

## **Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton**

Mohammed Al-Batineh

Yarmouk University, JORDAN

### ABSTRACT

Scholarly attempts to uncover plagiarism in literary retranslations have for the most part adopted a manual comparison of segments (Wright, 1904; Leighton, 1994). Recent computerized studies related to investigating plagiarism in retranslation also present methodological issues related to data scalability or to the methodological approach (Turell 2004, Şahin et al. 2015). The present paper discusses the difficulties of tracing plagiarism between retranslations and proposes a corpus-based target-text-oriented approach for detecting and investigating plagiarism in retranslation by blending methods from forensic stylistics and corpus linguistics. In this paper, I argue that investigating plagiarism in retranslation in this way highlights the potential of this method of analysis to reveal something of the linguistic ‘fingerprint’ of the original translator and so trace it in the plagiarized version(s). To this end, I reinvestigate a classical retranslation plagiarism controversy involving John Payne and Sir Richard Burton over the English retranslations of *One Thousand and One Nights*, as a case study. The combination of corpus qualitative and quantitative analyses used provided pieces of evidence that Burton had palatized in his translation. The findings of the case study suggest the potential viability of the proposed method in investigating and revealing plagiarism in retranslation, especially in cases where plagiarizers replace words with their synonyms to hide their plagiaristic act.

**KEYWORDS:** Arabian Nights, orientalism, corpus-based analysis, N-gram analysis, retranslation, translation plagiarism, plagiarism in retranslation, John Payne, Sir Richard Burton

## **Plagiarism in Retranslation: The Controversy of *The Arabian Nights***

Literary translation is a creative endeavor that gives, according to Benjamin (2000) an "afterlife" to the source text and introduces it to another cultural environment. Sometimes, the process of giving an "afterlife" to the source text inspires another attempt that manifests itself, as Antoni Berman (1990:1) puts it, through the act of retranslation. Retranslation can be narrowly defined as the art of producing another translation of a literary work that has been previously translated into the same target language (Şahin et al. 2015: 194). The phenomenon of literary retranslation is often considered a positive practice that diversifies and broadens the available interpretations of the source text (Tahir Gürçağlar 2009:233). Retranslation can also be an institutionalized act that is supported by certain institutions, such as the translation of the Holy Quran (See Herrag 2012 and Khosravi and Pourmohammadi 2016). In this regard, Lawrence Venuti (2004) explains that retranslations housed by institutions can "maintain and strengthen the authority of a social institution by reaffirming the institutionalized interpretation of a canonical text" (26). In other words, some institutions support retranslations in order to reaffirm the institution's interpretation of a specific text, to challenge other interpretations that go against the institution's ideology, or to enforce a particular ideology (Venuti 2004: 26).

Economic and commercial factors also play a role in driving retranslational acts. In some translation traditions, as explained by Venuti, commercial publishers opt for retranslating a canonical text rather than a 'new' or less known text. Reasons for this can be commercial: conducting a retranslation of an already translated text is sometimes cheaper than obtaining or purchasing translation rights from the translation rights holder (Venuti, 2004:30). Another market-driven practice related to retranslation that has recently emerged is known as "recycling." This process involves reprinting existing translations in a new format and design. The practice of "recycling" is attractive due to the "prestige, cost-effectiveness and guaranteed sales associated with the publication of literary classics" (Tahir Gürçağlar 2009:235). Lately, however, the practice of "recycling" has at times been adopted by some publishers to "reproduce" or to plagiarize existing translation of forty classical books and best sellers. Examples of this practice include, amongst many Dostoyevski's *The Idiot* William Shakespeare's *Macbeth* and Charles Dickens's *The Old Curiosity Shop*, and present them as new retranslations (Şahin et al. 2015). In

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

a decidedly digitized era, the means by which plagiarizers of translation can appropriate or ‘recycle’ online translated content and claim ownership of translations are clearly varied but ever more difficult to detect. The diversity of such questionable practices nonetheless draws attention to the importance of discussing, examining, and revealing potential plagiarism in translation, the principle reason being to highlight the importance of protecting translatorship rights and necessity of working to guard the creativity of one translator from being appropriated by another. Any discussion of plagiarism in translation involves, in the first instance, the investigating of how and by which means plagiarism in translation can be detected in the first place.

The process of interrogating the issue of plagiarism involves detecting where it occurs. There is a twofold difficulty in detecting plagiarism in retranslation, as argued by Teresa Turell (2004):

on the one hand, the fact that all translations will tend to reflect the author’s original form and content, and in so doing resemble that original piece of work; on the other, the circumstance that the more faithful to the original these translations are, the more difficult it is to detect their originality. (Turell 2004:1)

According to Turell (2004), there are two main scenarios for plagiarism in translation. The first is when an author/translator translates a work or a part of it into another language and claims that this work is his/her “original” and not a translation of or a copy from any source. This practice is called “Translational Plagiarism”. The second scenario is when a translator “copies” an existing translation or a part of it and claims that the translation is his/her new retranslation of the same original work. This is “Retranslational Plagiarism” (Turell 2004: 7).

In this paper, I discuss scenario two, retranslational plagiarism, in contexts of a case study drawn from one of the first plagiarism debates about retranslation within the Arabic-English literary tradition: the plagiarism controversy involving John Payne and Sir Richard Burton concerning the English retranslations of *One Thousand and One Nights* (*Arabian Nights* hereafter), one of the all-time most popular Middle Eastern and Indian collections of folk tales. Several tales from *Arabian Nights* such as “Aladdin” and “Sinbad” became very popular in the Western world manifested in Disney's adaptation of these tales into films. As a story about the stories *Arabian Nights* has travelled the world over numerous times through many languages, times and places.

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

John Payne's translation of *Arabian Nights* from Arabic into English is considered the first unabridged English rendition of the tales. It came out in eight volumes in 1882 under the title *The Book of the Thousand Nights and One Night*. Sir Richard Burton's translation was published 1885. Burton and Payne were English orientalist<sup>1</sup>, translators, poets, and writers. They were considered the greatest travelers and linguists documenting the manners and customs of the East, that England had ever produced (Wright 1904: 98). Their works and translations were well-known then and are well-known now, as their texts are currently used by anthropologists and students of Middle Eastern Studies until this day..

Despite the success of Payne's first version at the time of publication, Burton's later translation of *Arabian Nights* is also considered one the most publicly successful and well-known English translations of *Arabian Nights* (Wallen 2014:5). Burton's translation outsold Payne's and more copies of his translation were printed. However, Burton was accused of plagiarizing Payne translation of *Arabian Nights* in the early twentieth century. Joseph Campbell (1952), the editor of *The Portable Arabian Nights*, states that Payne's translation of *Arabian Nights* was "the first complete rendition of the work into a European language". However, Payne's version was, according to Campbell "appropriated straightway by his colourful friend, Captain Richard F. Burton, who immediately reissued it, slightly modified and garnished with a plethora of "anthropological" notes, under his name" (Campbell 1952: 31). Campbell continues that Burton followed Payne's translation "word for word, even semicolon for semicolon, not merely often, but particularly throughout" (Brodie 1984:341). We also find commentary on the possibilities of plagiarism on the part of Burton from Thomas Wright, who wrote the biographies of the two orientalist (Wright 1906). In his biography on Payne, Wright argued that Burton "copied whole pages from Payne merely altering one or two words" and that Burton's translation is a paraphrase of Payne's. He claims that Burton plagiarized "thousands of sentences from Payne, often without altering a single word" (Wright 1906:85). Wright bases his charges against Burton on an exchange of letters between Burton and Payne showing that Burton read Payne's translations before they were published. Wright also tried to prove the plagiarism by providing two parallel

---

<sup>1</sup> An Orientalist is someone from the West who studies the language, culture, history, or customs of countries in eastern Asia (Sinclair 2003: 923).

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

columns of the two translations to show that Burton's translation is "largely a paraphrase of Payne" (Brodie 1984:341).

In turn, other scholars tried to counter Wright and Campbell's charges against Burton. Norman Penzer (1997:319), for instance, argues that translators usually read and follow the steps of previous translations of the texts they want to translate and this is not evidence of translational plagiarism. Similarly, Mary Lovell claims that it was evident from the writings of Burton and his wife and from the statements of people who met him that "Burton had been collecting manuscripts of the Nights' stories and translating them, on and off, for over twenty-five years before he met Payne" (2000: 795–796). Lovell also argues that "Wright's claim that Burton had not done his own translation," but had "taken from Payne at least three-quarters of his entire work, is extraordinary" (2000: 796). Unlike Wright, who had provided letters and textual comparisons to back up his claims, Penzer (1997) and Lovell's (2000) counter-arguments were not based on any grounded methodology to provide pieces of evidence taken from the actual translations of Payne and Burton. Rather, their arguments were based on their own readings and judgments of each version without providing in their analysis any actual examples to show that Burton's translation was not plagiarized from Payne's.

This point brings me to the importance of researching alternative methodologies to detect and investigate plagiarism and re-translation. Firstly, it is important to note what sort of activity 're-translation' is. Retranslation is in fact a special form of creative writing which relies on an already existing text as a point of departure to create different target texts. This translational act can be unethically carried out by plagiarizers who can easily, with the availability of digital literary works, copy or manipulate previous translation and claim ownership of the translation. The very nature of retranslation makes it difficult for plagiarism investigators to prove plagiarism cases. This is because translators attempt, in many instances, to reflect the content and the style of the original text from which they translate. In this way, it is difficult to detect whether a stylistic choice originates from a translator's reading of the source text, or whether s/he is simply appropriating the work already carried out previously by another translator. For this reason, I argue in this paper that particular approaches are thus required to investigate the

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

issue of plagiarism in retranslations. Focusing on plagiarism in literary retranslations, I explain what these approaches could be by reinvestigating the plagiarism controversy of Payne and Burton, using forensic stylistics and corpus linguistics. I take this controversy as a case study for many reasons. It also has been a source of literary inspiration for English authors (see Al-Olaqi 2012) for centuries. *Arabian Nights* is considered the first “Romantic work of prose-fiction promoting Orientalism”. Burton’s translation of *Arabian Nights* has been widely used by scholars of orientalist and oriental studies as a reliable translation and source of data in (Sabouri 2011: 123). Aside from issues of textual accuracy and ethics of authorial credibility, an investigation of the Payne and Burton controversy is thus in itself very timely and relevant.

## **Investigating Retranslational Plagiarism**

When investigating retranslational plagiarism, three approaches are usually followed: manual investigation; computerized investigation and combining Forensic Stylistics and Corpus Analysis. The first approach relies on the manual investigation and close reading of disputed translations. It focuses on the careful reading of passages with an emphasis on lexis and syntax (Wright, 1904; Leighton, 1994). The second approach relies on computer applications to empirically trace plagiarism in a collection of retranslations (Turell 2004). A third hybrid approach to plagiarism detection combines manual qualitative and computerized quantitative approaches to analyze and investigate plagiarism in retranslations (Şahin et al. 2015; Şahin et al 2019). The following sections provide an overview of the methods used in investigating and detecting plagiarism in retranslation.

### *Manual Investigation of Retranslational Plagiarism: Critical Perspectives*

Manual analysis of a disputed document in retranslational plagiarism focuses on comparing an original text to its translation to reveal translational plagiarism or comparing different translations of the same source to spot instances of overlap and eventually translational plagiarism. One of the first attempts to manually investigate retranslational plagiarism was carried out by Wright (1904), as cited earlier, to prove his claim that Richard Burton plagiarized

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

from John Payne's English translation of *Arabian Nights*. Wright manually compared selected parts of Payne's and Burton's translations of *Arabian Nights* to reveal instances of overlap in the two translations. Based on the analysis of a small sample of parallel sections taken from their English translations, Wright concluded that Burton copied Payne's translation word for word in the tales he analyzed ("The Blacksmith who could Handle Fire without Hurt" and "Aboulhusn Ed Durraj and Abou Jaafer the Leper"). Despite the meticulousness of Wright's research, the task of tracing definitive instances of plagiarism in the entire disputed translated work was an almost unsurmountable task since Payne's translation of *Arabian Nights* was nine volumes long, each volume made up of around 250-350 pages. Another critical investigation into a case of retranslational plagiarism has involved the retranslation of the Russian-English translation of *The Bronze Horseman: Selected Poems of Alexander Pushkin* (Leighton 1994). Simon Karlinsky, professor of Slavic languages and literature, provided pieces of qualitative evidence, that were parallel examples, in which Thomas "erred exactly where Fennell erred and invented precisely where Fennell invented" (Karlinsky c.f. Leighton, 1994:74).

Reproducing or plagiarizing previous translations can however sometimes pass unnoticed, as was the case in the texts discussed above. Both Burton's translation of *Arabian Nights* and Thomas' retranslation of *The Bronze Horseman: Selected Poems of Alexander Pushkin* were very popular at their times of publication. Moreover, both translators were highly praised for their translations. In these instances, the purportedly plagiarized translation gained wider recognition and more commercial success than the initial translation of the original work. Five hundred copies of Payne's translation of the *Arabian Nights* were printed. More than one thousand copies were made and sold however from Burton's translation.

The translation and plagiarism cases reported by Wright (1904) and Leighton (1994) were investigated manually by the scholars comparing different translations of the same source to spot instances of retranslational plagiarism. While close reading might be useful for tracing plagiarism in translation, it is inapplicable to a large corpus of disputed texts such as the *Arabian Nights*. During Wright's time, the tools, technologies, and methods used for investigating retranslational plagiarism were very limited and thus led to limited conclusions. In this regard,

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

Matthew Jockers argues that close reading of a large corpus of work becomes “untenable as an exhaustive or definitive method of evidence gathering,” and as a result, “something important will inevitably be missed” (2013:9). That said, computerized methods for processing sizable numbers of digital texts have, however, been recently developed, enabling researchers to analyze large corpora of texts and to reach conclusions that are based on evidence drawn from quantitative-focused research methodologies. Such an approach of investigating plagiarism in translation relies on the power of computerized methods such as corpus analysis tools. Corpus methods give the researcher the ability to process a larger number of texts during shorter periods of time in comparison with the methods s/he uses for manual data analysis. Corpus methods also reduce the human bias as the computer processes the data consistently and accurately (Barnbrook 1996: 11). However, compiling a digital corpus from hardcopies is a very time-consuming process. Processing a larger corpus requires moreover a solid knowledge of corpus tools, and in some instances programming. The following section discusses in more detail the corpus methods used in investigating retranslational plagiarism and why such powerful tools also can have limitations.

### *Computerized Investigation of Retranslational Plagiarism: Corpus-based Analysis*

Mona Baker defines ‘corpus’ as a collection of digital texts or documents “held in an electronic form and analyzable automatically or semi-automatically” (1995:225). Corpus analysis tools allow for the processing of many texts in seconds, thus providing quantitative and qualitative pieces of evidence such as percentages of overlapping sentences or words in disputed documents to support plagiarism detection and analysis in very short periods of time. In her pioneering study on plagiarism in re-translations, Turell (2004) made use of computerized methods to investigate plagiarism in the case of Shakespeare’s *Julius Caesar* Spanish translation by Ángel Luis Pujante (1987) and subsequent retranslation by Vázquez Montalbán (1988). Like Burton, Montalbán was accused of plagiarizing and reproducing Pujante’s translation (Turell 2004: 2—3). In her study, Turell used *CopyCatch*, a software developed to detect non-originality in literary translation. With this software, the researcher can compare two documents and provide the percentage of similarity between them. In this way, Turell analyzed four Spanish retranslations of Shakespeare’s *Julius Caesar* by tracing instances of similarity in ‘overlapping vocabulary’ across all four versions, that is the percentage of shared vocabularies between the different

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

retranslations. The second set of traced evidence she investigated was *shared once-only words*, “words that occur only once in each retranslation (of the pairs compared)” but occur in both disputed translations (Turell 2004:9). Turell also traced *unique vocabulary* in each retranslation, which, she argues, helps reveal if the retranslation has been produced originally or copied from other retranslations (2004:10). The last piece of evidence examined was *shared once only phrases*, a method that reveals words overlapping at the phrase level.

Turell's study was one of the first studies to investigate plagiarism in retranslation using computerized methods. The result of her analysis supported the study's claim that Vázquez Montalbán's *Julius Caesar* Spanish translation was plagiarized from Pujante's. Turell reported that the methodology used along with the findings were strong enough to prove plagiarism. However, the potential limitation of using a computerized method is that the tool for detecting ‘overlapping vocabulary’ might not be enough to trace retranslational plagiarism in cases where a plagiarizer changes words using synonyms. That is, tracing overlapping vocabulary using corpus tools or other tools such as *CopyCatch* may not reveal a high percentage of overlap as the disputed translations because synonyms are not criteria flagged up by the program. As in such cases the plagiarism case cannot be proven outright, an alternative in-depth method of analysis is needed to detect and reveal where linguistic patterns of the plagiarized text intersect with patterns in the disputed text.

As a response to this issue, Şahin et al. (2015) adopted a hybrid approach to plagiarism detection by combining manual qualitative and computerized quantitative methods to analyze and investigate plagiarism in retranslations. Using such an approach, the researchers analyzed forty retranslations of classical books distributed as a part of a promotional campaign by *Star*, a national newspaper in Turkey. The researchers collected and analyzed data in three stages. The first stage included documentary analysis of the translators, publishers, and prices of the disputed retranslations. The second stage involved a paratextual analysis of footnotes and translators' prefaces in addition to a comparative content qualitative analysis of particular sections in the disputed retranslation and the original translation. They also manually analyzed five random pages of the actual retranslation in an attempt to spot instances of overlap between the original

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

translations and the plagiarized versions. Finally, the researchers scanned and searched the content of the first page of the retranslations online to trace their “twins” on the Internet.

Based on the manual comparative analysis of the disputed retranslations, the researchers claimed that the first retranslation was merely a reproduction of the previous translations with only a few lexical changes. The analysis also revealed that the plagiarizer replaced words with their synonyms and used contemporary words when producing the retranslation without changing the syntactic structure of the sentences. The researchers also found that the plagiarizers presented their retranslations without any the paratexts and footnotes. In a second case, the retranslation was a shortened version of the original translation with some changes in word choice and word order but without any changes to the syntactic structure (Şahin et al. 2015:208). The Internet search of the content of the first page revealed “absolute overlapping between the sentences” of the plagiarized translations and the online versions of original translations (206). The researchers concluded that the plagiaristic act is most obvious in the middle and last sections of the retranslations, and the activity is not a retranslation but a “reproduction” of previous translations.

Şahin et al. (2015)’s comparative analysis was based on five-page random pages taken from the beginning, the middle and the last part of the actual retranslations in an attempt to spot instances of overlap between original translations and the plagiarized versions. Increasing the size of the data set would have made it almost impossible to trace plagiarism manually. However, the power of corpus and computerized analysis methods did enhance and facilitate the process of plagiarism detection and can be used, according to Şahin et al,(ibid) to analyze larger data sets. They argue there that, combining qualitative and quantitative approaches to collect and trace evidence of plagiarism in translation can both save researchers time and potentially help reach solid conclusions. In some retranslational plagiarism cases, for example, plagiarizers opt for replacing words with their synonyms or for changing the syntactic structure of the plagiarized translation to hide the attempt to plagiarize. In such cases, tracing *vocabulary* or/and *page content overlap* would not help researchers reveal and detect retranslational plagiarism.

Therefore, an interdisciplinary approach that combines methodologies to reveal conscious and unconscious patterns of language use is needed to uncover such plagiarism. There is however an

element of re-translation which is yet to be researched in-depth: that of tracing 'unique translational style' Each literary translator has, in effect, his or her own style that combines conscious and unconscious habitual use of language to create a unique translational style. This style can form or emerge through detecting the use of certain vocabularies, phrases, or sentence structure. In contexts of this paper, I argue that a translator's habitual patterns of language use within a translation can be traced, counted and categorized using a combination of corpus tools and analyses. To contribute to an emerging field of research of corpus research in translation studies (e.g., Al-Batineh 2015; Baker 2000; and Saldanha 2011), I demonstrate how these corpus tools and analyses can be used, taking the re-translation of *Arabian Nights* as a case study. The following section sets out my proposed approach for detecting and investigating plagiarism in retranslation by combining methods from forensic stylistics and corpus analysis.

*Combining Forensic Stylistics and Corpus Analysis to Uncover Plagiarism in Re-translation*

The present study aims to reinvestigate a literary retranslational plagiarism controversy by combining methods from forensic stylistics and corpus linguistics to assess their viability as research tools in this respect. This method has never been used before when investigating plagiarism in literary retranslation. This study serves as a test for the viability of such a proposed approach. It builds on findings in the forensic stylistics field where each author has unconscious stylistic habits in his or her writing that are well grained and can be traced and identified. This means that no two authors write in the same exact way (Burrows 1992; McMenamin 2002).

*Forensic stylistics* is a field of study that is concerned with detecting the author by revealing the unconscious and idiosyncratic linguistic patterns of choice of the text's writer (Baker 2000:264). Baker makes a distinction between literary stylistics, which focus on what is "assumed to be conscious linguistic choices on the part of the writer," and *forensic stylistics*, which focuses on "quite subtle, unobtrusive linguistic habits which are largely beyond the conscious control of the writer and which we, as receivers, register mostly subliminally" (2000:246). These linguistic and stylistic "habits" of the original author are mostly unnoticed and usually left unchanged by plagiarizers when copying someone else's work. Therefore, tracing translator's unconscious use of language helps collect pieces of evidence in retranslational plagiarism cases when a

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

plagiarizer replaces words with their synonyms, for instance. The unconscious use of language can be revealed by tracing patterns in the use of Function Words (FW), which are “words having little semantic content of their own, such as prepositions, conjunctions, and articles” (Coyotl-Morales et al. 2006:844). In this context, Li, Zheng, and Chen (2006) argue that tracing FW in a specific text help reveal the author’s linguistic fingerprint or unconscious use of language. The conscious use of language, on the other hand, can be traced by revealing patterns in the use of content words that have meaning such as adjectives and nouns in a specific corpus.

In this study, I drew on *WordSmith Tools* (Scott 2015), a corpus analysis tool used to process texts to retrieve patterns in language use. The tool provides several corpus analyses such as *wordlists*, *concordance lines*, and word *n-grams* which can be used as forensic pieces of evidence to investigate plagiarism in retranslation. Two functions in *WordSmith* are used to reveal potential unusual overlap in the analyzed retranslations: *wordlist* and *concordance*. The first function provides a list of words or word *n-gram* contained in a specific text along with their frequencies. Word *n-grams* (word clusters or lexical bundles) are “words found repeatedly together in each other’s company, in sequence” (Scott 2010). Word *n-gram* is thus the sequence of words in a text or speech in *n* size; for instance, a *3-word gram* list would include a list of a three-word sequence in a specific corpus and their frequency. The second function, concordance, provides the context of each word or word *n-gram* in a *wordlist* retrieved from a specific corpus.

As the first study to investigate retranslation plagiarism using corpus analysis and forensic stylistics, I use this method to focus on revealing the translators’ unconscious use of language in disputed retranslations by means of content words and function word *n-grams*. It is also the first study that makes use of corpus tools and computerized methods of analysis to reinvestigate the Payne and Burton plagiarism controversy over the translation of *Arabian Nights*, one of the most iconic literary works. Ultimately, the proposed approach can be adopted as a methodological framework to investigate literary retranslational plagiarism cases. The following sections present the data collection and analysis methods using corpus tools and *forensic stylistics* in order to reveal the extent to which Burton’s translation is a copy of Payne’s work.

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

## Method and Data Analysis

The data for this study consists of two corpora of tales taken from the beginning (*Volume 1*) and the middle (*Volume 6*) of three English translations of the *Arabian Nights*. The first corpus, “The Hunchback,” contains three translations of “The Hunchback’s Tale” and other tales taken from *Volume 1* translated by John Payne (1882), Richard Burton (1885), and Jonathan Scott (1811) (See Table 1). Like Payne and Burton, Scott was also an orientalist, author, and a translator. He was known for his translation of *Arabian Nights*. As the translations by Payne and Burton relied on the same source text, Scott’s translation is used in this study as a reference corpus for the overlap comparison between the disputed versions (Payne’s and Burton’s translations). Other translations of *Arabian Nights* were ignored in favor of Scott’s (1811) translation because his was published, relatively, during the same period as Payne’s (1882) and Burton’s (1885) translations. It is also used to see the extent to which the disputed translations overlap with Scott’s translation and whether the overlap in the disputed versions is unusual. The second corpus contains three translations of “Sinbad Seven Voyages” by the same translators, taken from *Volume 6* (See Table 1). The first corpus was selected based on the findings of Şahin et al. (2015:208): that plagiarizers attempt to avoid plagiarism at the beginning of retranslations. The second corpus was thus selected because plagiarism is more likely to happen and to be more obvious in the middle section of a disputed translation. Both samples of the texts would illustrate the case study and the viability of the proposed method.

Table 1: Corpus Length in Words

Translator	First corpus		Second corpus	
	“The Hunchback’s Tales” Length in Words	Whole Volume Length in Words	“Sinbad Seven Voyages” Length in Words	Whole Volume Length in Words
<b>John Payne (1882)</b>	34,549	149,174	25,048	127,945
<b>Richard Burton (1885)</b>	38,969	131,023	32,572	116,686
<b>Jonathan Scott (1811)</b>	43,766	125,439	19,600	125,440

When conducting quantitative analysis to check the overlap between the disputed retranslations, the researcher should decide on the *similarity threshold*, which is a predefined percentage of

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

overlap above which the overlap between retranslations is considered ‘suspicious’, i.e., there is a chance that one translator plagiarized from the other one. In her pioneering study, *Textual kidnapping revisited: the case of plagiarism in literary translation*, Turell (2004) has set the *similarity threshold* to 70%. She argues that this threshold would help discover strong text relationships between the disputed retranslations. The same *similarity threshold* will be adopted in the present study, too. However, more studies are needed to confirm this value.

### *Quantitative Analysis*

The data analysis in this paper is built on the notion of *forensic stylistics* and develops a two-staged target-text-oriented approach for investigating plagiarism in retranslations. The first stage involves a corpus quantitative analysis of disputed translations to retrieve the following lists:

1) lists of *content words with only one occurrence*: This corpus analysis retrieves a list of words used only once by Payne in his translation and traces their frequency of occurrence in the other two translations. In corpus analysis, a *list of content words with only one occurrence* is derived from a longer *wordlist*. This longer list contains all words included in a corpus along with their frequencies arranged, in most cases, alphabetically. Then the researcher looks at the words with only one occurrence and builds a sub-list containing words occurring only once in the corpus. The present study randomly selects and analyzes a list of 10 words that occur only once in Payne’s translation and traces their occurrence in the other translations (See table 3). This is in recognition that it is very difficult to manually retrieve a list of words that occur only once in a specific text and compare their occurrence in other texts across a large disparate corpus, such as *Arabian Nights*.

2) list of *content word n-grams*: As explained earlier, content word *n-gram* is a sequence of words in a text or speech in *n* size. For instance, word 2-gram is the sequence of two words, and similarly, 3-gram is the sequence of three words. For example, a 2-gram analysis of the following sentence *Sam lives in Boston and Jim lives in NY* would reveal the following list:

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

Table 2: Example of Word 2-gram Analysis

<b>2-gram</b>	<b>Frequency</b>
Sami lives	1
lives in	2
in Boston	1
Boston and	1
and Jim	1
Jim lives	1
in NY	1

One important step in conducting word *n-gram* analysis is defining the value of *n* or the size of the cluster in words. In the present study, *n-gram* analysis is applied with  $n = 3$  and  $n = 4$ .

It is worth noting that sometimes content word *n-gram* analysis becomes infeasible when tracing retranslational plagiarism in situations where a plagiarizer copies a text or part of it and applies some changes such as replacing content words with their synonyms. In such cases, the researcher should trace plagiarism through a *forensic* analysis that reveals the unconscious use of language, which is made possible by tracing overlap in the use of function words. While translators and authors consciously pay attention to selecting content words, they unconsciously use function words, their use a pattern that could distinguish them from other authors (Houvardas and Stamatatos 2006; Stamatatos 2013; Li, Zheng, and Chen 2006, Al-Batineh 2015).

3) list of *function word bi-grams (2-gram)*: FW *bi-grams* have been used by several scholars as style markers to reveal the linguistic fingerprint of a document producer (Coyotl-Morales et al. 2006; Houvardas and Stamatatos 2006; Stamatatos 2013; Li, Zheng, and Chen 2006). Linguistic fingerprint in this context means the idiosyncratic use of language manifested in a set of style markers that makes the writing of a specific author different from that of other authors. As discussed earlier, FWs are considered style markers that are topic independent and are not subject to the conscious control of the text author.

### *Qualitative Analysis*

The second stage of analysis includes analyzing concordance lines retrieved from the disputed translations using the concordance search function in *WordSmith* tools. This function allows the

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

researcher to look for a term within its context in the sentence. The qualitative analysis is applied only if the overlap percentage of the traced evidence exceeds the predefined *similarity threshold* (>70%). In other words, the context of one *word with one occurrence*, one *function word n-gram*, and one *content word n-gram* will be retrieved and analyzed using concordance function in *Wordsmith*. The following sections report the results of the quantitative and qualitative analysis of data.

## Findings

### *Quantitative Analysis*

The findings presented in this section reveal the overlap percentage of *words with only one occurrence*, *3- and 4-word n-grams*, and *function words 2-grams* in the three translations produced by Payne, Burton, and Scott.

### *Words with only One Occurrence*

Using *wordlist* function in *Wordsmith*, wordlists containing content words and their frequencies in each retranslation were retrieved. The percentage of overlap of 10 words that occurred only once in Payne’s translation and in the other two translations was calculated (See Table 3). A sample of these words is presented in Table 4.

Table 3: Overlap Percentage of *Words with only One Occurrence*

<i>Comparison</i>	“The Hunchback”	“Sinbad”
	<i>Words with only One Occurrence Overlap</i>	<i>Words with only One Occurrence Overlap</i>
<i>Payne Vs. Burton</i>	90%	100%
<i>Payne Vs. Scott</i>	20%	3%
<i>Scott Vs. Burton</i>	25%	3%

After comparing Payne’s translation to the other two translations in terms of the overlap in *words with only one occurrence*, we notice that the overlap is very high in the Payne-Burton comparison. As indicated in Table 2, the overlap percentage in Payne-Burton comparison in the

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

first corpus (“The Hunchback”) is 90%. This means 9 out of the 10 words that were used only once by Payne in producing his translation were also used only once by Burton in his translation of the same sections. Tracing the same piece of evidence in the second corpus (“Sinbad”) revealed a very high overlap in the Payne-Burton comparison (100%). This means that the 10 words that were used only once by Payne in producing his translation of “Sinbad” are also used once by Burton in rendering the same tales. On the other hand, the overlap percentage in Payne-Scott and Scott-Burton comparison is lower and does not exceed the predefined the pre-set *similarity threshold* (>70%) in the comparisons of the first and second set of data.

Table 4: A Sample of *Words with only one Occurrence* in Payne’s Translation and Their Score in Burton’s and Scott’s Translations

<b>“The Hunchback”</b>		
<i>Words with only One Occurrence in Payne’s Translation</i>	Number of Occurrences in Burton’s Translation	Number of Occurrences in Scott’s Translation
Wottest	1	0
Wrath	1	2
Wrist	1	0
Written	2	1
Wrong	1	2
Wrought	2	2
Yielded	1	0
Yonder	1	0
Youngest	1	0
Zephyr	1	0
<b>“Sinbad”</b>		
<i>Words with only one occurrence in Payne’s Translation</i>	Number of Occurrence in Burton’s Translation	Number of Occurrence in Scott’s Translation
Theirs	1	0
Thirsty	1	3
Thrust	1	0
Tomb	1	0
Train	1	0
Vinegar	1	0
Vintages	1	0
Vipers	1	0
Weep	1	0
Woe	1	0

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

The high percentage of overlap in the Payne-Burton comparison in both corpora raises a red flag. Revealing the context of the above words in both retranslations would be very useful here to make the point clear. For instance, Payne uses the word "**vipers**" once in his translation of "Sinbad" in the following context:

...whereon neither iron nor steel hath effect, neither can we cut off aught therefrom nor break it, save by means of the leadstone. Moreover, the valley swarmed with huge snakes and **vipers**, as big as palm-trees

Burton also used the same word only once in his translation and in the same context:

... whereon neither iron nor hardhead hath effect, neither can we cut off aught therefrom nor break it, save by means of leadstone. Moreover, the valley swarmed with snakes and **vipers**, each big as a palm tree

It can be noticed that in this instance Burton's translation overlaps, to a great extent with Payne's. As mentioned earlier, this section was selected to provide quantitative analysis of the overlap between Payne and Burton's translation. More examples taken from both translations will be displayed and discussed in the Qualitative analysis section.

### *Word 3- and 4-gram*

The second traced piece of evidence is the overlap in content words *n-grams*. An examination of the overlap in larger chunks of texts from the three translations reveals that the overlap in word 3- and 4-grams (3- and 4-word sequence) between Payne's and the other two translations is very low in "The Hunchback" corpus. It is clear, however, that the highest score of overlap was in Payne-Burton comparison (Table 5).

Table 5: Overlap Percentage of Words 3- and 4-gram

<i>Comparison</i>	<i>"The Hunchback"</i>		<i>"Sinbad"</i>	
	Word 3-gram Overlap %	Word 4-gram Overlap %	Word 3-gram Overlap %	Word 4-gram Overlap %
<i>Payne Vs. Burton</i>	47.12%	42.76%	84.70%	82%
<i>Payne Vs. Scott</i>	25.51%	12.32%	17.58%	0
<i>Scott Vs. Burton</i>	19.66%	15.62%	17.62%	0

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

By contrast, the overlap percentages of word 3- and 4-grams in the Payne-Scott and Scott-Burton translations included in the second corpus was 0% respectively.

Table 6: Sample of Words 3- and 4-gram from Payne’s Translation of “The Hunchback” and their Overlap in Burton’s and Scott’s Translations

<i>Word 3-gram</i>						
<b>Word 3-gram</b>	Frequency in Payne	Frequency in Burton	Frequency in Scott	Payne Vs. Burton 3-gram Overlap	Payne Vs. Scott 3-gram Overlap	Scott Vs. Burton 3-gram Overlap
<i>and said to</i>	53	32	7	60.38%	13.21%	21.88%
<i>o my lord</i>	42	29	0	69.05%	0.00%	0.00%
<i>said to me</i>	37	16	3	43.24%	8.11%	18.75%
<i>said to him</i>	39	3	17	7.69%	43.59%	17.65%
<i>chief of the</i>	23	3	0	13.04%	0.00%	0.00%
<i>the old woman</i>	22	25	34	88.00%	64.71%	73.53%
<i>me and i</i>	21	6	22	28.57%	95.45%	27.27%
<i>the chief of</i>	21	6	0	28.57%	0.00%	0.00%
<i>to me and</i>	20	16	6	80.00%	30.00%	37.50%
<i>and went away</i>	19	10	0	52.63%	0.00%	0.00%
<b>Average</b>				<b>47.12%</b>	<b>25.51%</b>	<b>19.66%</b>
<i>Word 4-gram</i>						
<b>word 4-gram</b>	Frequency in Payne	Frequency in Burton	Frequency in Scott	Payne Vs. Burton 4-gram Overlap	Payne Vs. Scott 4-gram Overlap	Scott Vs. Burton 4-gram Overlap
<i>chief of the police</i>	15	0	0	0.00%	0.00%	0.00%
<i>at the end of</i>	12	8	4	66.67%	33.33%	50.00%
<i>o my lord replied</i>	10	4	0	40.00%	0.00%	0.00%
<i>and said to him</i>	17	10	6	58.82%	35.29%	60%
<i>and said to me</i>	16	6	0	37.50%	0%	0%
<i>chief of the police</i>	15	0	0	0%	0%	0%
<i>commander of the faithful</i>	11	13	6	84.62%	54.55%	46.15%
<i>said to me o</i>	10	3	0	30%	0%	0%
<i>she said to me</i>	10	6	0	60%	0%	0%
<i>to me o my</i>	10	5	0	50%	0%	0%
<b>Average</b>				<b>42.76%</b>	<b>12.32%</b>	<b>15.62%</b>

As can be seen in Table 6, the overlap percentage of content word 3- and 4-grams in all the comparison does not exceed the pre-defined *similarity threshold* (>70%).

Table 7: Sample of Words 3- and 4-gram from Payne’s Translation of “Sinbad” and Their Overlap in Burton’s and Scott’s Translations

<i>Word 3-gram</i>						
<i>Word 3-gram</i>	Frequency in Payne	Frequency in Burton	Frequency in Scott	Payne Vs. Burton 3-gram Overlap	Payne Vs. Scott 3-gram Overlap	Scott Vs. Burton 3-gram Overlap
<i>as soon as</i>	23	23	20	100%	86.96%	86.96%
<i>one of the</i>	18	21	10	85.71%	55.56%	47.62%
<i>o my lord</i>	16	23	0	69.57%	0%	0%
<i>the rest of</i>	15	12	5	80%	33.33%	41.67%
<i>there is no</i>	15	26	0	57.69%	0%	0%
<i>to me and</i>	15	20	0	75%	0%	0%
<i>a company of</i>	14	13	0	92.86%	0%	0%
<i>and said to</i>	14	14	0	100%	0%	0%
<i>had befallen me</i>	14	13	0	92.86%	0%	0%
<i>till one day</i>	14	15	0	93.33%	0%	0%
<b><i>Average</i></b>				<b>84.70%</b>	<b>17.58%</b>	<b>17.62%</b>
<i>Word 4-gram</i>						
<i>Word 4-gram</i>	Frequency in Payne	Frequency in Burton	Frequency in Scott	Payne Vs. Burton 4-gram Overlap	Payne Vs. Scott 4-gram Overlap	Scott Vs. Burton 4-gram Overlap
<i>all that had befallen</i>	12	9	0	75.00%	0%	0%
<i>that had befallen me</i>	12	11	0	91.67%	0%	0%
<i>me joy of my</i>	10	9	0	90.00%	0%	0%
<i>one day as i</i>	10	10	0	100.00%	0%	0%
<i>the rest of the</i>	10	9	0	90.00%	0%	0%
<i>till one day as</i>	10	8	0	80.00%	0%	0%
<i>as soon as it</i>	9	6	0	66.67%	0%	0%
<i>gave me joy of</i>	9	7	0	77.78%	0%	0%
<i>made an end of</i>	8	6	0	75.00%	0%	0%
<i>and gave me joy</i>	7	5	0	71.43%	0%	0%
<b><i>Average</i></b>				<b>82%</b>	<b>0%</b>	<b>0%</b>

On the contrary, as noted in Table 7, the overlap percentages of *and said to* and the *one day as I* in Payne’s and Burton’s translation of “Sinbad” is very ‘suspicious’ in that the overlap percentage and their occurrence in the retranslations is 100%.

One instance of *and said to* in Payne translation came in the following context:

When they saw I understood them not and made them no answer, one of them came forward **and said to** me in Arabic, "Peace be on thee, O my brother!

Burton also used the same word 3-gram in the same context, producing an identical translation.

When they saw I understood them not and made them no answer, one of them came forward **and said to** me in Arabic, "Peace be with thee, O my brother!

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

Similarly, the context of *one day as I* overlaps in both translations. The following excerpt includes one instance where *one day as I* occurred in Payne's translation:

I occupied myself thus in exploring the islands till, **one day, as I** stood in the port, with a staff in my hand, according to my wont, I saw a great ship, wherein were many merchants...

The same word 4-gram was also found in the same place in Burton's translation.

I occupied myself thus in visiting the islands till, **one day, as I** stood in the port, with a staff in my hand, according to my custom, behold, a great ship, wherein were many merchants...

The quantitative investigation of word *n-grams* can thus help and guide the search for overlap in disputed retranslations. Quantitative investigation of *n-grams* overlap is not enough to prove plagiarism in retranslation; it should be followed by a qualitative analysis that traces the context of the overlapping *n-grams*. The following section discusses *function words n-grams* as another piece of evidence that can also be viable when investigating plagiarism in literary retranslation.

### *Function Words N-Grams.*

As for the function words analysis, the most frequent FW in Payne's translation was retrieved (it), and this was then used to retrieve its most frequent 10 *bigrams* (2-word sequence containing the function word and the word that follows it, in addition to their frequency in the corpus). As a second step, the occurrences of the ten most frequent *to-bigrams* in Payne's translation were traced in the other two translations (Burton's and Scott's) in order to reveal their overlap percentage. It is worth mentioning that the use of FWs is unconscious and beyond the choice of the translator. They are found to be an idiosyncratic style marker that distinguishes one text's author from others (Stamatatos 2013).

### *Overlap Percentage of to-bigram*

The analysis of *to-bigrams* revealed that the overlap percentage is considerably high in Payne’s and Burton’s translations of “The Hunchback” and “Sinbad,” 83.97% and 88.09% respectively (Table 8).

Table 8: Overlap Percentage of *to-bigram*

<i>Comparison</i>	“The Hunchback”	“Sinbad”
	<i>to-bigram</i> Overlap %	<i>to-bigram</i> Overlap %
<i>Payne vs. Burton</i>	83.97%	89.29%
<i>Payne vs. Scott</i>	51.50%	50.61%
<i>Scott vs. Burton</i>	53.47%	44.94%

According to the pre-defined *similarity threshold* (>70%), the overlap percentages in the Burton-Payne comparisons constitute signs of retranslational plagiarism. On the other hand, the overlap percentages in Payne-Scott and Scott-Burton comparisons are lower and do not exceed the predefined *similarity threshold* (>70%). If we look at the sample of *to-bigrams* list provided in Table 8, we can notice that the overlap percentage of *to him bigram* in Payne’s and Burton’s translation of “The Hunchback” is very high compared to its occurrence overlap in the other two comparisons. Similarly, the occurrence overlap of *to thee* is very unusual as the overlap percentage in Payne’s and Burton’s translation exceeds 83%; however, this overlap is 0% in the other comparisons. This high percentage of overlap in the use of *to-bigrams* in Payne’s and Burton’s translation is also found in their translations of “Sinbad.” The overlap in the occurrence of the three *bigrams* listed in Table 8 exceeds 86%, while the overlap does not go beyond 47% in the other two comparison.

Table 9: Sample of *to-bigram* list from Payne’s translation of “The Hunchback” and “Sinbad” and Their Overlap in Burton’s and Scott’s Translations

“The Hunchback”						
<i>to-bigram</i>	Frequency in Payne	Frequency in Burton	Frequency in Scott	Payne Vs. Burton <i>to-bigram</i> Overlap	Payne Vs. Scott <i>to-bigram</i> Overlap	Scott Vs. Burton <i>to-bigram</i> Overlap
<i>to the</i>	171	175	168	97.71%	98.25%	96.00%
<i>to me</i>	103	67	48	65.05%	46.60%	68.66%
<i>to him</i>	80	82	40	98.77%	50.00%	48.78%
<i>to my</i>	47	46	58	97.87%	81.03%	79.31%
<i>to be</i>	34	29	104	85.29%	32.69%	27.88%
<i>to thee</i>	30	36	0	83.33%	0.00%	0.00%
<i>to his</i>	26	22	37	84.62%	70.27%	59.46%
<i>to eat</i>	18	13	11	72.22%	61.11%	84.62%
<i>to thy</i>	16	26	0	61.54%	0.00%	0.00%
<i>to a</i>	15	14	20	93.33%	75.00%	70.00%
<b>Average</b>				83.97%	51.50%	53.47
“Sinbad”						
<i>to-bigram</i>	Frequency in Payne	Frequency in Burton	Frequency in Scott	Payne Vs. Burton <i>to-bigram</i> Overlap	Payne Vs. Scott <i>to-bigram</i> Overlap	Scott Vs. Burton <i>to-bigram</i> Overlap
<i>to the</i>	106	129	91	82.17%	85.85%	70.54%
<i>to me</i>	52	60	19	86.67%	36.54%	31.67%
<i>to him</i>	31	34	13	91.18%	41.94%	38.24%
<i>to my</i>	29	33	9	87.88%	31.03%	27.27%
<i>to a</i>	22	20	12	90.91%	54.55%	54.55%
<i>to his</i>	15	17	11	88.24%	73.33%	64.71%
<i>to them</i>	12	13	6	92.31%	50.00%	46.15%
<i>to us</i>	11	13	5	84.62%	45.45%	38.46%
<i>to this</i>	8	9	7	88.89%	87.50%	77.78%
<i>to and</i>	4	4	0	100%	0%	0%
<b>Average</b>				89.29%	50.61%	44.94%

The quantitative analysis of content and function word wordlists has revealed that the overlap in the Payne-Burton comparisons are the highest and also exceed the pre-defined *similarity threshold*. It might be useful to display a sample of the *to-bigrams* in their contexts within Payne and Burton’s translations. Payne used *to eat bi-gram* in the following context in “The Story of the Hunchback”:

Each took what he wanted and the rest they buried in a corner of the room. Then they set on food and sat down **to eat.**

Burton used the same bi-gram in the same context in his translation of the same story.

Each took what he wanted and they buried the rest in a corner of the room. Then they set on food and sat down, **to eat.**

Mohammed Al-Batineh, *Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton*, 28-61.

It can be clearly noticed here that the overlap is very high between the two sentences above. A similar situation can also be noticed in the following example including *to and* taken from both translations of “Sinbad”:

Payne

but presently the fumes of the wine rising to his head, he became helplessly drunk and his every limb relaxed and he swayed **to and** fro on my back.

Burton

But presently the fumes of the wine rising to his head, he became helplessly drunk and his side-muscles and limbs relaxed and he swayed **to and** fro on my back.

Burton’s translation overlaps with Payne’s translation not only in occurrence and the context of *to and*, but also in the vocabularies that accompany the bigram as well as the word order and syntax. The high percentage of overlap in the use of the function words *bigrams* in Payne and Burton’s translations is more than a coincidence; the words’ context, as displayed in the above example, suggests that the overlap is even greater in terms of syntax and vocabulary. Besides, the overlap in the Payne-Scott and Burton-Scott comparisons did not exceed 51%, which is considered normal based on the predefined similarity threshold. The following section takes the data analysis a step further and traces the context of the overlapped style markers in Payne and Burton’s translations of “The Hunchback” and “Sinbad”.

### *Qualitative Analysis: “The Hunchback”*

As mentioned earlier, using the concordance search function in *WordSmith* tools, only a sample of the entire qualitative analysis will be presented in this section, including the context of one *word with one occurrence*, one *function word n-gram* and one *content word 3-gram*. Using the results in the previous quantitative analysis section, the qualitative overlap has been traced for the pieces of quantitative evidence exceeding the pre-defined *similarity threshold (>70%)*. Thus, only two pieces of evidence, *words with only one occurrence* and *FW bigram*, will be qualitatively analyzed and presented in this section. In other words, the contextual overlap of one

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

*word with one occurrence* and one *FW bigram* taken from Payne's and Burton's translations of the first corpus ("The Hunchback") and the context overlap of one *word with one occurrence*, one *word 3-gram*, one *word 4-gram*, and one *FW bigram* taken from Payne's and Burton's translations of the second corpus ("Sinbad") are analyzed.

The qualitative analysis attempts to reveal the extent to which the selected concordances from Payne's translation are similar to those of Burton's in terms of lexical items. Turell notes:

when a substantial proportion of these linguistic units in two texts under comparison are the same, one should suspect that either those texts have been produced by the same author, or that plagiarism has occurred between one text and the other and they both share a common source. (2012:539)

If the concordances of Burton's and Payne's work fit within this definition, the proportion of occurrence would affirm the claim that Burton's translation demonstrates an act of plagiarism.

### *Words with only One Occurrence*

The retrieved *wordlist* of *words with only one occurrence* has revealed that the underlined word (*wottest*), in the concordances below, is used only once by both Payne and Burton in their translations, while it is not used by Scott in his translation of the same parts of *Arabian Nights*. The occurrence of the archaic term, *wottest*, used only once in Payne's and Burton's translations in the same exact context is very suspicious.

**Payne** P.1. "...me with the hunchback that thou wottest of. This then is my story and peace be on thee.' Quoth the King"

**Burton** B.1 "...me with the Hunchback that thou wottest of. This then is my story, and peace be with thee! Quoth the King"

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

It can be noticed from P.1 (by Payne) and B.1 (by Burton) that the two translators use almost identical words to produce the translations. The only difference between Burton's and Payne's versions, as can be noticed, is in the use of one preposition. The overlap in the word order, syntax and vocabulary of the target text and in the use of *wottest* by both translators in the same context and position does illustrate an instance of plagiarism.

### *To-bigrams*

In the following concordances, the underlined *FW* bigram (*to thee*) appears 30 times in Payne's translation, 36 times in Burton's, and 0 times in Scott's.

**Payne** P.2 Wherefore it behoves thee to give thanks to God and not cross me, for I am a true counsellor to thee and take an affectionate interest in thee.

**Burton** B.2 Wherefore it behoveth thee to render thanks to Allah Almighty and not cross me, for I am a true counsellor to thee and benevolently minded towards thee.

The linguistic overlap in these concordances (P.2 by Payne and B.2 by Burton) is unusual in terms of word choice, punctuation marks, and use of underlined function words. It seems that Burton tried to hide his plagiarism by adding a foreign flavor to his translation and replaced *God* with *Allah Almighty*, a literal translation of the Arabic word for God (*Allah*). Noticeable is the use of the underlined function word bigram (*thee to*) being identical in P.2 and B.2. This analysis was applied to a set of stories taken from the first volume of *Arabian Nights*. The qualitative and quantitative analyses show that Burton's translation, to a great extent, overlaps with Payne's translation. The following section similarly utilizes traced style markers and investigates another set of data taken from the previously traced pieces of evidence in the middle section of *Arabian Nights*. This may include a higher degree of overlap since plagiarizers tend to plagiarize the middle of the work and avoid such activity in the beginning of retranslations (Şahin et al. 2015).

### *Qualitative Analysis: "Sinbad"*

#### *Words with only One Occurrence*

The underlined word *thrust* is used only once by Payne in his translation of the "Seven Voyages of Sinbad" once by Burton, and three times by Scott. Using *WordSmith* tools, the context of *thrust* in the disputed translations is revealed as follows:

**Payne** P.4 as he lay snoring on the bench, thrust them into his eyes and pressed upon them, all of us, with our might, so that his eyes were put out and he became blind.

**Burton** B.4 as he lay snoring on the bench, thrust them into his eyes and pressed upon them, all of us, with our united might, so that his eyeballs burst and he became stone blind.

These concordances show that *thrust* appears in Payne's (P.4) and in Burton's (B.4) versions in the same context, while it appears in three different places in Scott's translation. The above concordances also indicate that Burton's translation is a reproduction of Payne's in this particular example. In his translation, Burton applied some changes to Payne's, either to hide his plagiaristic act or to impose his stylistic preference on the plagiarized parts. Burton added one word *united* and replaced two *eyeballs* and *burst* in his version. It can be also noticed that the use of punctuation marks is also identical in the two translations

#### *Word 3-Grams*

The two concordances in the following table provide the context of one *word 3-gram*, (*had befallen me*), retrieved from Payne's and Burton's translations. This particular *word 3-gram* is used 14 times by Payne and 13 times by Burton, but it does not appear in Scott's translation.

**Payne** P.5 I sat on the beach, recalling all that had befallen me and saying, "I wonder if God will save me alive and restore me to my country and my friends!"

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

**Burton**                    B.5 I sat on the beach, recalling all that had befallen me and saying, "I wonder if *Allah* will save me alive and restore me to my *home* and *family* and friends!

These two concordances (P.5 by Payne and B.5 by Burton) are similar in structure, word choice, and use of punctuation marks. They also overlap in the use of the same word 3-gram in the same context which is, an unusual overlap. In his version, Burton applied some changes to hide his plagiaristic act and add his stylistic preferences. Burton used *Allah* instead of *God*, again adding a foreign flavor to his translation. He replaced *country* with a synonym *home* and added *family*.

#### *Word 4-Gram*

The two concordances below display the context of one word 4-gram (*me joy of my*) in Burton's (B.6) and Payne's (B.6) translations. *Me joy of my* is used 10 times by Payne, 9 times by Burton, and 0 times by Scott.

**Payne**                    P.6 whereupon both he and the merchants were certified of the truth of my story and recognized me and gave me joy of my deliverance, saying, "By Allah, we thought not that thou hadst escaped drowning!

**Burton**                    B.6 whereupon both he and the merchants were certified at the truth of my story and recognized me and gave me joy of my deliverance, saying, "By Allah, we thought not that thou hadst escaped drowning!

It is noticeable that Burton copied Payne's translation word for word, even comma for comma. The use of punctuation marks, which is considered a personal attribute in writing or translating, is identical in these two concordances. The only difference in the two translations, as can be seen, is in the use of one preposition where Burton replaced *of* with *at*.

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

### *To-bigram*

The last qualitatively traced piece of evidence is the context of one *to-bigram* in the disputed translations. The following concordances display the occurrence of *to them* in the context appearing 12 and 13 times in Payne's and Burton's translations and 6 times in Scott's.

**Payne**                    P.8 I filed them smooth and tinned them. Moreover, I made fast to them fringes of silk and fitted bridle-leathers to the bit. Then I fetched one of the best of the royal horses and saddling and bridling him...

**Burton**                    B.8 I filed them smooth and tinned them. Moreover, I made fast to them fringes of silk and fitted bridle-leathers to the bit. Then I fetched one of the best of the royal horses and saddling and bridling him...

These two concordances by Payne (P.8) and by Burton (B.8) exhibit an unusual overlap in word choice, punctuation marks, and use of conjunction and function words, Burton's version is a mere copy of Payne's translation without any addition, deletion, or modification in the latter's version. The qualitative and quantitative analyses of the second set of data thus reveal that Burton's translation is largely a reproduction or a copy of Payne's translation. The plagiaristic act in the second set of data is more evident and clearer compared to the first set of data.

## **Conclusion**

The present paper proposed a corpus-based target-text-oriented approach for detecting and investigating plagiarism in retranslation, blending corpus quantitative and qualitative analyses. The quantitative analysis was used as a first step to reveal lists of 'linguistic clues,' which constitute an overlap in the use of some individual lexical items or a series of lexical items. This first step leads to the second step of using concordance search to reveal the context of lexical items included in the wordlists. The second step involves analyzing concordance lines retrieved from the disputed translations. The suggested combination of qualitative and quantitative

analyses has provided, alternative points of evidence which propose new conclusions on the plagiarism case, Burton's vs. Payne's translations of *Arabian Nights*.

One conclusion I put forward is that by investigating the Payne and Burton controversy using this methodology, the claim made by Wright (1904) that Burton plagiarized in his translation of the tales is certainly not unfounded. In the initial set of data, (taken from the first volume of *Arabian Nights*) the corpus analysis of *content words n-grams* lists did not initially reveal significant overlap percentage in the disputed versions. As the concordance qualitative analysis has revealed, Burton opted for replacing words with their synonyms to hide his plagiaristic act. This leads to another conclusion: that investigating retranslational plagiarism by merely analyzing the overlap in the frequencies of vocabularies or phrases included in a *wordlist* is not enough unless the analysis is supported by contextual overlap analysis which can be easily carried out using the concordance search function in corpus tools.

The analysis of function words to reveal the unconscious use of language has also served as a strong piece of evidence by revealing a 'suspicious' overlap percentage in the first set of data. That is, the presence of Payne's unconscious use of language manifested in the use of function words, and their occurrence in context in the same place as Burton's translation is unusual. The qualitative results reached by the quantitative analysis of *FW bigrams* also revealed overlap in the use of function words.

The quantitative corpus analysis of wordlists of the second set of data ("Sinbad"), taken from the middle volume of *Arabian Night*, has also revealed a very high percentage of overlap in Payne's and Burton's translations in terms of both content and function words. The overlap was clearer and easier to detect in "Sindbad". This conclusion then is similar to that of Şahin et al. (2015: 208) in contexts of their research: plagiarism is often more obvious in the middle section of the plagiarized translations analyzed. Bringing these conclusions together can also help us argue that translation is, in fact, a form of writing whose target text is a stand-alone text containing idiosyncratic style markers that can distinguish one translator from another and which can be traced to investigate plagiarism in retranslation. This conclusion is important in the area of

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

retranslational plagiarism as it supports the argument that the investigation of the such cases can be target-oriented. In other words, analyzing the disputed translations alone, to spot instances of overlap, without referring to the original text, could help reveal potential cases of plagiarism.

There are two limitations to the present study which can be considered as venues for further investigation. First, the paper did not analyze a third set of data taken from the final volume of *Arabian Nights*. Another question inviting further research relates to determining the baseline or overlap cutoff percentage between the analyzed retranslations where the overlap is considered ‘unusual’. This would help translation researchers draw a line between what is considered ‘normal’ or ‘unusual’ overlap in plagiarism cases since all retranslations often attempt to reflect the content and style of the original and so a percentage of overlap is likely to be found in most cases. In conclusion, this study has revealed that more research is needed in the area of plagiarism in translation to set the boundaries of this unethical translational act. It is an important issue in a highly connected world where the availability, access, and manipulation of content, be it digital or printed, make it easier for plagiarizers to claim ownership or translatorship of texts.

## **Author Details**

Mohammed Al-Batineh, Yarmouk University, Irbid, Jordan  
PO Box: 22, Postal Code: 21163  
Mobile: (+962) 798 093 551  
Email: [M\\_bataineh@yu.edu.jo](mailto:M_bataineh@yu.edu.jo)

## **Acknowledgments**

I am immensely grateful to Professor Françoise Massardier-Kenney, Dr. Erik Angelone, Dr. Loubna Bilali, Dr. Haru Dewi and Mr. Riyad Tubaishat for their comments on earlier versions of the manuscript which improved it significantly. I would also like to extend my gratitude to the anonymous reviewers and to Dr. Ruth Abou Rached for her help and insights during the review process for *New Voices in Translation Studies*.

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

## References

- Al-Batineh, Mohammed (2015) *Latent Semantic Analysis, Corpus stylistics and Machine Learning Stylometry for Translational and Authorial Style Analysis: The Case of Denys Johnson-Davies' Translations into English*. Unpublished Doctoral Dissertation, Kent State University: Ohio, USA. Available online at [[http://rave.ohiolink.edu/etdc/view?acc\\_num=kent1429300641](http://rave.ohiolink.edu/etdc/view?acc_num=kent1429300641)] (accessed 1 January 2017)
- Al-Olaqi, Fahd (2012) 'The Influence of the Arabian Nights on English Literature: A Selective Study' *European Journal of Social Sciences*, 31(3), 384–396.
- Baker, Mona (1995) 'Corpora in Translation Studies: An Overview and Some Suggestions for Future Research' *Target*, 2(7): 223–43.
- Baker, Mona (2000) 'Towards a Methodology for Investigating the Style of a Literary Translator' *Target*, 12(2): 241–66. DOI:10.1075/target.12.2.04bak.
- Barnbrook, Geoff (1996) *Language and Computers: A Practical introduction to the Computer Analysis of Language*, Edinburgh: Edinburgh University Press.
- Benjamin, Walter (2000) 'The Task of the Translator; in Lawrence Venuti (ed) *The Translation Studies Reader*. London: Routledge.
- Berman, Antoine (2010) 'Translation and the Trials of the Foreign', in Lawrence Venuti *The Translation Studies Reader*, New York and London: Routledge, 284–297.
- Brodie, Fawn (1984) *The Devil Drives: A Life of Sir Richard Burton*, New York: W.W. Norton & Company Ltd.
- Burrows, John (1992) 'Not Unless You Ask Nicely: The Interpretative Nexus between Analysis and Information'. *Literary and Linguistic Computing*, 7(2), 91–109.
- Burton, Richard (1885) *The Book of the Thousand Nights and a Night*, London: Kama-Shastra Society.
- Campbell, Joseph (1952) *The Portable Arabian Nights*, New York: Viking Press.
- Coyotl-Morales, Rosa María, Luis Villaseñor-Pineda, Manuel Montes-y-Gómez, and Paolo Rosso (2006) 'Authorship Attribution Using Word Sequences' In José Francisco Martínez-Trinidad, Jesús Ariel Carrasco Ochoa, and Josef Kittler (eds) *Progress in Pattern Recognition, Image Analysis and Applications. CIARP 2006. Lecture Notes in Computer Science*, Heidelberg: Springer, 4225:844–53.
- Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

- Herrag, El Hassane (2012). *The ideological factor in the translation of sensitive issues from the Qurān into English, Spanish and Catalan*. Unpublished Doctoral Dissertation, The Autonomous University of Barcelona, Barcelona, Spain. Available online at [<https://core.ac.uk/download/pdf/19576529.pdf>] (accessed 1 September 2018).
- Houvardas, John, and Efstathios Stamatatos (2006) 'N-Gram Feature Selection for Authorship Identification' in Jérôme Euzenat and John Domingue (eds) *Artificial Intelligence Methodology Systems and Applications: Lecture Notes in Computer Science*, Heidelberg: Springer: 77–86. DOI:10.1007/11861461\_10.
- Jockers, Matthew (2013) *Macroanalysis: Digital Methods and Literary History*, Illinois: University of Illinois Press.
- Khosravi, Habibeh. & Pourmohammadi, Majid. (2016). 'Influence of Translator's Religious Ideology on Translation: A Case Study of English Translations of the Nobel Quran', *International Journal of English Language & Translation Studies* 4(4): 151-163.
- Leighton, Lauren (1994) 'Translation and Plagiarism: Puškin and D. M. Thomas', *The Slavic and East European Journal* 38(1): 69–83.
- Li, Jiexun, Rong Zheng, and Hsinchun Chen (2006) 'From Fingerprint to Writeprint', *Communications of the ACM* 49 (4): 76–82. DOI:10.1145/1121949.1121951.
- Lovell, Mary (2000) *A Rage to Live: A Biography of Richard and Isabel Burton*, New York: W.W. Norton & Company Ltd.
- McMenamin, Gerald (2002) *Forensic Linguistics: Advances in Forensic Stylistics*, London: CRC Press.
- Penzer, Norman (1997) *An Annotated Bibliography of Sir Richard Francis Burton*, Connecticut: Maurizio Martino.
- Payne, John (1882) *The Book of the Thousand Nights and One Night*, London: The Villon Society for Private Subscription and for Private Circulation.
- Sabouri, Hossein (2011) 'The Orientalist Readings of The Arabian Nights' *Ferdowsi Review*, 1(4), 123–132.
- Saldanha, Gabriela (2011) 'Translator Style: Methodological Considerations' *The Translator*, 1(17), 25–50.

*Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*

- Şahin, Mehmet, Derya Duman, and Sabri Gürses (2015) 'Big Business of Plagiarism under the Guise of ( Re )Translation: The Case of Turkey', *Target* 2(61): 193–218.
- Şahin, Mehmet, Derya Duman, Sabri Gürses, Damla Kaleş, and David Woolls (2019) 'Toward an empirical methodology for identifying plagiarism in retranslation', In Ö. Berk Albachten & Şehnaz Tahir Gürçağlar (eds), *Perspectives on Retranslation*. London/New York: Routledge, 166–191.
- Scott, Jonathan (1811) *The Arabian Nights Entertainments*, London: Pickering and Chatto.
- Scott, Mike (2015) *WordSmith Tools Users' Manual Version 6.0*. Liverpool: Lexical Analysis Software Ltd.
- Scott, Mike (2010) *WordSmith Tools*, Liverpool: Lexical Analysis Software.  
[http://www.sprak.gu.se/digitalAssets/1361/1361039\\_wordsmith.pdf](http://www.sprak.gu.se/digitalAssets/1361/1361039_wordsmith.pdf).
- Sinclair, John (2003) *Collins COBUILD advanced learner's English dictionary*. Glasgow: HarperCollins Publishers.
- Stamatatos, Efstathios (2013) 'On the Robustness of Authorship Attribution Based on Character N-gram Features' *Journal of Law & Policy* (21)2: 421–39.
- Tahir Gürçağlar, Şehnaz (2009) 'Retranslation' in Mona Baker and Kirsten Malmkjær (eds) *Routledge Encyclopedia of Translation Studies*, London: Routledge, 232-236.
- Turell, Teresa (2004) 'Textual Kidnapping Revisited: The case of Plagiarism in Literary Translation', *International Journal of Speech, Language and the Law*, 11(1), 1–26.
- Turell, Teresa (2012) 'Detection Approaches to Plagiarism in Literary Translation', in Juan J. Lanero Fernández, José Luis Chamosa (eds), *Language, Translation, Reception to Honor Julio César Santoyo* León: Universidad de León, 535–559.
- Venuti, Lawrence (2004) 'Retranslations: The Creation of Value', *Bucknell Review* 47 (1): 25–38.
- Wallen, John (2014) 'The "Terminal Essay" to Burton's Arabian Nights: Tasting The Forbidden Fruit', *The Victorian*, 2(2), 1–28.
- Wright, Thomas (1904) *The Life of John Payne*. London: T. Fisher Unwin.
- Wright, Thomas (1906) *The Life of Sir Richard Burton*. New York and London: Putnam's Sons and Evertt & Co.
- Mohammed Al-Batineh, Using Forensic Stylistics and Corpus Analysis for Detecting and Investigating Plagiarism in Literary Retranslation: The Plagiarism Controversy of John Payne and Sir Richard Burton, 28-61.*