es con la | |

¹² The identical or very similar strings are highlighted in **bold** and this fact helps the reader to see that they are much more numerous than the different strings which are normal type. The codes in brackets appearing at the beginning of the sentences analysed indicate the paragraph number and paragraph sentence of the text under analysis, and those appearing at the end indicate the matched paragraph and sentence in the text under suspicion where the identical or similar strings appear.

| | |
|--|--|
| mujeres ... si no es con el punzón. {P16 S3} | herramienta por delante. [P13 S3] |
| [P13 S5] Si es cuestión de medias suelas, yo los remedio. {P16 S6} | {P16 S6} Si el mal es de medias suelas, yo lo remedio. [P13 S5] |
| [P13 S6] Todo el que ha caminado sobre cuero ha pasado por mis manos. {P16 S7} | {P16 S7} Todos los hombres decentes que pisan sobre cuero han pasado por mis manos. [P13 S6] |
| [P14 S1] Pero, ¿por qué no estás hoy en el taller? {P17 S1} | {P17 S1} ¿Y por qué no estás hoy en el taller? [P14 S1] |
| [P15 S1] Pues, señor, para que gasten los zapatos y tener yo más trabajo. {P18 S1} | {P18 S1} Pues, señor, para que gasten sus zapatos y así tener yo más trabajo. [P15 S1] |

5.2. The use of corpora in plagiarism detection

Reference corpora are useful detection tools to indicate the rarity or otherwise of the choices made by authors, and thus establish the uniqueness or commonality of the texts produced, in this case, translations. Diachronic corpora can help the analyst establish the uniqueness of a translator's choices *vis-à-vis* the language used at the time the original writer was writing while synchronic corpora are useful to establish the rarity of a translator's choices in the contemporary uses of the language into which an original literary work is translated.

Table 2 (reproduced from Turell 2004) shows the results of the searches done in the *CREA* (*Corpus de Referencia del Español Actual*) and the *CORDE* (*Corpus Diacrónico del Español*) databases of the Spanish Corpus of the *Real Academia Española* (RAE), using both texts written in present-day Spanish and also texts written by Shakespeare's (1564–1616) contemporaries, such as Cervantes (1547–1616), in order to indicate whether the choices made by Pujante and then 'borrowed' by Vázquez Montalbán, were rare or otherwise. The contrastive units chosen to do the searches include examples found in the CopyCatch evidence, that is, words, phrases, greetings, whole sentences, idioms and puns.

Table 2. Corpora searches in the CORDE and CREA databases (RAE) – Turell 2004

| | <i>CORDE</i> | <i>CREA</i> |
|--|--------------|-----------------|
| <i>WORDS</i> | | |
| Carpintero | 2/2 (1) | 244/128 (1,9) |
| Mandil | 5/3 (1,6) | 90/52 (1,7) |
| Escuadra | 20/10 (2) | 319/220 (1,4) |
| Artesano | 0 cases | 315/185 (1,7) |
| Chapucero | 0 cases | 45/40 (1,1) |
| Oficio | 231/28 (8,2) | 3122/1401 (2,2) |
| Toga | 2/2 (1) | 129/85 (1,5) |
| <i>PHRASES</i> | | |
| zapatero remendón | 0 cases | 11/9 (1,2) |
| medias suelas | 0 cases | 9/7 (1,2) |
| el yugo de los tiempos | 0 cases | 0 cases |
| un aspecto famélico | 0 cases | 0 cases |
| <i>GREETINGS</i> | | |
| Buenas noches | 0 cases | 1052/389 (2,7) |
| <i>WHOLE SENTENCES</i> | | |
| Jamás podré apurar la amistad de | 0 cases | 0 cases |
| y quisieran que no fuese tan ciego | 0 cases | 0 cases |
| <i>IDIOMS</i> | | |
| Vivimos unos tiempos singulares | 0 cases | 0 cases |
| (Sin) aguijón | 0 cases | 175/112 (1,5) |
| no volveríais a mirarme a la cara | 40/40 (1) | 0 cases |
| Así es como los dioses censuran al cobarde | 0 cases | 0 cases |
| <i>PUNS</i> | | |
| A quien mal anda, lo consuela | 0 cases | 0 cases |

This table indicates that the majority of searched forms do not show any occurrence in either the diachronic (CORDE) or the contemporary (CREA) corpus, and in those cases where tokens occur, the density of the ratio between number of cases and number of texts is very low, between 1 and 2.7, except for one case which reaches 8.2. These results account for the rarity of the choices made by Pujante and thus confirm that if any of these same choices appear in Vázquez Montalbán's

translation as well, it is very unlikely that he would have produced them independently, that is, without having ‘borrowed’ them from Pujante.

6. A comparative evaluation of methods

Not all methods and approaches can be used in all cases. Each plagiarism case might require a different approach, but if the nature and amount of data are relevant and appropriate, more accountable results will be obtained when qualitative and quantitative methods are used in combination. Also, it is useful to begin an expert witness report by incorporating into it types of qualitative evidence that were used in similar cases, but pointing out the idiosyncrasy of the new case and the need for a different quantitative approach.

My evaluation of the methods and approaches that have been discussed in this article is done by pulling together the criteria used in the analysis, that is, a) the nature of the plagiarised units and categories and their degree of similarity or difference; b) the ease or difficulty of the different approaches, and c) the advantages and disadvantages of taking one method or the others.

a. Qualifying approaches

1. Qualitative approaches allow analysts of plagiarism in translation to propose a global characterization of the translations under analysis, in terms of literal, word-for-word, rhetorical, and flowery translations, among others. Only these type of approaches allow the accurate observation of the literary discourse, the stylistic and figurative resources, the echoes, quotations, and references from the original work that a translator has been able to decode and of the hyperbole, similes, metaphors, images, rhyme, syncope, ellipsis, and other figures of speech that he or she has been able to reproduce in the target language; furthermore, only using qualitative methods and observing these types of units and categories one is able to observe whether or not a new translation is a mere paraphrase of a previous one.

2. A qualitative approach also facilitates the observation of discrete units which are bound to be plagiarised in terms of taxonomies such as *one-word calques*, *multi-word calques*, *whole-verse calques* and *calques of a series of verse*.

3. This more qualitative approach also allows the researcher to detect “borrowed” distinctive markers, that is, syntactic, semantic and pragmatic structures which are difficult to locate with automatic detection tools.

4. And finally, qualitative observational approaches are useful to detect the most frequent translation strategies (*addition*, *omission* and *recreation*), which are further calqued by the plagiarist translator.

b. Quantifying approaches

1. The establishment of the statistical significance of *overlapping vocabulary*, shared *once-only words*, *words unique to each file*, *shared once-only phrases*, and identical or similar phrasing by means of using CopyCatch, allows the analyst to claim that the higher proportion of these counts in the comparison of translation pairs is an indication of plagiarism, particularly when they are above 70 per cent in the case of overlapping vocabulary.

2. Another advantage in using the CopyCatch approach is that it is possible to look at all these categories in context since they appear as marked units within the whole text, and also that there is an indication of the paragraph and sentence of the other text that matches the former. And this can be done for all translated pairs of texts under consideration.

3. The use of reference corpora is useful in establishing the rarity/expectancy (or commonality) of the choices made by a particular translator.

4. And finally, as far as plagiarism directionality is concerned, in order to establish which text is the source text (T1) and which is the derived one (T2), if two texts are contemporary, it would be necessary to use a third text (T3) to see whether T2 and T3 have more in common with T1 than with each other (Johnson (1997)).

References

- Astrana Marín, Luis. 1961. *William Shakespeare. Obras Completas*. Madrid: Ed. Aguilar.
- Coulthard, R. Malcolm. 2005. "Algunas aplicaciones forenses de la lingüística descriptiva". In *Lingüística forense, lengua y derecho. Conceptos, métodos y aplicaciones*, M. T. Turell (ed.). Barcelona: Publicacions de l'IULA. 249-273.
- Delabastita, Dirk. (ed.) 1997. *Traductio: Essays on Punning and Translation*. Manchester: St. Jerome.
- DeVoss, D. and Rosati, A.C. 2002. "It wasn't me, was it? Plagiarism and the Web". *Computers and Composition* 19 (2): 191-203.
- Downing, Angela. 1989. "Translating Grammatical Metaphor". In *Translation Across Cultures. La traducción entre el mundo hispánico y anglosajón. Relaciones lingüísticas, culturales y literarias. Actas XI Congreso AEDEAN*. J.C. Santoyo (ed.). Universidad de León. 87-96.
- Downing, Angela, and Locke, Philip. 1992. *A University Course in English Grammar*, A. Downing and P. Locke (eds.). Hemel Hempstead: Prentice Hall International.
- Flotow, Luise. 1997. *Translation and Gender: Translating in the 'Era of Feminism'*. St Jerome Publishing/University of Ottawa Press.
- Gala, C. (2004) *La Metamorfosi*. Barcelona: Ed. Columna.
- Gibbons, John and Turell, M. Teresa (eds.). 2008. *Dimensions of Forensic Linguistics*. Amsterdam/Philadelphia: John Benjamins.
- Kafka, Frank. 1980. *Die Verwandlung*. Frankfurt am Main: Fisher Verlag GmbH.
- Llovet, Jordi. 1985. *La Metamorfosi*. Barcelona: Edicions Proa.
- Johnson, Alison. 1997. "Textual kidnapping – a case of plagiarism among three students texts?" *Forensic Linguistics* 4 (2): 210–25.
- Menasche, Lionel. 1977. *Writing a Research Paper*. (Ann Arbor: The University of Michigan Press, 1977; revised Edition: Pitt Series in English as a Second Language): 55-60.
- Pozo, Cristina. 2006. *El plagi en la traducció de Kafka*. Unpublished report on forensic linguistics. Barcelona: Universitat Pompeu Fabra.

- Pujante, Ángel Luis. 1987. *William Shakespeare. Julio César. Traducción, introducción y notas*. Universidad de Murcia.
- Rodríguez Tapia, José Manuel. 1995. "Traduttore, traditore: Plagio y originalidad de una traducción". Comentario a la sentencia del Tribunal Supremo (Sala 1ª) (1993). *Revista General de Derecho* 604-5: 771-779.
- Roig, Miquel. <http://facpub.stjohns.edu/~roigm/plagiarism/Acknowledging%20the%20source.html>
- Turell, M. Teresa. 2004. "Textual kidnapping revisited: the case of plagiarism in literary translation". *The International Journal of Speech, Language and the Law* 11 (1): 1-26.
- Turell, M. Teresa. 2005. "El plagio en la traducción literaria". In *Lingüística forense, lengua y Derecho. Concepto, métodos y aplicaciones*, M.T. Turell (ed.). Barcelona: Publicacions de l'IULA. 275-298.
- Turell, M. Teresa. 2008. "Plagiarism". In *Dimensions of Forensic Linguistics*. J. Gibbons and M.T. Turell (eds.). Amsterdam/Philadelphia: John Benjamins. 265-299.
- Turell, M. Teresa. 2010. "The use of textual, grammatical and sociolinguistic evidence in forensic text comparison". *The International Journal of Speech, Language and the Law* 17 (2): 211-250.
- Vallduví, Enric. 1992. *The Informational Component*. New York/London: Garland Publishers.
- Valverde, José María. 1968. *William Shakespeare Teatro Completo*. Barcelona: Ed. Planeta.
- Van Lawick, H. y Sòria Parra, E. 1989. *La Metamorfosi*. Valencia: Edicions Bromera.
- Vázquez Montalbán, Manuel. 1988. *Julio César de W. Shakespeare*. Madrid: Centro Dramático Nacional.
- Woolfs, David. 2003. <http://www.jiscmail.ac.uk/lists/plagiarism.html>.
- Zozaya, Pilar. 1990 *Dictamen pericial* (manuscript). Barcelona.