

**TWO-DIMENSIONAL THEORY OF STYLE IN TRANSLATIONS:
AN INVESTIGATION INTO THE STYLE OF LITERARY
TRANSLATIONS**

by

DOMINIK VAJN

A thesis submitted to
The University of Birmingham
for the degree of
MASTER OF PHILOSOPHY (B)

The School of English, Drama and
American & Canadian Studies
College of Arts and Law
The University of Birmingham
S e p t e m b e r 2 0 0 9

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

The aim of this thesis was to address the issue of authorship of translations from two perspectives: firstly, the aim was to engage in the debate, and propose a theory on the authorship of translations; and secondly, to propose a method that would support the theory with empirical evidence. For this purpose a corpus of two translations of Plato's Republic was used, and the two translations were compared for their style.

Firstly, a method that ensures a step-by-step analysis of the texts, rather than the analysis of the whole texts was developed. The results showed that the method successfully compared two translations, demonstrating high degrees of correlation between the styles of the two translations. In other words, the method demonstrated that the unitary style of a translation is in fact a fusion of the translator's and the author's styles.

Thus, this research demonstrated the two-dimensional nature of the style of translations. Consequently, since the style and authorship are closely related, the research showed that a translation is in fact a result of the collaboration between the author and the translator. In other words, a translation is a result of co-authorship.

Dedication

For my family and friends

Acknowledgments

Firstly, I would like to thank my supervisor, Professor Wolfgang Teubert, for all the advice and guidance he provided regarding the research and writing of the thesis. Furthermore, I would like to thank Dr Nicholas Groom and Dr Pernilla Danielsson, who introduced me to corpus linguistics, showed me the ways to apply it, and helped me with this research with their advice and support.

Special thanks to John Olsson, with whom I had many fruitful discussions on the concept authorship, and who often helped me with my writing, often pointing me to my spelling and grammatical mistakes.

I would also like to thank the Ministry of Science, Education and Sports of the Republic of Croatia for their financial support, namely their contribution towards my tuition fees. Additional thanks to my former supervisors and lecturers at the University of Osijek, Croatia, Dr Marija Omazić, Dr Mario Brdar, Dr Gabrijela Buljan, Tanja Gradečak-Erdeljić and others, for making me interested in linguistics.

Finally, I would like to thank my family: my father, Krešimir, my mother, Dragica, and my sister, Katarina, who have supported me, both financially and emotionally, throughout all these years.

Table of Contents

INTRODUCTION.....	1
LITERATURE REVIEW.....	4
ON AUTHORSHIP.....	4
AUTHORSHIP OF TRANSLATIONS	12
STYLE AND AUTHORSHIP	18
A CASE OF PLAGIARISM.....	22
TWO-DIMENSIONAL THEORY OF THE STYLE OF TRANSLATIONS.....	26
METHODOLOGY.....	35
RESEARCH ANALYSIS	38
STANDARDISED TYPE-TOKEN RATIO.....	38
TYPE-TOKEN RATIO.....	39
SHARED VOCABULARY.....	41
RELATIVE VOCABULARY RICHNESS.....	44
DISCUSSION	46
CONCLUSION.....	53
APPENDIX 1.....	56
APPENDIX 2.....	59
APPENDIX 3.....	60
BIBLIOGRAPHY	61

List of Figures and Tables

Figures:

FIGURE 1 - PERCENTAGES OF OVERLAPPING VOCABULARY BETWEEN THE PAIRS OF TRANSLATIONS	24
FIGURE 2 - STANDARDISED TTR VALUES THROUGH SEGMENTS.....	29
FIGURE 3 - NUMBER OF SHARED TYPES THROUGH SEGMENTS	31
FIGURE 4 - NUMBER OF TYPES USED IN EACH TRANSLATION THROUGH SEGMENTS.....	31
FIGURE 5 - sTTR VALUES THROUGH STEPS	38
FIGURE 6 - TTR VALUES THROUGH STEPS.....	40
FIGURE 7 - THE NUMBER OF SHARED TYPES AND TOTAL NUMBER OF TYPES THROUGH STEPS	42
FIGURE 8 - THE NUMBER OF SHARED TYPES THROUGH STEPS.....	43
FIGURE 9 - RELATIVE VOCABULARY RICHNESS THROUGH STEPS.....	44

Tables:

TABLE 1 - CORRELATION BETWEEN sTTR VALUES OF THE TWO TRANSLATIONS	30
TABLE 2 - CORRELATION BETWEEN THE NUMBER OF SHARED TYPES AND THE TOTAL NUMBER OF TYPES	32
TABLE 3 - CORRELATION BETWEEN sTTR VALUES.....	39
TABLE 4 - CORRELATION BETWEEN TTR VALUES.....	41
TABLE 5 - CORRELATION BETWEEN THE NUMBER OF SHARED TYPES AND THE TOTAL NUMBER OF TYPES	43
TABLE 6 - CORRELATION BETWEEN RVR VALUES	45

Introduction

The notion of authorship has been addressed by many researchers with different scientific backgrounds as well as different motivation and approaches to the issue, ranging from linguistic and literary scholars to computer scientists and legal theorists. The motivation for most of the research comes from the need to advance the notion and/or application of the rights of intellectual property. Therefore, most of the research has until now focussed on defining the scope of the intellectual property or methods for identifying authorship, or its infringement. However, due to the complexity of the notion of authorship, there is still a great potential and need for further research. Specifically, one of the issues that has barely been addressed in authorship studies is the authorship of translations.

Thus far, the notion of authorship of translation has been the primary concern of translation theorists, and some legal theorist, but only to the extent of necessity, viz. to define the intellectual rights of the translator. Several translation theorists have proposed some theories on authorship of translations; however, none of these have been supported by empirical research that would confirm these assumptions.

The aim of this thesis is to address the issue of authorship of translations from two perspectives: firstly, the aim is to engage in the debate, and propose a theory on the authorship of translations; and secondly, to propose a method that would support the theory with empirical evidence.

The research on the method aims to present a method for analysing the style of translations. Specifically, as the notions of style and authorship are related, and authorship attribution is based on stylistic analysis, a method for analysing the style of translations would be essential for an empirical research into the issue of authorship of translations. Based on the empirical evidence obtained by the application of such method, the issue will be discussed from a more theoretical aspect. Namely, it will be discussed which of the theories can be related to the results of the analysis, and whether these results support some theories and oppose the others, or whether they offer a completely new perspective on the issue.

It is important to point out that this research builds upon the results of a preliminary research. That research has resulted with a method to analyse the style of

translations as well a theory on that style, i.e. the two-dimensional theory of the style of translations. However, as the method was rather crude and experimental, the results obtained by it can only serve as a guide for further research rather than evidence to support the theory derived from them. Furthermore, since the theory is derived from a preliminary research that has not been previously or published previously, it is necessary to present and explain the theory in detail together with the method which it has been derived from. Thus the two-dimensional theory will serve as a hypothesis that will have to be demonstrated using a more sophisticated method.

As this research aims to address the issue of the authorship of translations, the research will focus on literary translations rather than non-literary ones, as non-literary translations are not always subject to intellectual property, e.g. translations of *acquis communautaire*. Thus, for the purpose of this research I have built what Olohan (2004) calls 'a monolingual comparable corpus' consisting of two translations of 'The Republic' by Plato. The first text is a nineteenth century translation by Benjamin Jowett and the second one is Robin Waterfield's translation from 1990s. Even though one would expect that all translations would be quite similar, at first glance, these two seem quite different. Namely, Jowett's translation consists of about 110,000 words and ten chapters, whereas Waterfield's translation consists of about 130,000 words and fourteen chapters. The gap between the productions of these translations, as well as their difference in length, makes them an interesting pair for a comparison on their styles.

The thesis will be divided into several sections. First section will be the literature review, which will be followed by a section on a case of a plagiarised translation. Then two-dimensional theory will be presented together with the method used in the preliminary research. The methodology section will present a new method that the theory will be tested with, and the results of the research will be presented in the following section. These will be followed by sections on discussion and finally conclusion.

In the literature review, firstly, the notion of authorship will be discussed, both from legal and linguistic perspective. A historical perspective of the concept of authorship will be presented with the aim of demonstrating the influence of former theories on the contemporary view on authorship. This will be followed by a discussion on the historical and contemporary views on authorship of translations. Namely, theories from several

translation theorists will be discussed, as well as some of the legal views on the issue. The aim is to present contemporary theories on authorship of translations so that they can be later critically assessed and discussed both in theory and in practice. Additionally, the relation between style and authorship will be discussed as well as some issues related with this research, e.g. statistical methods used in the research.

Subsequently, a case of plagiarism of translations reported by Turell (2004) will be presented, with the aim of demonstrating some of the authorship issues of translations, i.e. the high degree of similarity between translations of the same original text. The section will also focus on critically assessing the concept of authorship, which the case has been based on. Namely, the effects of the concept on the analysis and interpretation of results will be discussed, and followed by a discussion on the effects that an alternative concept of authorship might have had. The aim is to stress the difference between these theories, and to demonstrate the importance of an empirical research which would contribute to a better understanding of this complex notion.

The central part of this thesis will present the results of the preliminary research as well as the two-dimensional theory derived from them. Then, the theory will be related to the existing theories on the authorship of translations. Furthermore, the caveats of the preliminary methodology will be discussed, stressing the issues where the method can and has to be improved.

This will be followed by a detailed description of the new methodology applied in the research. This will include a presentation of a new measure to be used in corpus linguistics, namely the *relative vocabulary richness*. Subsequently, the results of the analysis will be presented and interpreted.

Finally, in the discussion section the results of the research will be related to the proposed theory. Furthermore, the two-dimensional theory of the style of translations will be compared with other theories of authorship of translation, regarding the processes of production of translations and the translator's influence on the translations. Additionally, the findings will be related to the issues identified in the plagiarism detection study. The section will also present some alternative explanations of these results, as well as point out all the possible caveats of this research. The thesis will end with a summary of the research that will be presented in the conclusion.

Literature Review

On authorship

Authorship is a very complex notion; it is abstract and created by man. However, the authorship concept was not created in one particular instance, but developed through human history. Therefore, the question one often asks is ‘Who was the first author?’ Another question that comes to mind is in what sense that person was the first author, as the concept was not always perceived and understood as it is today. Thus it is important to look at the history of authorship concept so as to have a better understanding of the contemporary concept.

As authorship is a very complex notion, the question of who was the first author is not an easy one to answer. Is Shin-eqi-unninni, the supposed author of the epic of *Gilgamesh*, the first author because his name was next to the earliest literary text we have found thus far? Additionally, the myth of Gilgamesh precedes Shin-eqi-unninni by 700 years, thus challenging the hypothesis of him being the creator of the epic. In similar fashion, Homer is considered to be the first Western author with his *Odyssey* and *Iliad*. However, just like with Shin-eqi-unninni and *Gilgamesh*, the events depicted in *Odyssey* and *Iliad* precede Homer by a couple of centuries. Thus the ‘Homeric question’, namely the debate over the authorship of *Odyssey* and *Iliad*, was raised.

Instead of proposing that there might be one author of these epics, Lord (1960) proposed that Homer should be considered both as a poet and as an oral tradition. Lord (ibid) argues that even though Homer might have brought *Odyssey* and *Iliad* to the level of perfection, he inherited both the stories and his compositional techniques from a centuries old oral tradition. Specifically, in the oral tradition, every performance is unique, as the singer is both a creator (*poietes*) and a reciter (*rhapsode*). At the same time, every performance is a repetition and a part of the oral tradition (Lord 1960). Thus authorship of these epic lies in the oral tradition, and not within a single poet. Hence, Lord (1960:147) declares that Homer is not an individual author, but *the tradition*.

Nagy (Nagy 1996, cited in Bennett 2004) proposes that the figure of Homer is a result of back-formation or recomposition. Specifically, Nagy (ibid) suggest that in oral

tradition, a poem progresses from one poet to the other, each of them building on the work of the previous poet, and at some point the differentiation ends with a certain poet, as there is no more need for further elaboration or development. Thereafter the aim is to preserve it as it is, since the poem is considered to be finished (ibid). Thus each poet in the oral tradition is effectively the co-author of the final product. However, at the point when differentiation ends, the last performer is mythologized as the author and considered the originator of the poem, whereas all following poets in the tradition are considered mere reciters (rhapsodes) of the poem (ibid), whereas all the previous ones were long forgotten. The aim of these reciters is to preserve the poem, not only by retelling it as it is, but also by impersonating the poet. Thus Homer became an archetypical poet through a process of retrospective figuration in the oral tradition (ibid), namely he was mythologized as the originator by the reciters who wished to preserve his version of the epic and the manner it was recited.

However, the names of Homer and Shin-eqi-unninni were more an exception than general practice, as normally, literary works were not assigned to any author. On the other hand, with historians, legislators and even philosophers, it was quite common to find their names on their writings. Namely, such people had the *authority* in their societies, whereas literary writings were regarded unimportant to the society, because they were made by ordinary people, such as poets and bards. Thus, for example, Herodotus, Plato, Aristotle, Plutarch, Cicero and many others who had the authority in the society had their names written on their texts with the intent to assign that authority to their texts.

In the Middle Ages the authorship concept was similar to the ancient concept, and in practice authors were seen as authoritative writers, whose writings contributed to the knowledge and wisdom of humanity (Bennett 2004:39). Thus, just as in ancient times, the author was a person who possessed *authority* and made authoritative statements which had credibility and were to be believed (ibid). However, contemporary vernacular medieval writers were hardly considered to be authors, as the great authors of the past, both Christian and pagan, have already said almost everything there was to say (Burrow 1982:32, cited in Bennet 2004:39). In that sense there were no contemporary vernacular authors, and all the authors were dead, even the languages they wrote in, i.e. Greek and

Latin, were dead, and thus the task of contemporary writers was to only interpret these authorities (ibid).

Medieval vernacular writers were mostly anonymous, not only because in most cases they did not have the authority, but also due to practical reasons. Namely, in the Middle Ages, the process of copying books was done manually, and, normally, copies were only distributed amongst limited circle of people. Thus there was no need to add the writer's name to a manuscript as everyone knew the writer (Bennett 2004). On the other hand, if the manuscript was distributed more widely, the name of the writer became irrelevant as the writer was unknown to readers outside his community (ibid).

Generally, the name of the writer was irrelevant, as the contemporary vernacular writer was not seen as an authority, but just as someone who transmits a story, and the readers were interested in the work and the truths it revealed and not in the writer (Saunders 1964:19-20, cited in Bennett 2004:41).

Although the medieval concept of authorship was fundamentally different from the modern concept, the beginnings of the modern concept of authorship can be traced back to the late Middle Ages (Kimmelman 1991:21, cited in Bennett 2004:41). Namely, contemporary vernacular writers wanted to be considered authors, they wanted their work to be just as good as the works of the authors. Thus they imitated the old authors to the best of their abilities, but in their own language. In that sense, becoming an author involved both self-assertion and submission to the tradition (ibid). Kimmelman also noticed that one did not become an author just by putting their name on the text, but through the very craft of authorship (Kimmelman 1999:7, cited in Bennett 2004:41).

Specifically, in the late Middle Ages, writers, such as Chaucer, were becoming known to the readers through their writings, thus they were effectively becoming authors in the modern sense. However, the medieval conception was still present in Chaucer's time, which can be observed from Chaucer's *Prologue to The Canterbury Tales*, where Chaucer declares himself as the compiler of the tales and not the author (Bennet 2004:42). Chaucer's decision to disavow authorship, and give it to his own characters is an example of the transitions and contradictions in the understanding of authorship in the late Middle Ages (Trigg 2002:54-5, cited in Bennett 2004:42-3). Nevertheless, with Chaucer started

the development of the modern sense of the author, which was soon to be boosted by the development of printing press.

Before the printing press, there was a problem of stabilising texts, as in several successive copies there were often discrepancies between the 'original' and the copy. More importantly, the problem for contemporary writers was the fact that scribes often put their name of the copy (Bennett 2004), thus the 'original' was lost among copies. Additionally, in the process of manuscript transmission, it was quite normal for texts to be edited because the copying of a manuscript was a form of discourse (Marotti 1995:135, cited in Bennett 2004:44), just like a poem would become a part of discourse, and thus subject to changes, in the earlier oral tradition.

With the discovery of the printing press, texts have become more stabilised, but also more accessible to people because the process was much cheaper and less time consuming than manual transcription. However, the importance of printed texts was in that they have derived their authority from a different source than manuscripts. Specifically, manuscripts derived their authority from the place of production, such as courts, monasteries and universities, whereas printed texts had to rely on their content to deserve authorisation (Wall 1993:8, cited in Bennett 2004:47).

In the beginning, there was a lot of prejudice against printed texts as they were considered to undermine the fragile class boundary between the aristocracy and lower gentry (Bennett 2004:46). Nevertheless, printing offered writer's a possibility to be recognised as authors by expressing their individuality and talent, and, in return, successful authors removed the stigma from printed texts, not only by publishing their work in print, but also by demonstrating the potential financial benefit of print (Bennett 2004:49).

For centuries the Church controlled the production of books and held the authority over all publications, but with the discovery of the printing press, the authority of the Church over books gradually comes to an end. Namely, with the act of Queen Mary in 1557, the exclusive right to publish was given to the Stationers' Company, whose members were given monopoly on publishing of any book. However, this meant only that the right to publish was transferred from the Church to the state, and the individual writer still had no control over printing and copying of his work (Olsson 2009:28). In other

words, even though some writers were starting to be recognised as authors, the notion of intellectual property was yet to be established, and the right to copy (publish), or copyright, was exclusive to the Stationers' Company.

About the end of 16th century, more and more writers used to put their name on their work, but it is uncertain whether it was their own decision to do so, or the publisher's. As mentioned above, once the a piece of writing given to be published, all the right over publishing a piece of work was in the hands of the publisher. Thus it is uncertain whether it was the publisher's or the writer's decision to name the author of the work. Namely, as writer's established themselves as authors among their readers, or on the market, as seen in the eyes of the publisher, having a name of a popular author written on the front page was a good marketing trick. Thus the publisher cooperated with the author with the intent to secure exclusive rights on publishing their further work.

Every publisher that was a member of the Stationers' Company had copyright on all of their publications. However, eventually, publishers raised the question of literary property focussing on the limitation of the then perpetual copyright term. This was essentially a commercial issue between publishers, as authors had no rights over their work once they have sold it to the publisher (Rose 1993:4). The debate on literary property resulted with the Statute of Anne of 1710, which was the world's first copyright law.

The Statute confirmed the rights of the Stationers' Company, but also introduced two important changes: 1) the copyright term was limited, and 2) authors were recognised as possible proprietor of their work (ibid). Thus the Statute of Anne marks the birth of the modern concept of authorship, as authors have been legally recognised as owners of their work. The concept of literary, or intellectual, property, resulted much from the debate between publishers, but another factor also played a major role in its development, namely, the Romantic view on author as a Genius.

About the time the author's rights to intellectual property were being discussed, the practice of authorship was seen as an act of original creation, the work of original Genius. Unlike learning, which distinguished man from all other beings, Genius distinguished the man of genius from all other people. Thus authors were seen to have

Genius, and/or being Geniuses, because they stopped repeating what has already been said before, and started to create something new.

Derrida (2002:168) described the medieval practice of authorship *revelatory invention*, or the production of what already was, as everything that was being said was just interpretation of what the great authors had already said, whereas the work of ‘Genius’, as the Romantic author was often perceived, he described as creative invention, or the production of what is not. Authors at the time were seen as creators of something new out of nothing, and thus were seen as having divine properties, which later became known as Genius. In 1815, Wordsworth described genius as ‘the introduction of a new element into the intellectual universe’ (quoted from Wordsworth’s *Essay, Supplementary to the Preface* in Woodmansee and Jaszi, 1994:16).

Edward Young, in his *Conjectures on Original Composition* (1759) related authorship with originality, and hence with genius, insisting on the true author as independent self-creating persona (cited in Bennett 2004:59). Author was seen as an isolated figure, but as a person of Genius above the society rather than outside; that was the myth of the author – an isolated creator and originator.

The contemporary concept of authorship stems from this romantic myth of the author as the creator and originator, one that produces what is not, one that creates *ex nihilo*. However, more recent studies have challenged the romantic myth of the author. Namely the introduction of the concepts of polyphony (Bakhtin, in Morson and Emerson, 1990) and intertextuality (Kristeva, 2002) challenged the myth of the author as the creator and originator.

Apart from the more philosophical works of Bakhtin and Kristeva, many other contemporary authorship studies challenged the myth. Thus, for example, Rose (1993:8) states that “authors do not really create in any literal sense, but rather produce texts through complex processes of adaptation and transformation.” Barnstone (1993:87) adds that in reality there is no true author of any text, since there is no work of literature which is not an adaptation from an earlier source.

Even though these modern theories have contributed to a better understanding of the complexity of the concept of authorship, in many areas of the modern society, the notion of Genius still has a major impact on the understanding of notion of authorship,

which is still often described just as a creative and originary act. Specifically, by looking up the word ‘author’ in contemporary dictionaries one will find some of the following definitions:

- ‘(1) *a person who writes books, or the person who wrote a particular book; (2) the person who creates or starts something, especially a plan or idea*’ – Oxford Advanced Learner’s Dictionary of Current English (2000)
- ‘*The author of a piece of writing is the person who wrote it*’ – Collins COBUILD English Dictionary for Advanced Learners (2001)
- ‘*an originator or creator*’ – Collins Essential English Dictionary (2nd Edition) (2006)

It is unlikely that the authors of these examples were not aware of the issues in authorship studies like e.g. intertextuality, polyphony, collective authorship, adaptation, parody and translation, but still, probably due to practical reasons, or tradition, they have rather embraced the still-present romantic myth.

More importantly, the romantic notion of authorship is not only found in contemporary dictionaries, but in contemporary legal writings as well. Depending on the subject area of the law, there are six possible ways of looking at authorship from the legal perspective:

- ***Authorship is mind over muscle*** – *it is the directing mind which is the author, not the person who carries out the author’s instructions;*
- ***Authorship is mind over machine*** – *not all machines are capable of assisting in authorship. For example, a word processing programme is not an author. However, a camera – used with the intention of creating an original photograph is an essential part of the author’s toolkit, because without it the picture wouldn’t exist;*
- ***Originality is synonymous with authorship*** – *apparently straightforward, as Ginsburg notes, but problematical in that different jurisdictions set different levels for originality (e.g. civil vs. common law jurisdictions);*

- ***Perspiration is sufficient, originality is not necessary*** – this implies that it is not necessary to be skilful, as long as the author exerts him/herself (some common law traditions);
- ***The author must intend to be an author, to create.*** Accidental creations do not count;
- ***Authorship rests with those who commission, and invest money in, a work*** (common law tradition).

(Olsson 2009:5, adapted from Ginsburg 2003:11-27).

From the above examples, it can be observed that the law acknowledges most of the modern concepts of authorship. However, as it can be observed in the third point above, the Romantic concept is still present. Thus as Bracha notes (2006:2-8) the law still oscillates between the “valorization of the individual genius and a painful realization of the much more complex realities authorship.” Thus, when necessary, the courts of law often embrace the metaphor of individual creation *ex-nihilo* for purely practical reasons.

Furthermore, it can be observed from the examples above that the concepts of author and authorship did not appear in one historical event, but they were developed through time. This development was hardly radical, and new theories have always built upon the previous ones. This is best exemplified by the examples above, where in contemporary views on authorship one can find the relics of the past, namely the romantic myth of the author as the creator and originator of a piece of work. In all but one of the points above, authorship is seen as original creation and it rests with the creator. However, in the last point, the concept is compared with physical property, like a house or a car, and in that sense authorship rests with the owner, buyer, or commissioner of the creative work.

To conclude, the aim of this chapter was to demonstrate the complexity of authorship concept by presenting the history of its development. From it one can even observe the intertextuality in the process of the development of the concept, since the development of the concept was always based on previous assumptions. Furthermore, this short historical overview was to demonstrate the sources of our contemporary conceptions and misconceptions on authorship. Thus the aim of the chapter was to demonstrate the rationale for a study into authorship of translations, which is still one of the major issues in many authorship studies.

Authorship of translations

Thus far, there have been many studies (see Hermans 1996; Schiavi 1996; Boase-Beier 2006; Venuti 1995, 1998; and others) on authorship and/or style of translations. However, most of these studies focussed mainly on the translator's impact on the original text, and not as much on the issue of their authorship, since they were primarily done in the scope of translation, and not authorship studies. Nevertheless, the works of translation theorist like Theo Hermans, Giuliana Schiavi, Lawrence Venuti and others, offered some views on the process of production of translation and translator's influence on the final product of translation, which are closely related to authorship. Furthermore, Lawrence Venuti (1998) also presented a short historical overview on the status of translations in terms of copyright and authorship, focussing on the intellectual rights of the translator vis-à-vis the author.

Focussing on past court rulings, Venuti (1998) aims to demonstrate the legal status of translations before the Romantic myth of the author started to dominate the law. He (Venuti, 1998:56) reports that, before the Copyright Act of 1911, the author was only given right to reproduce, or copy, his work, whereas he had no rights over works derived from his own. Additionally, translation was not considered as derivative, but original, since it resulted from translator's labour (ibid). Venuti (ibid) supports this claim quoting a court ruling (Wyatt vs. Bernard, 1814) which states that: "Translations, if original [...] could not be distinguished from other Works," and thus the translator or his employer have copyright on the translation, "unless that translation is a copied translated text."

Citing another case, Venuti (1998:56-57) reports that defendant's counsel in *Burnett vs. Chetwood* (1720) observed that the Statute of Anne protected only the form and not the content of the author's work, and thus excluded the translation from author's copyright. Specifically, the content, the ideas in the work, was considered as public knowledge upon publication. Thus the translation only disseminated ideas which were by then considered public knowledge. On the other hand, the translation due to the translator's skill in language had its own style and expressions, which made the translator owner of the text (ibid). Thus authorship was based on the form, i.e. the words used, as

neither the author nor the translator had any intellectual property rights on their ideas once the work was published.

That the translation is the work of the translator got its most extreme articulation in 1852, in a famous American case over unauthorised German translation of Harriet Beecher Stowe's *Uncle Tom's Cabin* (Venuti, 1998:57). In *Stowe vs. Thomas*, the judge decided that the translation did not infringe her copyright in the English text, adding that the same conceptions written in "another language cannot constitute the same composition," because the making of a good translation "often requires more learning, talent and judgment than was required to write the original" (ibid). The judge added that the control over translations would interfere with the circulation of Stowe's ideas, "thereby contradicting the constitutional view of authorial copyright as legal means "to promote the Progress of Science and useful Arts"" (ibid). However, Venuti (1998:58) reports that *Stowe vs. Thomas* never had the authority of a precedent, but rather considered an eccentric case in the history of copyright. Namely, about that period the Romantic concept of authorship was adopted in the law, thus "dooming translation to the ambiguous legal status it currently occupies" (ibid). This is best exemplified by the Copyright Act of 1911 which defined translations as mere copies by giving the author the exclusive right to produce or publish any translation of the work (ibid).

Venuti (1998) asserts that the presence of the Romantic concept of authorship in law diminishes the legal status of all derivative forms, and, as it denies intertextuality, at the same time conceals the fact that any underlying work is itself derivative. Namely, the Romantic concept does not acknowledge the fact that every work appropriates pre-existing cultural materials, whereas a translation appropriates the foreign text and its own domestic cultural materials (Venuti, 1998:60). However, Venuti (1998) does not discuss intertextuality in the Barthesian sense, i.e. that there are no real authors, but as part and parcel of the modern authorship concept. Thus, by focussing on the form rather than sense, a translation is just another form of intertextuality.

Furthermore, Venuti (1998:60) asserts that even though translations are bound to be more closely related to the underlying work than other derivative forms, the close relation does not imply that the two works are identical, or that the translation itself is not a work of authorship (ibid). Additionally, as both texts draw upon different cultural

materials and contexts, the significance of the original novel in its own language and literature will never be exactly equivalent to the significance of the translation designed for another language and literature (Venuti, 1998:61-62). According to Venuti (ibid) this explains why bestsellers in one language sometimes do not repeat their success when translated.

Although not directly related to authorship, Venuti (1995, 1998) proposed the concept of 'translator's invisibility' from which his views on authorship of translation may be inferred. Namely, Venuti (ibid) argued that in the Anglo-American translation tradition the focus was on creating translations that would look, or read, as original pieces of writing in the target language, thus making the translator invisible to the reader. On the other hand, the translator's visibility Venuti (ibid) described as the foreignness of the text, manifested by contextual references foreign to the reader, and it is the translator's own choice to decide on the visibility or invisibility of translating.

Due to such domesticating approach in Anglo-American translations, Venuti (ibid) attributes authorship to the translator. Although, Pym (1996:170-1) questions the claim that focus on fluency and domestication can be described as exclusively Anglo-American. Nevertheless, Venuti (ibid) describes the production of a translation as a derivative process since it imitates another text. Therefore, translating is a form of authorship, but authorship redefined as derivative, and not self-originating. However, Venuti (1998:63) rejects the theory of collaborative authorship of translations as it presupposes the author's and translator's intention to collaborate at the time of writing. Venuti (ibid) argues against collaborative authorship, or joint work, by pointing out the spatial and temporal distance between the author and the translator, and also emphasising the lack of shared intention. Namely, foreign authors address a certain readership which does not include the readers of translation, whereas translators address their domestic readership "whose demand for intelligibility in the terms of translating language and culture exceeds the foreign author's intention" (ibid). Thus Venuti (1998:64) asserts that the translation is inevitably different from the foreign text as it relies on target language approximations, adding, that even though the translation "is required to imitate the entire foreign text, their linguistic and cultural features are sufficiently distinct [...] to be considered as an autonomous work."

In his essay on the translator's discursive presence in the translated text, Theo Hermans (1996) compared translation with interpreting, and argued that, just like interpreters, translators write in someone else's name. Hermans (ibid) proposed that in translation, just as with interpreting, one can observe two 'voices'; one belonging to the original author, and the other to the translator, which may be more or less overt in the translated text. However, this raised the question whose voice gets to the reader upon reading a translation (Hermans, 1996:26). Hermans also raised the question of the style of translations:

“Does the translator, the manual labour done, disappear without textual trace, speaking entirely ‘under erasure’?” (1996:26)

However, Hermans did not apply stylistic analysis to detect the translator's presence in translation, but resorted to narratology to answer this question.

In an article written alongside with Hermans', Giuliana Schiavi (1996) argued that the notion of translator's voice had been neglected in translation theory. Also by resorting to narratology, Schiavi (ibid) described the process of production of translations as co-production of discourse between the translator and the original author. Basically, as the original is written for some implied readership, the translator must also pose the question of readership of the translation, since the implied reader of the translation is not the same as of the original (Schiavi 1996:15). Thus the translator takes the role of the implied reader, but is also aware, hopefully, what the author wanted the reader to be, and since the translator's task is to produce a text, this awareness will be expressed in the translation (ibid). Thus, the translator intercepts the communication, processes it, and transmits it to the new reader (ibid). In other words, as the translator is the reader of the text, in the process of the production of the translation, there is a kind of discourse between the original author and the translator, which is then transmitted to the reader of the translation.

Based on Schiavi's model of the narrative of translations, Hermans (1996:42) concluded that the presence of the translator in the discourse of translations is reasonable and necessary assumption. Hermans (1996) argued that the translator's voice is always present in a translation alongside the author's, thus also taking the view of co-production

of translations. However, he argued that sometimes this voice may be hidden behind the voice of the narrator, and thus on occasions impossible to be traced (ibid).

Even though Hermans and Schiavi (1996) did not provide any concrete linguistic evidence that would demonstrate the presence of the translator's voice within the discourse, at least not in terms of some linguistic features characteristic of the translation, it proposed and argued a view on the production of translations. Hermans and Schiavi (1996) propose a different view on authorship of translation from Venuti (1995, 1998); specifically, they view the process of translation as co-production, or co-authorship, between the source author and the translator.

Regarding the translator's presence in the translation, Mona Baker (2000) proposed a hypothesis that translating is creative work and thus a translator must leave a personal imprint on the new text. Similar view is taken by Boase-Beier (2006:1) who also agrees that apart from the style of the original text, in the 're-creative' process, the translator's individual style is bound to become a part of the translation. Baker (2000:244) accounted this with an analogy that:

“it is as impossible to produce a stretch of language in a totally impersonal way as it is to handle an object without leaving one's fingerprints on it.”

However, Baker (2000) did not provide any evidence of translator's style in translations, but rather discussed a potential approach for research on the style of translations. Namely, Baker (2000) proposed an analysis of the style of translations so as to identify linguistic features typical of the translation, in contrast to the linguistic features found in the original work. Bosseaux (2007:23), taking her cue from Baker (2000), in her analysis of translations of two novels by Virginia Woolf, also proposes that the translators' presence in translations can be uncovered by looking at their style, which is defined as characteristic language use typical of a translator. Baker (2000:245) defined style as an individual profile of linguistic habits, or linguistic fingerprint, and suggested that this fingerprint can be detected specifically for the author and the translator.

However, linguistic fingerprint is in reality an unhelpful metaphor, inasmuch as, unlike individual's fingerprints or DNA, language is not an innate property. One is not born with language, but acquires it through years, and thus the language of an individual

is not fixed, but subject to change, which makes the detection of one's linguistic fingerprint very difficult, if not impossible (Olsson, 2004:32).

Nevertheless, Baker raised several important questions hitherto unanswered either by translation or authorship studies, namely:

Is a translator's preference for specific linguistic options independent of the style of the original author?

Is it independent of the general preferences of the source language...?

(Baker, 2000:248)

From the above example it can be seen that Baker (2000) is aware of the issue that the style of the translation is probably determined not only by the author and the translator, but also by the source and target language, and source and target society. Therefore Baker (ibid) described the style of translation as the result of a combination of two 'authors', two languages and two sociolects. Baker (ibid) is also aware of the lack of methodology that would be able to successfully isolate the translator's linguistic features from those which reflect the original author. However, Baker (ibid) proposed a method which can be used as a guideline for further research on translator's presence in translations, viz. the application of stylistics.

From the above mentioned theories, on one hand, one can conclude that translator's influence on the final product of translation is reasonable and generally well-accepted concept. On the other hand, there is still much debate over the authorship of translations, with authorship being attributed solely to the translator or both to the translator and the author. However, none of these theories have provided the answer to questions of in what linguistic forms does this voice or visibility of the translator manifest in the final product of translation, or how one can observe and measure it, and most importantly they have not provided any concrete linguistic evidence of its existence at all. Therefore, it is necessary to start a search for evidence of this voice, or propose another theory which could be supported by linguistic evidence.

Style and Authorship

According to Olsson (2008:28), the word 'style' is probably one of the most widely abused words in linguistics. It can be used to describe a set of mannerisms which fulfil a certain communicative need, or to describe the individual's marked way of using language, or idiolect, as opposed to the language of society, or sociolect. On the other hand, sociolect is language developed due to geographical or social separation of one group of language users from another, and it is usually described vis-à-vis the standard variety of language (McMenamin, 2002:26). Since we are a part of society, our idiolect is always a part of our sociolect but sociolect is not simply a large collection of individual idiolects, but a synthesis of shared elements (McMenamin, 2002:53). Furthermore, there is a mutual dependency between idiolect and sociolect; one changes as the other does (Olsson, 2004:34). This means style, as a marked use of language, is not an exclusive property of an individual, but "a complex, fluid interaction of the individual, language and society" (Olsson, *ibid*).

What is nowadays referred as idiolect, sociolect, and language in general, was already noticed by Coleridge in 1817:

"Every man's language has, first, its individualities; second, the common properties of class to which it belongs; and third, words and phrases of universal use."

(Coleridge, 1817:53, cited in McMenamin 2002:52)

Style is described by Gläser (1998: 131) as "the choice of linguistic means made by the *individual* [my emphasis] speaker or writer in accordance with the requirements of the communicative situation." The choice of these "linguistic means" is often influenced or even determined by external conditions, or "communicative situations", and therefore, we make conscious, subconscious, and unconscious choices which result in habitually using one word instead of another (McMenamin 2002: 164). These habits become predictable, and thus, what we say or write sounds or looks like something we would be likely to say or write, and our language becomes associated to us, as the way we speak, as our own *style*. Therefore, even though we can use any word, we make typical and individuating selections choosing the words we prefer.

Since style is often described as the individual's marked way of using language as opposed to language of society, it is easily related with authorship. As already mentioned, the habit of choosing some linguistic means rather than some others defines the style of an individual; thus, if recognised, these choices become style-markers. Thus, as every author has a style that can be recognised, one can recognise the author of a piece of writing by recognising his or her style, namely by applying forensic stylistics.

Forensic stylistics is the scientific study of patterns of variation in written language. The object of the study is the language of a single individual with the aim to describe his or her identifying linguistic characteristics (McMenamin, 2002:163). These patterns, or style-markers, are present in unique combination in the style of every writer, and can be usually described and measured, thus making author identification possible.

However, it is highly unlikely that one style-marker could be used on all occasions, discriminating between all users of language, or even successfully identifying one author's idiolect on several occasions. For example, Shakespeare can be distinguished from Dickens by word-length distributions. But the two authors are also separated by two centuries long time span, and the marker may not necessarily discriminate between authors who are socio-linguistically closer, e.g. Shakespeare and Marlowe (Grant and Baker, 2001:70-71).

In practice, forensic linguists do not attempt the difficult task of identifying reliable and valid markers of authorship, they rather test a range of statistics to identify discriminates for that particular case (Grant and Baker, 2001:76). Nevertheless, Grant has tested 170 markers proposed by others and found out that most of them are indeed case-specific (Coulthard, 2006:17)

Nevertheless, there are some measures, such as type-token ratio, average word-length and average sentence length, which can often be used, mostly with longer texts, to determine if two or more texts are of the same authorship.

Type-token ratio (TTR) is a measure of lexical density and is calculated by dividing the total number of different words (types) by the total number of used words (tokens) in a text. Since TTR changes with the size of the sample, when comparing samples of different sizes it has to be standardised, so as to make the results comparable. Thus, Standardised TTR (sTTR) is calculated for every 1000 tokens and the average value

is taken as the ratio for the whole sample (Olohan 2004). Testing the texts for type-token ratio is also a commonly-used method in the detection of plagiarism (see Chaski 2001). Usually, it is used to detect sudden changes in style, which are attributed to copying several passages from another author.

The percentage of overlapping vocabulary is a method specifically applied in plagiarism detection, and is calculated by dividing the number of lexical, or content, words shared between the samples with the total number of lexical words in the sample. The method was developed by Alison Johnson (1997), who noticed that some of the essays suspect of plagiarism shared a considerably higher amount of vocabulary than non-plagiarised texts. Even though the percentage of shared vocabulary is not normally used as a style-marker, but rather as an indicator of plagiarism, when applied on two translations it might demonstrate the freedom each translator had with the choice of vocabulary when translating.

For the sake of clarity, it is necessary to point out that the very concept of plagiarism is based on the concept of authorship, and plagiarism detection undoubtedly overlaps, in several aspects, with authorship identification (Grant and Baker 2001).

The main advantage with TTR and overlapping vocabulary is that they are quantifiable data, whereas some other language characteristics are sometimes very difficult to identify as countable units; and sometimes counting certain features of language makes little sense and does not seem to relate to their identifying importance (McMenamin, 2002:138). Quantitative analysis is important for at least two reasons: it makes decision making and hypothesis testing more precise, and it meets methodological and judicial requirements for scientific evidence (*ibid*). Thus, in cases of authorship identification, quantification is often necessary for the application of statistical tests which evaluate the significance of the relationship of the variables (style-markers) across comparison writings (*ibid*).

However, the issue with the quantitative analysis is to determine the sufficient sample size. There are many studies demonstrating various sample sizes (see McMenamin, 1993), but we can say that the language sample is sufficient when stylistic variables occur frequently enough to establish patterns of variation (McMenamin, 2002:171). Olsson (2004:63-6) demonstrated the significance of sample size, with his

experiment to identify the true authorship of a text, suspected to be made by Trollope. His test on a 400 word excerpt identified Trollope as the author of the unknown writing with the level of probability of 23%, while the 1000 word excerpt produced the level of probability of 97%, and thus successfully attributed the text to Trollope. Therefore, one can assume that the larger the sample, the more reliable the results of both qualitative and quantitative analysis will be.

There are various statistical tests which are applied in forensic linguistics. For example, for evaluation of potential relationship among variables in comparison writings, several statistical tests are used, such as t-test, chi-square, z-score, ANOVA, correlation tests, etc.

The statistical test used in this research is Spearman's rank correlation coefficient which is normally used with non-continuous variables, namely, ordinal variables, which can be ranked. For example, if comparing TTR values extracted for each chapter from two translations of the same underlying work, TTR values from each chapter are ranked according to the highest value. If the chapters are ranked in the same order in both translations, then they are correlated. With a larger number of pairs for comparison, this coefficient tells us if the variables from the compared samples are correlated and it also tells us the significance of the correlation (Oakes, 1998). As an alternative Pearson's correlation coefficient can be used, since the result from both tests are usually similar.

A Case of Plagiarism

As mentioned in the introduction, this research has been inspired by a case of plagiarism of a literary translation. Turell (2004) reported the first case of a linguistic analysis of a plagiarised translation. Namely, in 1993 the Supreme court in Madrid, Spain, found Vázquez Montalbán, a fairly well-known Catalan writer, guilty of having plagiarised Pujante's translation of Shakespeare's *Julius Caesar* (ibid).

From the legal point of view, namely the Spanish Law of Intellectual Property, a translation of a literary work is considered itself a literary work. Specifically, a literary translation involves a transformation of a prior piece of work and is thus subject to intellectual property (ibid). However, this intellectual property does not include the title, the plot, characters or proper names, since these are derived from the underlying work, and not created by the translator (ibid). Thus the translator is given rights only over the form and not the content.

From this it can be observed that the Spanish Law of Intellectual property partly coincides with Venuti's (1995, 1998) view on authorship of translation, that is translating is seen as a derivative form of authorship, and the translator's rights are based on the form of their translation. Furthermore, the presence of the Romantic concept of authorship can also be detected in the law, inasmuch as the translator cannot be given the right over the title, the plot and the characters since they are original to, and created by the author.

That the 'originality' of translations is based on the form rather than the content can be observed from the court ruling of that case, which states that in longer translations there is much more space for originality than in the short ones (Turell 2004:2). Thus clearly this 'originality' must reside in the form, since all translations of the same underlying work are bound to communicate the same ideas. Interestingly, the authorship of translations resides in the form, whereas the authorship of the underlying work resides in its ideas, thus giving the author the right over translations. Had the authorship of these two forms been discussed on equal terms, translations would have been considered either as mere copies of the original, with the same ideas reproduced in another language, or as original literary works since their form would be absolutely different from the underlying work.

The practical aim of Turell's (2004) study was to attempt a linguistic analysis of a plagiarised translation and also to test an analytical tool for detecting plagiarism, CopyCatch. However, the study also tried address the question of what makes plagiarism between translations different from other types of plagiarism and thus more difficult to detect.

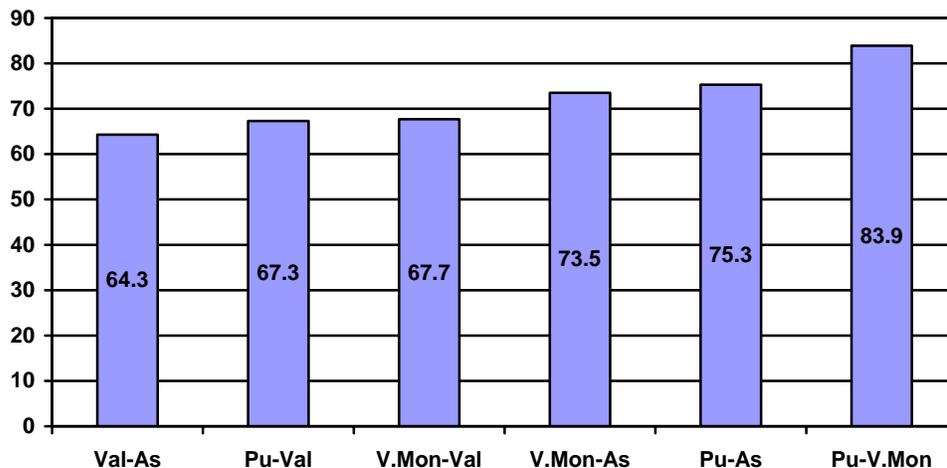
In her introduction, Turell (2004) reports that in terms of a linguistic enquiry, namely plagiarism detection, translations are very difficult to deal with for two reasons: first, they all tend to reflect the original author's form and content, and second, the more faithful to the original they are, the more difficult is to establish their originality.

The first task for Turell (ibid) was to establish the degree of similarity between translations of the same text, and the point at which this similarity becomes suspicious. To do so, Turell (ibid) looked at the percentage of overlapping vocabulary between the two texts, which is a reliable measure commonly used in plagiarism detection. Specifically, Turell (ibid) used CopyCatch software to calculate the overlapping vocabulary, but made a stop list which excluded all Spanish function words and patronymic and toponymic expressions from her analysis.

When testing text for plagiarism, for example, comparing two student essays, the more similar the set of linguistic items, the greater the likelihood that one of the texts was derived, at least in part, from the other (or, of course, that both were derived from a third text), rather than composed independently (Coulthard: 2005:17). One of most common methods for detecting plagiarism is to determine the percentage of shared lexical items.

In practice, if two texts share more than 50% vocabulary, it is suspected that they may not have been produced independently. However, since translations are inherently more similar to one other, Turell (ibid) raised the threshold to 70%. Afterwards, the disputed texts were compared, along with two other non-disputed translations written by Valverde and Astrana. The results of Turell's analysis may be observed in Figure 1.

Figure 1 - Percentages of overlapping vocabulary between the pairs of translations



It can be seen from the above figure that all of the compared translations show a high percentage of overlapping vocabulary, well above the 50% threshold, and half of the pairs are even above the 70% threshold. Even though the percentage of overlapping vocabulary was highest in the disputed pair, the other pairs also shared a considerable percentage of vocabulary. Thus this method failed to distinguish between plagiarised and non-plagiarised translations conclusively. Furthermore, it is unclear how Turell managed to obtain only one percentage, since if the texts are of different sizes, overlapping vocabulary would take up a different percentage of each of the paired texts.

Thereafter, Turell (2004) continued with her study to conclusively demonstrate plagiarism in the case, which is not the topic of this research, and thus the other methods will not be discussed here. However, the results of the analysis reported here are very important for the discussion on authorship of translations.

Even though at first glance intriguing, the high percentages of overlapping vocabulary between the non-disputed translations are not that surprising. Specifically, the high percentage of overlapping vocabulary indicates that the compared text have been either derived from one another or from another source. With translations this is absolutely true, since all of them have been derived from their underlying work, in this case from the ‘original’ English version of Julius Caesar.

However, the question posited by Turell (2004) of what makes translations so similar is only partly answered by this, since the question of the nature of authorship of translations remains unanswered. Specifically, if authorship of translations is based on the form, how come the form is so similar between them, and can derivative authorship account completely for such similarities?

Assuming the derivative nature of the authorship of translations, translations would be on a par with adaptations. Thus Shakespeare's *Antony and Cleopatra* would be on a par with Thomas North's translation of Plutarch's *Life of Markus Antonius*. However, Shakespeare's work is considered an adaptation, with its authorship based on the sense, whereas Thomas North's work is considered a translation, with its authorship based on the form. Interestingly, even though Shakespeare derived his work from Thomas North's translations (Barnstone 1993:87) and sometimes copied several passages from it, his work still regarded as original and creative, whereas Thomas North's work is seen 'merely' as a translation. Thus one is right to question whether the theory of derivative authorship of translations can account for the similarities between them.

To conclude, the aim of this chapter was to present the application of authorship theories of translations in practice together with all the implications these theories carry. The case of plagiarism was used to demonstrate the practical problems that arise even when applying tested methods due to the application of a possibly misconceived theoretical framework. Therefore it is necessary to revise the theory of the authorship of translations, and propose one that can be empirically demonstrated.

Two-dimensional Theory of the Style of Translations

The aim of this chapter is to propose a theory on authorship of translations that could be demonstrated empirically by a linguistic analysis. Since authorship is inherently related to style, the research will discuss the style of translations, which can be linguistically tested, rather than theories on the processes of production of translations.

Looking at a translation one can observe its style, just like with any other form of authorship. However, if one is to assume the co-authorship of translations, as an alternative to derivative authorship, one has to regard the style of translations as the blend of the author's and the translator's style. Therefore the style of a translation consists of linguistic features belonging both to the author and the translator. Thus, even though only one style is being observed, this style has two dimensions: one of the author, and the other of the translator.

The idea behind the two-dimensional approach to the style of translations is that all translations are co-authored, or as Boase-Beier (2006:41) said, the product of translation “is a text which *fuses*¹ the original author with the author of translation.” As mentioned before, Hermans (1996) and Schiavi (1996) take a similar view on the authorship of translations and have described the process of production of translations as co-production between the translator and the author. Barnstone (1993) also considers the translation as a collaboration, and even calls it a ‘double art’ since the translation is in actuality the work of two artists. Specifically, he describes the translation as “a collaboration under suspect nameship, a splendid forgery (and thoroughly forged as one) of the art of at least two creators in one new language” (Barnstone 1993:87). Consequently, this view is opposed to Venuti's (1998) view of the translation as a form of derivative authorship authored solely by the translator.

Therefore if translations are considered as co-authored, then the author of a translation is the team that has ultimately produced it (Barnstone 1993:13), e.g. Shakespeare-Pujante as the author(s) that produced the plagiarised translation of *Julius Caesar* in the above mentioned case. Thus, for example, when comparing two translations

¹ my emphasis

of Shakespeare's *Julius Caesar*, one is comparing the styles of Shakespeare-Pujante and Shakespeare-Astrana which are bound to be similar as Shakespeare will always resemble Shakespeare.

The style of the translation is unitary; however, it is defined by the author and the translator. If one changes so does the other. Hence, on the one hand, the styles in the translations of the same underlying work by different translators will always be different. On the other hand, the style of the translator will change with each translation of another work, as he or she will collaborate with different authors.

Even though the two-dimensional style of translations is a reasonable assumption, supported by theoretical works of Barnstone (1993), Boase-Beier (2006), Hermans (1996) and Schiavi (1996), to demonstrate these dimensions one has to compare two translations of the same underlying work. Thus, the two-dimensional theory is a result of a preliminary stylistic analysis of two translations of Plato's *Republic* and only supported by the above mentioned theories.

Stylistic analysis of translations was first suggested by Baker (2000:245) who proposed an analysis of several translations by the same translator in order to "capture the translator's characteristic use of language." Baker (2000) was aware of the lack of a method that would be able to differentiate the translator's style from the style of the original. Nevertheless, Baker (2000) tested two sets of translations by two different translators for type-token ratio (TTR) and noticed a general difference between them, specifically, one translator having a higher TTR than the other. Furthermore, Baker (ibid) realised the problem of attributing the observed stylistic differences, since there were two 'authors', two languages, and two sociolects involved in the styles of each translation. Thus in her conclusion, Baker (ibid) questioned whether it would be more useful to analyse different translations of the same underlying text in order to keep the variables of the author and source language and sociolect constant.

Furthermore, when comparing translations of the same underlying work made by different translators, one can observe, on the one hand, the similarities or the consistencies in these translations, which constitute the translator's dimension of style, and, on the other hand, the differences that result from the translators' own styles, which constitute the translator's dimension.

Taking by cue from Baker's research, I have decided to test two translations of 'The Republic' by Plato, one written by Benjamin Jowett and the other by Robin Waterfield, for their TTR values. However, as the compared texts were of different size, i.e. Jowett's translation consisted of about 110,000 words, whereas Waterfield's translation consisted of about 130,000, the texts were compared for Standardised TTR (sTTR) values, since TTR depends on the sample size. Furthermore, sTTR values were not calculated for the whole texts, as proposed by Baker (2000), but rather throughout the texts so that the language change within translations may be observed.

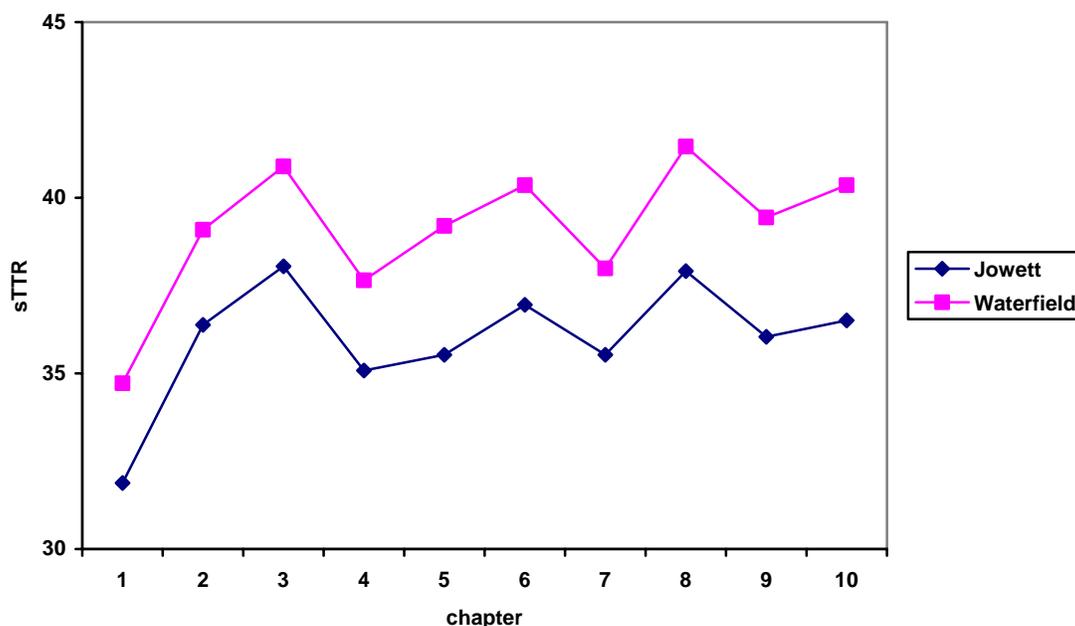
The first task of the preliminary research was to clear the translations of any meta-language, that is, the translator's own language in form of comments, footnotes or introduction to the chapters. This was necessary because these elements are not in the scope of this research, since they were not a part of the underlying of work, and thus do not represent the style of the translation.

The initial idea was to compare the translations chapter by chapter. However, this was not possible since there was a discrepancy in the number of chapters between the two translations. Specifically, Jowett's translation consisted of 10 chapters, just like the underlying Plato's work, whereas Waterfield's translation was divided into 14 chapters. Thus it was necessary to make the two translations comparable.

To do so, it was necessary to divide each translation into an equal number of equal segments. Specifically, each translation was divided into 10 equal segments each representing 10% of the whole text. This meant that size of the segment was determined by the length of the translations, and that the corresponding segments from the compared translations were not of the same size, due to the difference in length between these translations. Nevertheless, each of the segments in both translations roughly represented 10% of the translation, and consequently 10% of the underlying work, which made them comparable. For this purpose a *Perl* script which automatically divided each translation into 10 equal segments was made (see Appendix 3).

Having divided each text into chapters, it was possible to calculate standardised type-token ratio (sTTR) for each segment and compare it with the corresponding one from the other translator. To calculate sTTR, Wordsmith Tools corpus access program was used, since, as Olohan (2004:80) pointed out, this program automatically calculates sTTR.

Figure 2 - Standardised TTR values through segments



Just as in Baker's (2000) research, sTTR values were different between the two translators, with one translator showing higher values in each of the compared segments. Not surprisingly, sTTR values varied considerably between the segments. However, the variation of sTTR in both translations followed the same pattern (see Figure 2). This pattern indicated that the values of sTTR between the translations were correlated, which was possible to be demonstrated statistically.

Specifically, to demonstrate the correlation between sTTR values from one translation with the values from the other, Spearman's rank correlation coefficient was calculated. This coefficient is normally used with ordinal variables, such as sTTR values in this case, and it indicates if the two variables are correlated, namely by comparing if the values are ranked in the same order. Furthermore, this coefficient provides us with the significance of the correlation (Oakes, 1998). The coefficient was calculated using SPSS statistical analysis program. Namely, the data obtained by using Wordsmith Tools were used as input for SPSS which calculated the coefficient automatically.

Table 1 - Correlation between sTTR values of the two translations

			sTTR_Waterfield	sTTR_Jowett
Spearman's rho	sTTR_Waterfield	Correlation Coefficient	1.000	.939(**)
		Sig. (2-tailed)	.	.000
		N	10	10
	sTTR_Jowett	Correlation Coefficient	.939(**)	1.000
		Sig. (2-tailed)	.000	.
		N	10	10

**** Correlation is significant at the 0.001 level (2-tailed).**

It is necessary to point out that the significance results showed in Table 1 are not equal to nil (.000), but just outside the three decimal span of the computer program that was used for the analysis, viz. SPSS. Actually, the significance of the correlation was shown at the 0.001 level, which meant that there was less than 1/1000 chance that the correlation of sTTR values between the segments was coincidental.

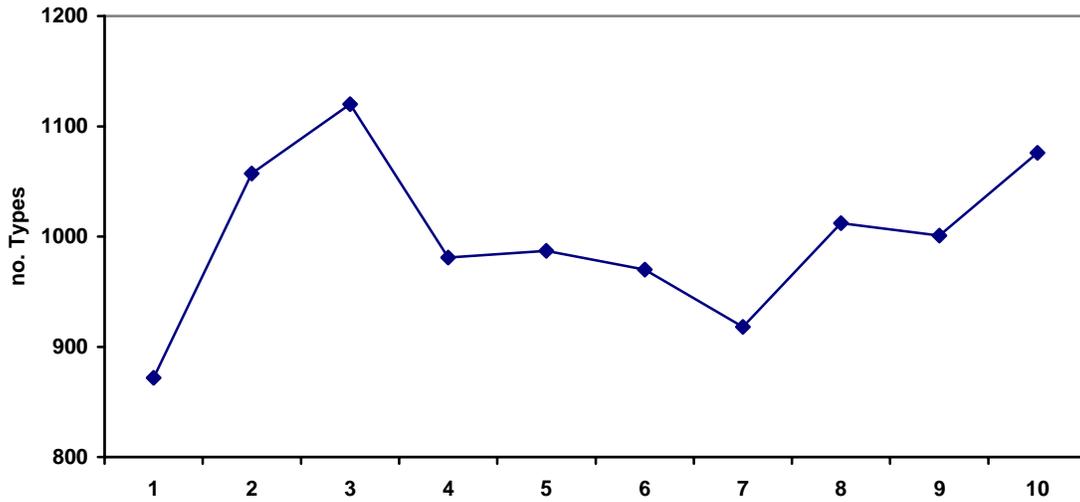
Thus even though sTTR values were different between the translations these values were correlated. From this one may conclude that the sTTR values were a result of the original author's and the translators' styles. Namely, the difference between the values from each translation can be attributed to the difference in style between the translators, whereas the pattern in which they change was determined by the original author's style.

With the idea of two-dimensional style in mind, I decided to address the issue of shared vocabulary between the translations. Specifically, if sTTR values, which inherently demonstrate the vocabulary of a text, were partially determined by the original author, then the percentage of shared vocabulary might also be partially determined by the author. Thus, in segments where both translators showed higher sTTR values one would expect the percentage of shared vocabulary to be higher as well.

Shared vocabulary between the two translations was calculated by using CopyCatch Gold² program. Specifically, by using CopyCatch, raw frequencies of shared vocabulary were extracted rather than percentages. The reason being is that percentage of shared vocabulary is different for each of the texts since they are of different sizes. Thus the number of types occurring in both translations was calculated for each segment. From Figure 3, one can observe that the number of shared types per segment changed in a similar pattern as sTTR values, but still somewhat different.

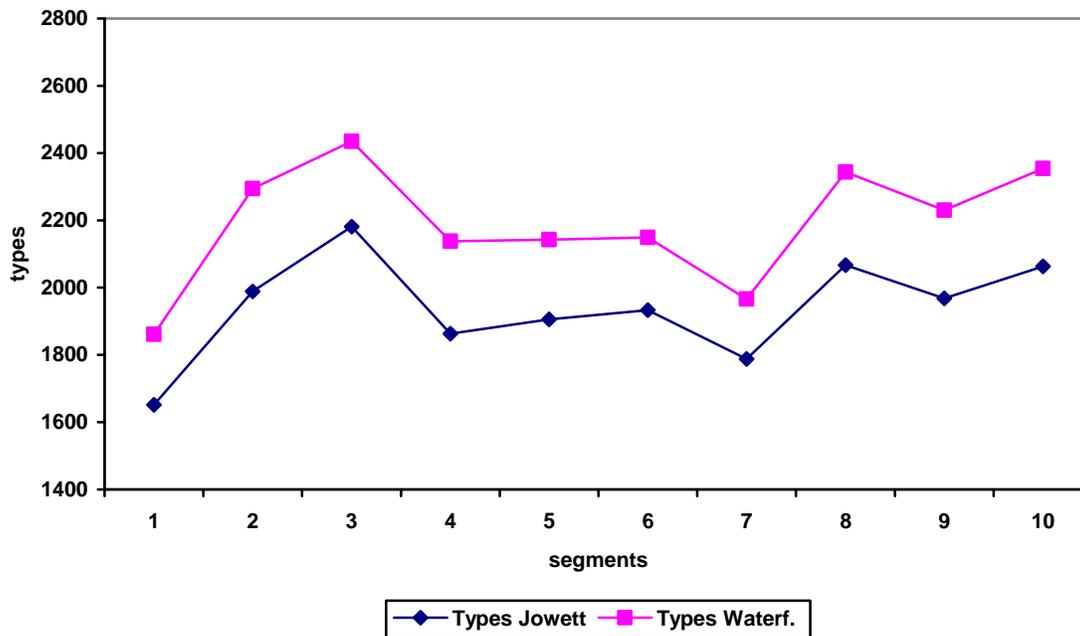
² a free version of the program can be found at www.cebe.heacademy.ac.uk/learning/Plagiarism/index.php

Figure 3 - Number of shared types through segments



However, the number of shared types was calculated on whole segments, i.e. it was not standardised on 1000 word samples like sTTR was. Thus to observe the correlation between two more comparable values, it was necessary to compare the number of shared types, with the total number of types used by each translator in each segment. The number of types for each segment was, just like sTTR, calculated with Wordsmith Tools, and the results are presented in Figure 4.

Figure 4 - Number of types used in each translation through segments



As it can be observed from the above figure, the number of shared types per segment followed pretty much the same pattern as the total number of types used by both translators. However, for the sake of clarity, correlation between all three variables, i.e. the number of shared types and the number of types used by each translator, will be calculated. Since the correlation between the translators has already been demonstrated by correlating their TTR values, the results from both translators are used in this analysis primarily to correlate the number of shared types with both translators, not just one of them.

Table 2 - Correlation between the number of shared types and the total number of types

			Types Jowett	Types Waterfield	Shared Types
Spearman's rho	Types Jowett	Correlation Coefficient	1.000	.988(**)	.927(**)
		Sig. (2-tailed)	.	.000	.000
		N	10	10	10
	Types Waterfield	Correlation Coefficient	.988(**)	1.000	.952(**)
		Sig. (2-tailed)	.000	.	.000
		N	10	10	10
	Shared Types	Correlation Coefficient	.927(**)	.952(**)	1.000
		Sig. (2-tailed)	.000	.000	.
		N	10	10	10

**** Correlation is significant at the 0.01 level (2-tailed).**

It can be observed from Table 2 that the number of shared types is correlated significantly with the number of types used by each translator. Again the significance levels seem to be nil (.000), but as already mentioned the significance level is just out of the decimal span of SPSS, and the significance is actually shown at the 0.001 level.

This correlation supports the two-dimensional theory in the sense that it demonstrates that the choice of vocabulary in a translation, and hence the shared vocabulary between translations, is only partly determined by the translator and is limited by the author.

Specifically, in this case, the shared vocabulary between the two translations is the result of the comparison of the styles Plato-Jowett and Plato-Waterfield. Thus, what is in common with these two translations is that they share one co-author, i.e. Plato, and a certain amount of vocabulary. On the other hand, what they do not share is the rest of the

vocabulary and the other co-author. Thus it is reasonable to attribute the similarities to the shared co-author, Plato, and the differences to the translators as the other co-authors.

This correlation between sTTR values, and between shared vocabulary and vocabulary used by the translators, demonstrates that the vocabulary of translation is predetermined partly by the original author and that the similarity between translations is indeed a result of two texts being of common authorship. This accounts for the issue raised by Turell (2004), who wondered what makes translation so similar and thus more difficult to detect plagiarism between them. On the other hand, the lack of this correlation might indicate that the vocabulary in one of the translations, and hence the percentage of shared vocabulary between them, is not determined by the original author, which would then indicate textual 'collusion'. However, inasmuch as this study was not carried out on translations which were known to be plagiarised this hypothesis could not have been empirically tested.

At the end of this chapter it is important to notice that there were several caveats in this research that have to be addressed, with the most important being the issue of segmenting of the translations. Namely, as the translations were of different sizes, segmenting was done by dividing the in terms of percentages so that each segment represented 10% of the translation. However, even though one translator generally used more vocabulary than the other, this does not mean that the difference in vocabulary richness between the translators was constant throughout the translations, since, in some parts of the text, the translators were given more freedom to elaborate than in some others. Thus the first and the last segments were most comparable since they represented the first or the last 10% of each translation. However, this did not mean that the segments were 100% comparable, since, e.g. one translator within his 10% of the vocabulary may have 'reported' 12% of the content of the underlying work, whereas the other may have 'reported' 9% of the content. Since the following segments started at the end of the previous one, this issue was exacerbated about the middle of each translation. Thus it was possible that two compared segments, each representing 10% vocabulary, were only comparable in 7-8% of the vocabulary with the rest distorting the results of the analysis. In this case, the issue was not as significant since the segments were large in terms of the number of words so that the incomparable parts did not alter the results significantly,

which was demonstrated with high levels of correlation of sTTR values between the two translations. Nevertheless, since the segments had to be large to ensure comparability, the more subtle changes in the vocabulary could not have been detected with this method.

Furthermore, the size of the segments limited the total number of segment that could have been compared. Thus, as segments had to be large to ensure compatibility, the two translations had to be divided into 10 segments each. This meant only 10 pairs of values could have been compared, which is practically the bare minimum for a statistical analysis. Thus the correlation results were also more difficult to obtain, and also less reliable.

To conclude, the preliminary research shows that through one style-marker, in translations, one can observe two styles, or more accurately, two dimensions of style. This indicates that the style of translations is, as Boase-Beier (2006:41) said, a fusion of the author's and the translator's styles. Consequently, if both the author and the translator are responsible for this style, it is then right to postulate that translations are co-authored, or co-produced, as Hermans (1996) and Schiavi (1996) proposed. However, the methodology in the preliminary research had some caveats that have to be readdressed, so that the two dimensions of style would be demonstrated with more conclusive empirical evidence.

Methodology

As mentioned previously, to demonstrate the two dimensions of style of translations it is necessary to compare the two texts step-by-step rather than as a whole. This was done by dividing the each text an equal number of segments which would be comparable. However, due to discrepancies in the size of the compared texts, the corresponding segments were firstly of different sizes, and secondly not absolutely comparable. To account for these issues the texts had to be divided into fairly large segments, which resulted with the method being fairly crude and unable to detect more subtle changes in style. However, if the segment size was reduced some of the segments might have become non-comparable. Thus it was necessary to develop a more sensitive method which could detect more subtle changes in the style of a text, but also ensure the texts were comparable for analysis. Specifically, the problem was to find a way to increase the sensitivity of the method and ensure comparability at the same time.

As it is easier to do a stylistic analysis on a large language sample, it was decided that the size of the language sample, namely 10% of the text, be kept. Consequently, it was necessary to improve both the compatibility of segments and the sensitivity of the method by changing the method of segmenting the text. Therefore, instead of splitting the text into 10 'pieces' and analysing each $1/10^{\text{th}}$ of the text for its style, I have decided to conduct an ensuing analysis of the translations by analysing the texts through several 'steps' using equal language samples. Namely, the language samples were analysed after every n^{th} word, instead of conducting each subsequent analysis where the previous one ended. Thus, for example, if the language sample is 10% of the text and the analysis is conducted in 50 steps, then a language sample of $X/10$ words is analysed after every $X/50^{\text{th}}$ word, with 'X' being the total number of words in a text. In this way, a text is analysed 50 times, or after every 2% of language, instead of just 10 times, as in the preliminary research.

In this way, the method effectively shows the style of the text for every 2% of its total vocabulary, even though 10% samples are used. Namely, the size of the language sample only ensures that there is enough data for a stylistic analysis. By analysing the text

after every 2% of the vocabulary the sensitivity of the method to detect more subtle changes in the style of the analysed language is increased.

Furthermore, inasmuch as the method is more sensitive to detect sudden changes in style, it demonstrates the style more accurately. Consequently, the comparability between the two translations is increased since the style is demonstrated more accurately. Additionally, due to the larger number of steps, the correlation test is also more reliable.

Since the language samples are still based on percentages of the total vocabulary, the language samples will vary in size between the two translations. Therefore, to demonstrate the two dimensions of style, this method, just like the preliminary one, uses Standardised Type-Token Ratio (sTTR) which, apart from demonstrating the style of the translation, is to account for the difference in the size of the samples between the two translations. Once calculated, sTTR values are analysed statistically so that the correlation of sTTR values between the two translations is demonstrated.

However, even though the difference between the language samples of the two translations would normally necessitate the application of Standardised TTR, the standardisation may have an adverse effect on the accuracy of these results. Namely, the aim of the method is to make it more sensitive to detect more subtle and sudden changes in style, and if these changes are standardised with the rest of the language sample, they will be more difficult to demonstrate. On the other hand, it is true that due to the difference in sample sizes the values of TTR between the two translations will not demonstrate the difference in vocabulary richness as much as sTTR would. Therefore, the texts will also be compared for TTR values.

In similar fashion as with the preliminary research, the shared vocabulary between the two translations will be calculated. Namely, the number of types occurring in both of the compared segments will be calculated. This procedure is also necessary to be recalculated using the new methodology since the method ensures better comparability of segments. For example, when comparing the first pair of segments, towards the end of the segment of one translation, there might be some words which are, by the initial method of segmenting, found in the second segment of the other translation, thus making the corresponding segments less comparable. By moving the next step of comparison by only 2% of the text, 8% of the text is being reused in the comparison; thus the words that did

not have a counterpart to be compared with in the first segment will find it in the second one, and so on. In this way all the shared types will be detected.

Again as in the preliminary research, the aim is to correlate the number of shared types with the total number of types used by each translator, so as to demonstrate that the shared vocabulary between translations is the result of the author's dimension of style, unless of course they are plagiarised.

The research will also attempt to demonstrate solely the author's dimension of style by neutralising the translator's dimension, which accounts for the differences between the styles of the two translations. Thus, if the translator's dimension is neutralised, the style in both translations would represent solely the author's dimension, and the style in both translations would effectively be the same.

To do so, the style in each of the segments will be observed as relative to the style of the whole translation, thus demonstrating a style independent of the stylistic features of the translator.

Again the focus will be on the vocabulary of the translations, with the idea to calculate the relative vocabulary richness of both translations. This will be achieved by calculating the proportion of segment vocabulary and total vocabulary of the translation. Specifically, the number of types in a segment will be divided by the total number of types used by that translator. Thus the vocabulary richness of each segment would be independent of that translator's general vocabulary richness. I have provisionally named this measure '*relative vocabulary richness*' (RVR). If the values of relative vocabulary richness from both sets of translations are similar and correlated, a common style between the translations, namely the author's dimension of style, would be demonstrated.

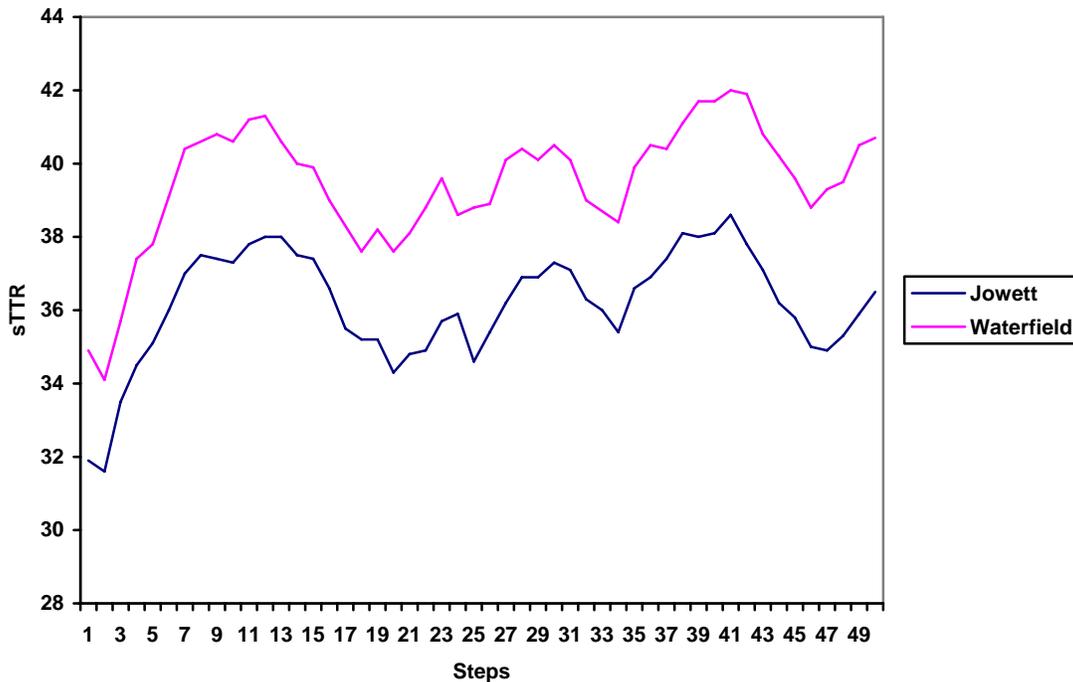
In all analyses, just like as in the preliminary research, the correlation between the two sets of values will be determined by Spearman's correlation coefficient, which will be calculated by using SPSS statistical analysis program. The segmentation and the step by step analysis of the translations with the extraction of style-markers will be done using a program specially designed for this purpose. However, as I do not have the intellectual property rights over the program, it cannot be included with this research.

Research Analysis

Standardised type-token ratio

The first task of the analysis was to compare the two translations for their sTTR values, so as to observe if these values are indeed correlated as the preliminary research suggested. Before conducting a statistical analysis, the results were put into a graph to observe if there is any pattern which would need to be confirmed by statistics. The results can be observed in Figure 5 (for a list of sTTR values of both translations, see Table 1 in Appendix 1).

Figure 5 - sTTR values through steps



As it can be observed from the above figure, sTTR values, although different in each translation, followed a similar pattern, just as with the preliminary research. However, even though the values followed a similar pattern, by taking a closer look at the above figure, it can be observed that the changes in values did not always occur in the corresponding segments. Nevertheless, the new method ensured that these changes were detected in the following step of the analysis. Although a general pattern can be observed

from the above figure, the results had to be analysed statistically to determine whether they are correlated, and if so whether that correlation is significant. The values of Spearman's correlation coefficient can be observed in Table 3.

Table 3 - Correlation between sTTR values

			sTTR Jowett	sTTR Waterfield
Spearman's rho	sTTR Jowett	Correlation Coefficient	1.000	.899(**)
		Sig. (2-tailed)	.	.000
		N	50	50
	sTTR Waterfield	Correlation Coefficient	.899(**)	1.000
		Sig. (2-tailed)	.000	.
		N	50	50

**** Correlation is significant at the 0.01 level (2-tailed).**

The statistical analysis demonstrates that sTTR values from the two translations are significantly correlated. Specifically, the correlation is significant at the 0.001 level. Even though from the above table it seems that the significance of the correlation is nil (.000), the significance value is only out of the decimal span of SPSS, which was used to calculate the correlation.

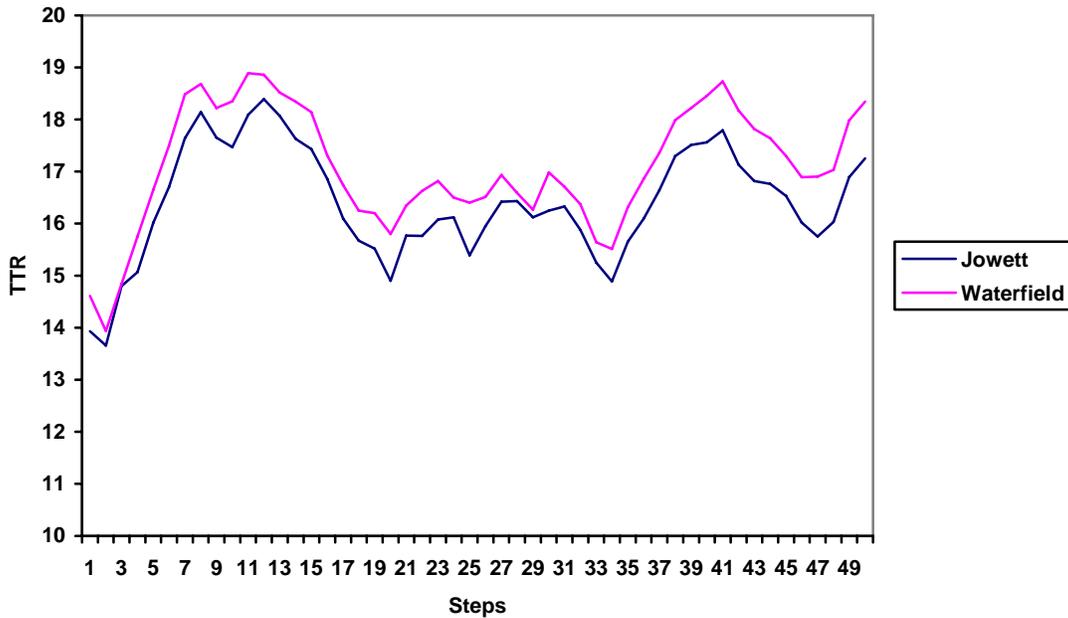
To conclude, the assumptions derived from the preliminary research, namely that sTTR values from the two translations are correlated, have been confirmed with this analysis. Furthermore, these results are more conclusive inasmuch as the method used is more sensitive and accurate than the one used in the preliminary research. Additionally, the larger number of sTTR values for comparison ensures that the correlation test is also more reliable.

Type-token ratio

Having demonstrated the correlation between sTTR values, the following task was to repeat the analysis by using non-standardised TTR values. As the correlation of sTTR values was significant, the results of this analysis are not expected to provide results that would differ much from the comparison of sTTR values. Nevertheless, it will be interesting to observe the difference between these two analyses. For the sake of

comparison, the results of this analysis will be presented in the same way as the results of the previous one. TTR values may be observed in Figure 6 (for the list of TTR values see Table 1 in Appendix 1).

Figure 6 - TTR values through steps



From the above figure it can be seen that the difference between the results from the two translations is much lower, due to the difference in size of the compared samples not being accounted for. Thus this analysis fails to effectively demonstrate the much higher vocabulary richness in Jowett’s translation. Additionally, the difference between the two sets of values seems less constant than with sTTR values. However, the values of TTR between the two translations follow the same pattern somewhat more accurately than sTTR values did. This will be demonstrated if the correlation coefficient is higher than with sTTR values. Table 4 shows the results of the statistical analysis.

Table 4 - Correlation between TTR values

			Jowett TTR	Waterfield TTR
Spearman's rho	Jowett TTR	Correlation Coefficient	1.000	.956(**)
		Sig. (2-tailed)	.	.000
		N	50	50
	Waterfield TTR	Correlation Coefficient	.956(**)	1.000
		Sig. (2-tailed)	.000	.
		N	50	50

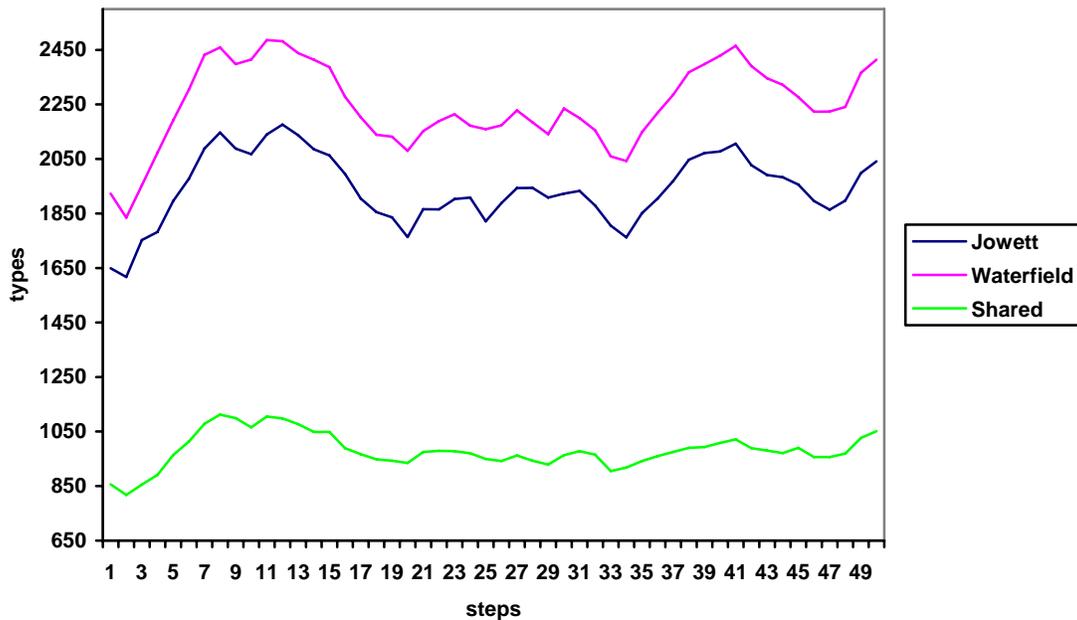
**** Correlation is significant at the 0.01 level (2-tailed).**

As assumed, from the table above it can be observed that TTR values correlated more closely than sTTR values. However, since the correlation was very significant with both analyses, it is better to compare sTTR values as they demonstrate the difference between the translations more accurately.

Shared vocabulary

The third task of the analysis was to correlate the amount of shared vocabulary with the total amount of vocabulary of both translators. To do so, first it was necessary to calculate the number of types in each of the segments for both translations. Second, it was necessary to compare each the vocabularies of each corresponding segments so as to determine the number of types occurring in both of the compared segments. Figure 7 shows the number of types used in each of the translations, as well as the number of types shared between them (for a list of the number of shared types and the number of types used in each of the translations, see Table 2 in Appendix 1).

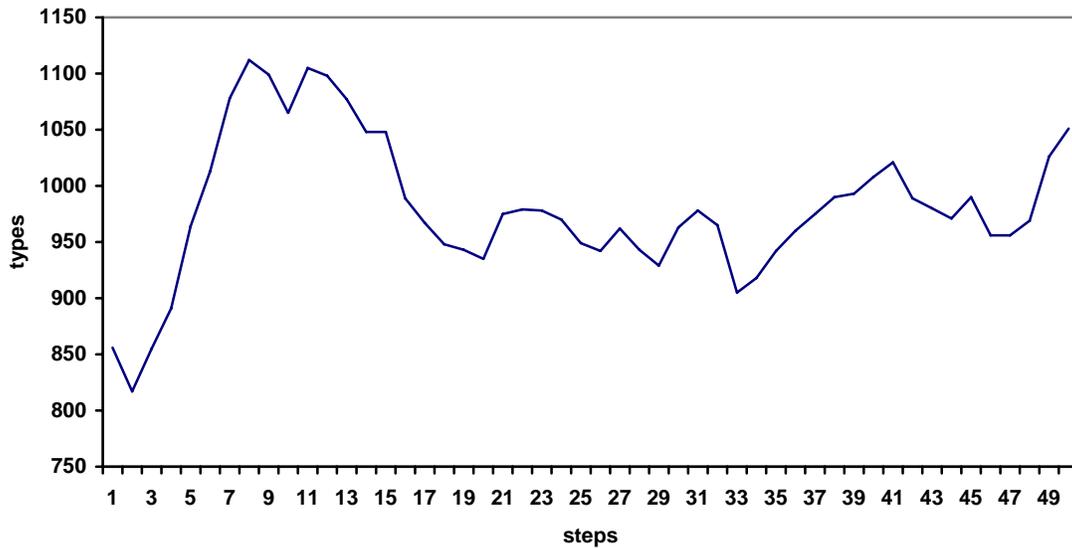
Figure 7 - The number of shared types and total number of types through steps



The two top lines, demonstrating the number of types used by each of the translators, follow a similar pattern as the lines from Figure 6, which demonstrate TTR values. On the other hand, the distance between the lines, namely the difference in the number of types used by each translator, is more striking, and also rather constant. Furthermore, values between the two translations seem to follow the same pattern more accurately than sTTR values.

However, the aim of this analysis was to determine if the number of shared types coincide with the total number of types used by each of the translators. Thus the bottom line, in the above figure, shows that the number of shared types follows a similar pattern as the total number of types used in each translation. However, since the graph is scaled to show three sets of values, the line representing shared types seems, at first glance, more 'flat' than the two lines demonstrating the total number of types used by each of the translators. Thus, in Figure 8, the number of shared types is shown on its own so as to observe the change in the number of shared type more easily.

Figure 8 - The number of shared types through steps



The above figure shows that distribution of shared types follows a similar pattern as the distribution of the total number of types used by the translators. Nevertheless, all the sets of values will be analysed statistically so as to demonstrate the correlation between them. (see Table 5)

Table 5 - Correlation between the number of shared types and the total number of types

			Jowett Types	Waterfield Types	Shared Types
Spearman's rho	Jowett Types	Correlation Coefficient	1.000	.956(**)	.915(**)
		Sig. (2-tailed)	.	.000	.000
		N	50	50	50
	Waterfield Types	Correlation Coefficient	.956(**)	1.000	.925(**)
		Sig. (2-tailed)	.000	.	.000
		N	50	50	50
	Shared Types	Correlation Coefficient	.915(**)	.925(**)	1.000
		Sig. (2-tailed)	.000	.000	.
		N	50	50	50

**** Correlation is significant at the 0.01 level (2-tailed).**

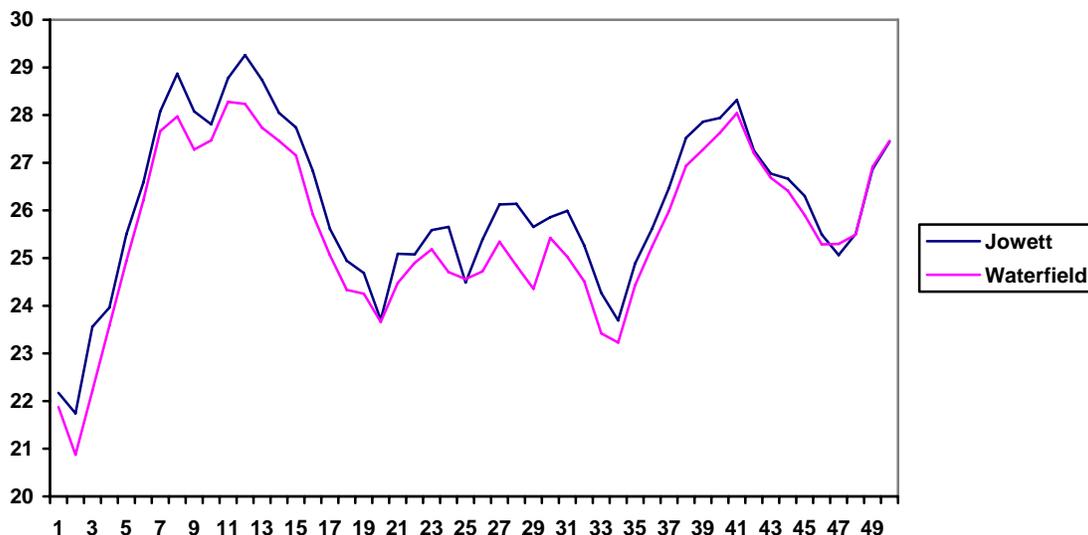
The statistical analysis demonstrated that the number of shared types correlates with the number of types used in both translations. Again the correlation is very significant, with the significance level being again out of the decimal span of SPSS. The correlation

between these three sets of values indicates that they are all determined by something in common to these texts, namely, the underlying work.

Relative vocabulary richness

The last task of the analysis was to neutralise the translator's dimension and compare the translations. As mentioned in the methodology chapter, the aim is to calculate the relative vocabulary richness (RVR) of each translation and compare these values. Relative vocabulary is simply the ratio between the number of types in a particular segment and the total number of types used in that translation. In this it does not matter which translator has a richer vocabulary, since the aim is to observe the way that vocabulary is distributed. We can observe the results of the analysis in Figure 9 (see the list of values in Table 3 Appendix 1).

Figure 9 - Relative vocabulary richness through steps



As already mentioned, the aim of RVR is to demonstrate that the vocabularies used in the compared texts are distributed evenly throughout the text. Thus, in an ideal situation the values between the texts should be equal. However, this is literally impossible, as in that case the whole idea of introducing steps to analysis to account for the uneven distribution of vocabulary throughout the text would be unnecessary. Thus the values of RVR between

to translations are, as demonstrated in Figure 9, expected to be as close as possible. As the values of RVR are very similar, they follow the same pattern, and are thus correlated. Nevertheless, for the sake of clarity, Table 6 shows the results of a correlation analysis.

Table 6 - Correlation between RVR values

			Jowett RVR	Waterfield RVR
Spearman's rho	Jowett RVR	Correlation Coefficient	1.000	.956(**)
		Sig. (2-tailed)	.	.000
		N	50	50
	Waterfield RVR	Correlation Coefficient	.956(**)	1.000
		Sig. (2-tailed)	.000	.
		N	50	50

** Correlation is significant at the 0.01 level (2-tailed).

Discussion

This research has provided some very interesting results, which support the preliminary research, and hence the two-dimensional theory of the style of translations, as well as the co-authorship theory of translations. However, as this research draws upon a debate on the authorship of translations, and proposes a new method which is to provide empirical evidence to support one of the theories, both the theoretical and methodical aspects of this research will have to be addressed. Firstly, the applied method will be critically assessed for all its advantages and disadvantages. Then the results obtained by it, together with their implications on authorship of translations will be discussed. Finally, alternative explanations of these results will be discussed together with respect to the other theories on authorship of translations.

Firstly, this research has demonstrated that, for an analysis of the style of translation, it is important to compare translations of the same underlying work rather than to compare the translations by the same translator. This is supported Baker (2000), who mentioned a caveat in her comparison of the style of translations by the same translator; specifically, that it was unclear how much of that style can be ascribed to the translator, and how much was being transposed from the original. Thus, the influence of the author can only be observed by comparing the translations of the same underlying work. In this way, one can observe the consistencies, or the author's dimension, and the differences, the translator's dimension, between the styles of the two translations.

Nevertheless, one of the main advantages of the proposed method is that it ensures an ensuing analysis of texts, thus allowing the researcher to observe the use of language throughout the text instead of observing the texts as whole. Furthermore, with respect to the method used in the preliminary research, the method is more sensitive to detect sudden and more subtle changes in style. Specifically, the method ensures that in a sufficient language sample for the extraction of various style-markers. On the other hand, steps ensure that the style is observed frequently enough to detect sudden or more subtle changes. However, the description of style depends rather on style-markers than on the method for the ensuing analysis.

As the compared translations were of different sizes, it was reasonable to assume that they would differ in their vocabulary richness. Thus, type-token ratio was chosen as a style-marker since it demonstrates vocabulary richness of a text. Furthermore, standardised type-token ratio was used to account for the differences in the size of language samples taken from the two translations so as to demonstrate the difference in vocabulary richness more effectively.

The results showed (see Figure 5) that the two translations had different values of sTTR, with one translation consistently having higher sTTR throughout the text. However, as with the preliminary research, these values followed a similar pattern, and were in fact correlated (see Table 3). Thus, these results support the two-dimensional theory of the style of translations, proposed in the preliminary research, as they demonstrate that the values of sTTR are only to some extent determined by the translator, since the results from both translations follow the same pattern, even though they are produced independently, and are actually quite different.³

Similarly, the results of comparison of non-standardised TTR values support the two-dimensional theory in the same way. However, as TTR values from the two translations are more similar to each other than sTTR values, they do not demonstrate the difference between the styles of the two translations as much as sTTR values do. Nevertheless, since they are not standardised, TTR values demonstrate more subtle changes in style, which was demonstrated by a higher correlation coefficient between them (see Table 4).

The analysis of shared vocabulary provided some very interesting results regarding the degree of similarity between translations. Specifically, it showed that the amount of shared vocabulary in the compared segments was determined in the same way the vocabularies of these segments were. However, more striking was the fact that the percentage of shared vocabulary comprised about 43% or 52% of the vocabulary (see Figure 7, or Table 2 in Appendix 1), depending on the translation compared with. Thus these translations did not seem to be nearly as similar as the translations compared by Turell (2004) (see Figure 1), indicating that the styles of the translations used in this

³ Just to remind the reader that there is a gap of 100 years between the productions of these translations, as well as a difference of about 20,000 words in their length.

analysis must have been very different. Still, the analysis showed that that was not true as their TTR values, whether standardised or not, were in fact correlated, thus demonstrating a certain degree of similarity.

It is important to notice that the principles of calculating shared vocabulary and overlapping vocabulary, used by Turell (2004), were quite different. Turell (2004) when applying CopyCatch to calculate the overlapping vocabulary made a stop list which excluded all Spanish function words and patronymic and toponymics from her analysis. However, it is even more surprising that Turell managed to demonstrate a considerably higher level of shared vocabulary, with all the words that were bound to appear in all translations excluded, than did my comparison of types used in both translations.

Another difference between the two approaches was the approach in which the vocabulary was compared, i.e. comparing the texts as whole or step by step. Additionally, CopyCatch compares texts for their tokens rather than types. The decision to compare texts for their tokens may be justified with shorter language samples, but, as it has been demonstrated with TTR values, the longer the texts are the more repetitive their vocabulary gets. Therefore two considerably long texts on the same topic would overlap in much more vocabulary than two shorter texts on the same topic. Thus even with a stop list excluding patronymic and toponymic expressions and function words from the comparison, with sufficiently large samples, like whole dramas or novels, the percentage of overlapping vocabulary would be much higher than in smaller excerpts of these texts. Thus the segmentation of texts, primarily so as to observe them step by step, also serves as a way of standardisation of these values, just as TTR is standardised by being subsequently calculated after every 1000 words.

Interestingly, the court ruling in the case reported by Turell (2004:2) concluded that with longer translations, such as translations of literary works, there is more space for the translator to be creative and original. However, it is unclear, as the authorship of translation is based on the form, what is meant by this “space for originality”. Nevertheless, the method used in Turell’s (2004) analysis shows exactly the opposite, i.e. the longer the texts are the more they are going to overlap in their vocabulary.

However, the percentage of overlapping vocabulary cannot be standardised in the sense that it is calculated after every 1000 tokens since the compared texts most likely will

not have the same number of tokens to compare. For the same reason, namely the difference in length, comparing the texts for their tokens segment by segment also would not ensure accurate comparison. Thus the idea of this research was to compare the texts for the types they share rather than tokens.

The main reason for comparing the texts for types rather than tokens was the fact that the compared segments, as explained previously, were not 100% comparable. Specifically, due to a large number of tokens, some of the tokens may be matched even though their matches would not actually be their counterparts; thus distorting the values and resulting with a higher percentage of overlapping vocabulary. This is avoided by a comparison of types, since once the types have been matched the number of times they occur in that segment does not make any difference on the result.

Furthermore, as mentioned previously, the percentage of shared tokens depends on the size of the language sample, which is also true for the percentage of shared types, but to a lesser extent. Thus, translations and non-translations will not necessarily differ in the percentage of shared vocabulary. However, translations will differ from non-translations in the sense that the percentages of shared vocabulary will be correlated with the vocabularies of the two translations, which will also correlate themselves (cf. analysis of shared vocabulary using larger samples, in Appendix 2).

The results of the final analysis, i.e. the comparison for the values of relative vocabulary richness, also demonstrate strong correlation, and are, as expected, quite similar. The aim of this measure was to demonstrate that even though the vocabularies of the compared translations were different in size, the way they are used throughout the texts is pretty much the same. Specifically, the aim was to demonstrate that roughly the same percentage of the total vocabulary will be used per segment for each of the translations. This was to demonstrate that there is an underlying consistency between the two translations, namely, the author's dimension.

However, as a careful reader may have observed, or assumed, the correlation between TTR, raw frequency of types, and RVR values between the two translations was exactly the same. Both TTR and RVR values are, in fact, based on the number of types used in each of the translations, but just divided by different numbers, the number of tokens and the number of types respectively. This posits the question of whether to

include the number of tokens in the analysis, and then, due to the difference in their number between translations, standardise the type-token ratio to actually observe the difference in the number of types between translations. Namely, the raw number of types sufficiently demonstrates the difference in lexical richness between the two translations. On the other hand RVR just neutralises this difference so as to demonstrate that roughly the same proportion of vocabulary is used in translation no matter how rich the translator's vocabulary is.

From the results of these analyses, one may conclude that, when analysing translations, one style-marker represents two-dimensions of style: the author's and the translator's. Thus, as Hermans (1996) states, one can observe two voices in a translation, which indicates that a translation is a result of co-authorship.

This co-authorship is, in a way, a result of a dialogue between the author and the translator. Consequently, derivative authorship would be then described as the translator's monologue inspired by the underlying work. In the former case, it would be reasonable to detect the style of the underlying text in the translation, since it is a result of a dialogue between two authors, thus both 'voices' are being observed. On the other hand, in the latter case, between two translations of the same underlying work, one would normally expect similar concepts to occur, however, the styles of such monologues would not be correlated. Therefore, co-authorship theory accounts for the possibility that the style of the original may be partly transferred to and hence observed in another language, whereas the derivative theory would ascribe such phenomena to imitation. This case would then demonstrate that both of the translators were indeed very good imitators of the style of the original, since their styles have been so successfully correlated throughout the text.

As an alternative to the two-dimensional theory, one can interpret these results as merely representing the ideas of the author, and not the dimension of his or her style. Specifically, as these analyses used style-markers that actually demonstrate the choice of vocabulary, one can argue that the correlation of values between two translations only demonstrates that these translations report the same ideas, which would support the derivational theory of authorship of translations. Thus these results would not undermine Venuti's (1995, 1998) derivative theory, or the legal view on authorship of translations.

However, as it has been noticed in the theoretical review, by assuming the derivative theory, the authorship of translations resides in the form, whereas the authorship of the 'original' resides in the originality of its ideas, as well as the form, but only within that language. Thus, the two forms of authorship, one derivative, the other creative, are quite different from each other, since they are based on completely different grounds.

On the other hand, if one assumes that a translation is a result of co-authorship between the author and the translator, then the authorship of the translation resides in both the form and the sense, since the author provides the original ideas and the manner of expression, which are then expressed in the form provided by the translator, i.e. the foreign language. Thus, in case of co-authorship, just as with any other form of authorship, the authorship resides in both the form and the sense.

This makes the theory of co-authorship sound more plausible, as it offers a theoretical framework which treats the authorship of translation in the same way as any other form of authorship. Furthermore, the co-authorship theory supports, and is supported by the two-dimensional theory of style at the same time, since one presupposes the other. Namely, if one accepts that the results of this research indeed demonstrate a relation between the styles of the two translations, and hence two dimensions of style, then the style of translation is a result of co-authorship. In that sense, the style and authorship of translations are related.

On the other hand, derivative authorship cannot account for the consistencies between the styles of two translations, since it does not attribute the authorship of the translation to the author, but solely to the translator. Hence the style of translation has to be the translator's. In that case, if the style demonstrates two dimensions, the author's and translator's, and the authorship is attributed just to the translator, then style and authorship are not related, which goes against the main principles of authorship studies.

Finally, as only a small number of style-markers have been tested in this research, there is a lot of potential for further research. Specifically, only one pair of translations has been analysed in this way, and the two-dimensional theory still has to be tested on various other translations using various style-markers.

Furthermore, these results may encourage a more in-depth analysis of the style of translations, with intent to detect which linguistic elements are responsible for the correlation of the styles of translations. Thus if there really are some linguistic elements common to all translations of the same underlying work, they would actually be the style-markers of the author's style. However, as the style of translations has been described as a fusion of styles rather than a conglomeration of two sets of style-markers, it is unlikely that some elements of the author's style can be translated without being altered. Still, Baker (2000) proposed an analysis of the style of translations so as to identify the style of the translator, namely to determine which linguistic features are the result of the manner of expression typical of translator, in contrast to the stylistic features of the original work. Therefore, it might be possible through a cross-linguistic comparison for such style-markers to be detected.

Another potential application of this method may also be in the detection of plagiarism of translations. Specifically, the fact that, in translations, each style-marker represents two styles at the same time may be applied in plagiarism detection. Because of it is possible to demonstrate that the two translations represent the same original work by using exactly the same style-markers one would use to demonstrate that the texts are of different authorship. However, if one of the translations is partly plagiarised, both tasks would be harder to prove. Namely, since the style of the translation is a fusion of the translator's and the author's styles, when one 'borrows' someone else's translation, the change of style does not only affect the translator's style dimension, but the original author's dimension as well. Therefore if the author's dimension is altered, the plagiarised translation will not correlate with non-plagiarised. However, this theory still needs to be tested on a translation known to be partly plagiarised.

Conclusion

The aim of this research was to readdress the issue of the authorship of translation. Thus far there has been much debate on this issue; however, no definite conclusions have been made. The conclusions of this research are twofold: firstly, it presented a new method for the analysis of the style of translations, and possibly for the analysis of any texts; secondly, it contributed to the debate on the authorship of translations, supporting the claims with some empirical evidence.

Before the application of the method, the issue of authorship of translations was addressed from the very beginning, i.e. the very concept of authorship and its history together with its application on translations. Additionally, a case of plagiarism of translations was presented so that the issue is observed in practice.

Since it was based on the preliminary research, the aim of this research was firstly methodological. Specifically, the aim was to develop a method that could successfully compare translations for their styles. The results show that the applied method successfully compares two translations, as high degrees of correlation between the translations have demonstrated.

Furthermore, the results of the analysis demonstrate that, in translations, one style-marker can represent two styles, or more precisely two-dimensions of one fused style. This indicates that the style of translations is, as Boase-Beier (2006:41) said, a fusion of the author's and the translator's style. Consequently, if both the author and the translator are responsible for this style, then the translation is co-authored, or co-produced, as Hermans (1996) and Schiavi (1996) proposed. Specifically, if a translator derives a translation from an original, authoring a work of his own, as Venuti (1995, 1998) proposed, then one would not be able to observe the correlation between the styles of two translations, i.e. the author's dimension. Therefore, the two-dimensional style, which can be empirically demonstrated, indicates that the process of production of translations is indeed a process of co-authorship.

Alternative interpretations of these results have been offered, however they are not as plausible since they cannot account for the issues raised by Baker (2000) and Turell (2004). Namely, one of the issues raised by Baker (2000) and mentioned above, is

whether the translator linguistic preferences, or the style of his translation, are independent of the style of the original? Analysing two translations of the same original text, one can observe that translators have a degree of freedom, since the styles of these translations are somewhat different, namely the values for particular style-markers are different for each translation. On the other hand, the changes between these values of the compared translations are correlated to each other, indicating that both styles are dependent on something else, namely the original style. Therefore, two-dimensional theory demonstrates that translators only have a certain degree of freedom in creating the style of translation, since they are merely co-authors of it.

The other issue, raised by Turell (2004), i.e. what makes translations so similar, can also be accounted for by the two-dimensional theory, or more precisely by the co-authorship theory. Namely, since the translator is only the co-author of the translation, the choice of lexicon is limited by the original author and hence it could not have been produced independently in any of the translations. Thus we can conclude that the higher degree of shared lexicon when comparing two translations can be attributed to original author's dimension of style. Whereas due to the translator's dimension these translations will not be identical, i.e. they will not share the vocabulary completely.

To sum up, this research has demonstrated the two-dimensional nature of the style of translations. In other words, it has demonstrated that the unitary style of a translation is in fact a fusion of the translator's and the author's styles. Consequently, since the style and authorship are closely related, the authorship of a translation is in fact the result collaboration of the author and the translator, namely, translation is a result of co-authorship.

However, before the end of this thesis, it is necessary to point out that this research has included only one pair of translations, and thus its conclusions still have to be confirmed by subsequent research. Furthermore, due to the small scale of the research, the method was tested on only one sample, and it needs further testing to demonstrate its re-test reliability. Nevertheless, the validity of the method can be inferred from the theoretical grounds it is based on, namely the two-dimensional theory, which is at the same time the direct result of the method, and the theoretical works of Hermans (1996), Schiavi (1996), Barnstone (1993), Baker (2000), and Boase-Beier (2006).

Finally, the issue of feasibility of the method may be raised, due to the lack of corpus access tools which would be able to apply it automatically, since manual application on longer text might be too time-consuming. However, as this research has demonstrated such tools can be made for linguists to use them. Once such tools are available, it is necessary to test more style-markers using this method to test the method itself, but also to develop a better understanding of the style of translations.

Appendix 1

Table 1 – Type-token ratio values through steps

Step	sTTR Jowett	sTTR Waterfield	TTR Jowett	TTR Waterfield
1	31.9	34.9	13.93	14.61
2	31.6	34.1	13.66	13.94
3	33.5	35.7	14.80	14.84
4	34.5	37.4	15.07	15.75
5	35.1	37.8	16.02	16.65
6	36	39.1	16.70	17.50
7	37	40.4	17.64	18.48
8	37.5	40.6	18.14	18.68
9	37.4	40.8	17.65	18.22
10	37.3	40.6	17.47	18.35
11	37.8	41.2	18.09	18.89
12	38	41.3	18.39	18.86
13	38	40.6	18.07	18.52
14	37.5	40	17.63	18.34
15	37.4	39.9	17.43	18.14
16	36.6	39	16.86	17.31
17	35.5	38.3	16.10	16.74
18	35.2	37.6	15.67	16.25
19	35.2	38.2	15.52	16.20
20	34.3	37.6	14.90	15.80
21	34.8	38.1	15.77	16.35
22	34.9	38.8	15.76	16.63
23	35.7	39.6	16.08	16.82
24	35.9	38.6	16.12	16.50
25	34.6	38.8	15.39	16.40
26	35.4	38.9	15.95	16.51
27	36.2	40.1	16.42	16.93
28	36.9	40.4	16.43	16.59
29	36.9	40.1	16.12	16.27
30	37.3	40.5	16.25	16.98
31	37.1	40.1	16.33	16.71
32	36.3	39	15.88	16.37
33	36	38.7	15.25	15.64
34	35.4	38.4	14.89	15.51
35	36.6	39.9	15.65	16.32
36	36.9	40.5	16.10	16.86
37	37.4	40.4	16.64	17.36
38	38.1	41.1	17.30	17.99
39	38	41.7	17.51	18.22
40	38.1	41.7	17.56	18.45
41	38.6	42	17.79	18.73
42	37.8	41.9	17.13	18.17
43	37.1	40.8	16.82	17.82
44	36.2	40.2	16.76	17.64
45	35.8	39.6	16.53	17.30
46	35	38.8	16.02	16.89
47	34.9	39.3	15.75	16.90
48	35.3	39.5	16.03	17.03
49	35.9	40.5	16.89	17.98
50	36.5	40.7	17.25	18.34

Table 2 – Number of shared types and the total number of types

Step	Shared Types	No. Types – Jowett	% of shared voc. in Jowett	No. Types – Waterfield	% of shared voc. in Waterfield
1	856.00	1649	51.910249	1923	44.513781
2	817.00	1617	50.525665	1835	44.523161
3	855.00	1752	48.80137	1953	43.778802
4	891.00	1782	50	2073	42.981187
5	964.00	1896	50.843882	2192	43.978102
6	1,013.00	1977	51.239251	2304	43.967014
7	1,078.00	2088	51.628352	2432	44.325658
8	1,112.00	2147	51.7932	2459	45.221635
9	1,099.00	2088	52.6341	2398	45.829858
10	1,065.00	2068	51.499033	2415	44.099379
11	1,105.00	2140	51.635514	2486	44.448914
12	1,098.00	2176	50.459559	2482	44.238517
13	1,077.00	2137	50.397754	2438	44.175554
14	1,048.00	2086	50.239693	2414	43.413422
15	1,048.00	2063	50.799806	2387	43.904483
16	989.00	1995	49.573935	2278	43.415277
17	967.00	1905	50.761155	2203	43.894689
18	948.00	1855	51.105121	2139	44.319776
19	943.00	1836	51.361656	2132	44.230769
20	935.00	1763	53.0346	2080	44.951923
21	975.00	1866	52.250804	2152	45.306691
22	979.00	1865	52.493298	2189	44.723618
23	978.00	1903	51.392538	2214	44.173442
24	970.00	1908	50.838574	2172	44.6593
25	949.00	1821	52.114223	2159	43.955535
26	942.00	1888	49.894068	2173	43.350207
27	962.00	1943	49.511065	2228	43.177738
28	943.00	1944	48.50823	2184	43.177656
29	929.00	1908	48.689727	2141	43.390939
30	963.00	1923	50.078003	2235	43.087248
31	978.00	1933	50.59493	2200	44.454545
32	965.00	1879	51.357105	2155	44.779582
33	905.00	1805	50.138504	2059	43.953375
34	918.00	1762	52.099886	2042	44.955926
35	942.00	1851	50.89141	2148	43.854749
36	960.00	1905	50.393701	2219	43.262731
37	975.00	1969	49.517522	2285	42.669584
38	990.00	2047	48.363459	2368	41.807432
39	993.00	2072	47.92471	2398	41.409508
40	1,008.00	2078	48.508181	2429	41.498559
41	1,021.00	2106	48.480532	2465	41.419878
42	989.00	2027	48.791317	2391	41.363446
43	980.00	1991	49.221497	2346	41.773231
44	971.00	1983	48.966213	2322	41.817399
45	990.00	1956	50.613497	2277	43.478261
46	956.00	1896	50.421941	2223	43.004948
47	956.00	1864	51.287554	2224	42.985612
48	969.00	1897	51.080654	2241	43.239625
49	1,026.00	1998	51.351351	2366	43.364328
50	1,051.00	2041	51.494366	2414	43.537697

Table 3 – RVR values through steps

Step	RVR Jowett	RVR Waterfield
1	22.172918	21.874645
2	21.742638	20.873621
3	23.557886	22.215902
4	23.961275	23.580935
5	25.49415	24.934591
6	26.5833	26.208622
7	28.075836	27.664658
8	28.869167	27.971788
9	28.075836	27.277899
10	27.806911	27.471277
11	28.775043	28.278921
12	29.259109	28.233421
13	28.734705	27.732908
14	28.048944	27.459902
15	27.739679	27.152769
16	26.825333	25.912865
17	25.615168	25.059721
18	24.942854	24.331703
19	24.687374	24.252075
20	23.705795	23.660563
21	25.090763	24.479582
22	25.077316	24.900467
23	25.588274	25.184849
24	25.655506	24.707087
25	24.48568	24.559208
26	25.386581	24.718462
27	26.126125	25.344103
28	26.139572	24.84359
29	25.655506	24.354454
30	25.857201	25.423729
31	25.991663	25.025595
32	25.265564	24.513708
33	24.270538	23.42168
34	23.692348	23.228302
35	24.889069	24.43408
36	25.615168	25.241724
37	26.475729	25.992493
38	27.52454	26.93664
39	27.860697	27.277899
40	27.941374	27.630531
41	28.317869	28.040041
42	27.255613	27.198271
43	26.771547	26.686384
44	26.663977	26.413378
45	26.300928	25.901489
46	25.49415	25.287226
47	25.063869	25.298601
48	25.507597	25.49198
49	26.865671	26.913889
50	27.443861	27.459902

Appendix 2

Table 1 – Shared vocabulary on segments the size of 50% of vocabulary

Step	Shared Types	No. Types – Jowett	% of shared voc. in Jowett	No. Types – Waterfield	% of shared voc. in Waterfield
1	2,938.00	5106	57.54	5962	49.28
2	2,919.00	5086	57.39	5966	48.93
3	2,911.00	5096	57.12	5925	49.13
4	2,901.00	5093	56.96	5930	48.92
5	2,891.00	5078	56.93	5925	48.79
6	2,913.00	5129	56.79	5936	49.07
7	2,928.00	5147	56.89	5971	49.04
8	2,932.00	5156	56.87	5976	49.06
9	2,962.00	5211	56.84	6002	49.35
10	2,986.00	5233	57.06	6025	49.56
11	2,996.00	5236	57.22	6022	49.75
12	3,003.00	5245	57.25	6015	49.93
13	2,989.00	5232	57.13	5992	49.88
14	2,980.00	5205	57.25	5985	49.79
15	2,985.00	5200	57.40	5945	50.21
16	2,991.00	5216	57.34	5951	50.26
17	2,940.00	5148	57.11	5922	49.65
18	2,932.00	5123	57.23	5891	49.77
19	2,940.00	5136	57.24	5918	49.68
20	2,941.00	5142	57.20	5907	49.79
21	2,932.00	5128	57.18	5873	49.92
22	2,897.00	5080	57.03	5837	49.63
23	2,851.00	5019	56.80	5787	49.27
24	2,833.00	5004	56.61	5766	49.13
25	2,818.00	5006	56.29	5765	48.88
26	2,815.00	5006	56.23	5741	49.03
27	2,811.00	5002	56.20	5751	48.88
28	2,796.00	5006	55.85	5735	48.75
29	2,777.00	4976	55.81	5712	48.62
30	2,793.00	4982	56.06	5754	48.54
31	2,796.00	4985	56.09	5765	48.50
32	2,817.00	5006	56.27	5811	48.48
33	2,822.00	5017	56.25	5829	48.41
34	2,834.00	5052	56.10	5858	48.38
35	2,813.00	5032	55.90	5815	48.37
36	2,804.00	5025	55.80	5826	48.13
37	2,789.00	5015	55.61	5819	47.93
38	2,783.00	5016	55.48	5796	48.02
39	2,798.00	5015	55.79	5844	47.88
40	2,814.00	5039	55.84	5860	48.02
41	2,818.00	5046	55.85	5857	48.11
42	2,816.00	5030	55.98	5875	47.93
43	2,814.00	5035	55.89	5865	47.98
44	2,815.00	5035	55.91	5875	47.91
45	2,838.00	5075	55.92	5896	48.13
46	2,824.00	5064	55.77	5896	47.90
47	2,822.00	5050	55.88	5898	47.85
48	2,843.00	5100	55.75	5939	47.87
49	2,863.00	5142	55.68	5976	47.91
50	2,879.00	5151	55.89	6015	47.86

Appendix 3

Perl script used to split texts into segments:

```
print "which file are you chopping?";
chomp($infile=<>);
open($in,"$infile")||die "infile 1";
open($out1,">out1.txt")||die "something went wrong with out1";
open($out2,">out2.txt")||die "something went wrong with out2";
open($out3,">out3.txt")||die "something went wrong with out3";
open($out4,">out4.txt")||die "something went wrong with out4";
open($out5,">out5.txt")||die "something went wrong with out5";
open($out6,">out6.txt")||die "something went wrong with out6";
open($out7,">out7.txt")||die "something went wrong with out7";
open($out8,">out8.txt")||die "something went wrong with out8";
open($out9,">out9.txt") ||die "something went wrong with out9";
open($out10,">out10.txt")||die "something went wrong with out10";
```

```
$nr=0;
while(<$in>){
    $inp=$_;
    @inputlist= split(' ', $inp);
    foreach $w (@inputlist){
        $nr++; }
close($in);
print "$nr\n";
$newnumber=($nr/10);
open($in2,"<$infile") ||die "wrong input 2";
```

```
$no=0;
while(<$in2>){
    $input=$_;
    @inputword= split(' ', $input);
    foreach $word (@inputword){
        $no++;
        if($no <= $newnumber){
            print $out1 "$word "; }
        elsif($no <= (2 *$newnumber)){
            print $out2 "$word }
        elsif($no <= (3 *$newnumber)){
            print $out3 "$word }
        elsif($no <= (4 *$newnumber)){
            print $out4 "$word }
        elsif($no <= (5 *$newnumber)){
            print $out5 "$word }
        elsif($no <= (6 *$newnumber)){
            print $out6 "$word }
        elsif($no <= (7 *$newnumber)){
            print $out7 "$word "; }
        elsif($no <= (8 *$newnumber)){
            print $out8 "$word "; }
        elsif($no <= (9 *$newnumber)){
            print $out9 "$word }
        else{
            print $out10 "$word } } }
```

Bibliography

- Baker, M. (2004) A Corpus-Based View of Similarity and Difference in Translation. **International Journal of Corpus Linguistics**, 9 (2): 167-193.
- Baker, M. (2000) Towards a Methodology for Investigating the Style of a Literary Translator. **Target**, 12 (2): 241-266.
- Baker, M. (1992) **In Other Words : a coursebook on translation**. London: Routledge.
- Barnstone, W. (1993) **The Poetics of Translation: History, Theory, Practice**. New Haven; London: Yale University Press.
- Bennett, A. (2004) **The Author**. London: Routledge.
- Biber, D. Conrad, S. and Reppen, R. (1998) **Corpus Linguistics: Investigating Language Structure and Use**. Cambridge: Cambridge University Press.
- Boase-Beier, J. (2006) **Stylistic Approaches to Translation**. Manchester: St. Jerome.
- Bracha, O. (2006) **The Ideology of Authorship Revisited**. University of Texas, Public Law Research Paper No. 82
- Bosseaux, C. (2007) **How Does It Feel? : Point of View in Translation: The Case of Virginia Woolf into French**. Amsterdam; New York, N.Y.: Rodopi.
- Burrow, J.A. (1982) **Medieval Writers and Their Work: Middle English Literature and its Background, 1100-1500**. Oxford: Oxford University Press
- Chaski, Carole E. (2001). Empirical evaluations of language-based Author identification techniques. **Forensic Linguistics: The International Journal of Speech Language and the Law**. 8(1), 1-65

Coulthard, Malcolm (2006) **...and then... Language description and author attribution.**

Found at: <<http://www.aston.ac.uk/lss/staff/coulthardm.jsp>> 27th April 2008

Coulthard, Malcolm (2005) **Some Forensic Applications of Descriptive Linguistics.**

Found at: <<http://www.aston.ac.uk/lss/staff/coulthardm.jsp>> 27th April 2008

Coulthard, M. (1994) **Advances in Written Text Analysis.** London: Routledge.

Coulthard, M. and Johnson, A. (2007) **An Introduction to Forensic Linguistics: Language in Evidence.** London: Routledge.

Derrida, J. and Anidjar, G. (2002) **Without Alibi.** trans. Peggy Kamuf. Stanford: Stanford University Press.

Gibbons, J. (2001) "Language and the Law" In Davies, A. and Elder, C. (eds.) **The Handbook of Applied Linguistics** Oxford: Blackwell. pp. 285-303.

Ginsburg, J.C. (2003) **The Concept of Authorship in Comparative Copyright Law.** Public Law & Legal Theory Research Paper Group. Columbia Law School. Paper 03-51.

Gläser, Rosemarie (1998) The Stylistic Potential of Phraseological Units in the Light of Genre Analysis. in Cowie, A. P. (ed.) **Phraseology. Theory, Analysis and Applications.** Oxford: Oxford University Press.

Grant, Tim, and Baker, K. L. (2001) Identifying reliable, valid markers of authorship: a response to Chaski. **Forensic Linguistics: The International Journal of Speech Language and the Law.** 8(1), 66 - 79.

Hatim, B. and Mason, I. (1997) **The Translator as Communicator.** London: Routledge.

Hemmungs Wirtén, E. (2003) **No Trespassing: Authorship, Intellectual Property Rights, and the Boundaries of Globalization.** London: University of Toronto Press.

Hermans, T. (1996) The Translator's Voice in Translated Narrative. **Target.** 8 (1): 23-48.

- Hoey, M. (1993) **Data, Description, Discourse: Papers on the English Language in Honour of John McH Sinclair on his Sixtieth Birthday**. London: Collins.
- Hunston, S. (2002) **Corpora in Applied Linguistics**. Cambridge: Cambridge University Press.
- Kimmelman, B. (1999) **The Poetics of Authorship in the Later Middle Ages: The Emergence of the Modern Literary Persona**. New York: Peter Lang
- Kinnear, P.R. and Gray, C.D. (2000) **SPSS for Windows Made Simple: Release 10**. Hove: Psychology Press.
- Kristeva, J. (2002) "Nous Deux" or a (Hi)story of Intertextuality. **The Romanic Review**, 93(1).7+. Found at:
[http://www.questia.com/read/5002035257?title=%22Nous%20Deux%22%20or%20a%20\(Hi\)story%20of%20Intertextuality](http://www.questia.com/read/5002035257?title=%22Nous%20Deux%22%20or%20a%20(Hi)story%20of%20Intertextuality)> 2nd April, 2009.
- Lefevere, A. (1992) **Translation, History, Culture : a sourcebook**. London: Routledge.
- Lord, A.B. (1964; 1960) **The Singer of Tales**. Vol. 24 Cambridge, Mass; London: Harvard University Press; Distributed by Oxford University Press.
- Malmkjaer, K. (2003) What Happened to God and the Angels: An Exercise in Translational Stylistics. **Target**, 15 (1): 37-58.
- Malmkjaer, K. (2004) Translational Stylistics: Dulcken's Translations of Hans Christian Andersen. **Language and Literature**, 13 (1): 13-24.
- Marco, J. (2004) Translating Style and Styles of Translating: Henry James and Edgar Allan Poe in Catalan. **Language and Literature**, 13 (1): 73-90.
- Marrotti, A.F. (1995) **Manuscript, Print, and the English Renaissance Lyric**. Ithaca: Cornell University Press.

- McMenamin, Gerald R., (2002) **Forensic Linguistics: Advances in Forensic Stylistics**. Boca Raton: CRC Press LLC
- Millan-Varela, C. (2004) Hearing Voices: James Joyce, Narrative Voice and Minority Translation. **Language and Literature**, 13 (1): 37-54.
- Morson, G.S. & C. Emerson (1990) **Mikhail Bakhtin: Creation of a Prosaics**. Palo Alto: Stanford University Press.
- Nagy, G. (1996) **Homeric Questions**. Austin: University of Texas Press.
- Oakes, M.P. (1998) **Statistics for Corpus Linguistics**. Edinburgh: Edinburgh University Press.
- Olohan, M. (2004) **Introducing Corpora in Translation Studies**. London: Routledge.
- Olsson, J. (2008) **Forensic Linguistics**. 2nd ed. London: Continuum.
- Olsson, J. (2004) **Forensic Linguistics : An Introduction to Language, Crime and the Law**. London: Continuum.
- Olsson, J. (unpublished) **Author Anonymous**. (a course book for forensic linguistics certificate course).
- Plato and Jowett, B. (1908) **The Republic of Plato**. 3rd ed. Oxford: Clarendon Press.
- Plato and Waterfield, R. (1994) **Republic**. Oxford: Oxford University Press.
- Rose, M. (1993) **Authors and Owners : The Invention of Copyright**. Cambridge, Mass.; London: Harvard University Press.
- Saunders J.W. (1964) **The Profession of English Letters**. London: Routledge and Kegan Paul.
- Schiavi, G. (1996) There Is Always a Teller in a Tale. **Target**, 8 (1): 1-21.

- Sinclair, J.M. (1991) **Corpus, Concordance, Collocation**. Oxford: Oxford University Press.
- Sinclair, J.M. and Carter, R. (2004) **Trust the Text : Language, Corpus and Discourse**. London: Routledge.
- Toury, G. (1995) **Descriptive Translation Studies and Beyond**. Amsterdam: John Benjamins.
- Trigg, C. (2002) **Congenial Souls: Reading Chaucer from Medieval to Postmodern**. Minneapolis, University of Minnesota Press.
- Turell, M.T. (2004) Textual Kidnapping Revisited: The Case of Plagiarism in Literary Translation. **The International Journal of Speech, Language and the Law**, 11 (1): 1-26.
- Venuti, L. (1998) **The Scandals of Translation: Towards an Ethics of Difference**. London: Routledge.
- Venuti, L. (1995) **The Translator's Invisibility: A History of Translation**. London: Routledge.
- Wall, W. (1993) **The Imprint of Gender: Authorship and Publication in the English Renaissance**. Ithaca: Cornell University Press.
- Woodmansee, M. and Jaszi, P. (1994) **The Construction of Authorship: Textual Appropriation in Law and Literature**. Durham, NC: Duke University Press.