A Corpus Stylistics Approach to the Characterisation of Magwitch and Miss Havisham in Dickens’s *Great Expectations*

Alumna: Antziñe Aramburu Zabala

Tutora: Izaskun Elorza Amorós

Salamanca, 2016
A Corpus Stylistics Approach to the Characterisation of Magwitch and Miss Havisham in Dickens’s *Great Expectations*

This thesis is submitted for the degree of English Studies

Date 4th July

Tutor: Izaskun Elorza Amorós

Vº Bº

Signature
ABSTRACT [EN]

The aim of this essay is to provide a corpus-based analysis of two characters in Dickens’s *Great Expectations* in order to disclose which purpose they serve in the novel. Taking into account that the novel is a *Bildungsroman*, special attention has been paid to how the chosen characters, Magwitch and Miss Havisham, are related to the main character, Pip. By embracing an innovative approach to study literature, namely corpus stylistics, not only are we able to support results that previous analyses have already provided, but also to disclose patterns that, even through close reading, would pass unnoticed. In this sense, this paper explores how corpus stylistics can contribute to Dickens’s studies.


RESUMEN [ES]

El objetivo de este ensayo es presentar un análisis basado en corpus de dos personajes de la novela *Grandes Esperanzas* de Dickens, para descubrir la función de ambos dentro de la novela. Teniendo en cuenta que la novela es un *Bildungsroman*, se ha prestado especial atención a cómo los dos personajes escogidos, Magwitch y la Señorita Havisham, están relacionados con el personaje principal, Pip. El innovador campo de la denominada estilística de corpus no solo permite confirmar los datos proporcionados por la crítica literaria, sino que revela detalles que, incluso a través de una lectura en profundidad, pasarían inadvertidos. En este sentido, este ensayo examina cómo la estilística de corpus puede contribuir a los estudios sobre Dickens.

AKNOWLEDGEMENTS

To my tutor, Izaskun Elorza, whose help has been essential throughout the process of writing this paper. I am immensely grateful for her commitment, advice and critical opinion, as well as for her constant support and positivity, which has made everything easier.

To my family, who encourages me to follow my dreams and never give up.

To all those who have helped me in difficult times.
TABLE OF CONTENTS

1. Introduction...................................................................................................................4

2. Framework of the study

2.1. Corpus linguistics and literature..................................................................................4

2.1.1. Defining corpus linguistics.....................................................................................4

2.1.2. Corpus stylistics......................................................................................................5

2.2. Corpus stylistics in Dickens’s *Great Expectations*: characterisation of Magwitch
    and Miss Havisham

2.2.1. Describing Dickensian characters.........................................................................7

2.2.2. The construction of Magwitch and Miss Havisham in Dickens’s *Great
        Expectations*.............................................................................................................8

2.2.3. Analysing Magwitch and Miss Havisham

2.2.3.1. Corpus compilation..........................................................................................10

2.2.3.2. Results................................................................................................................11

2.2.3.3. Concluding remarks..........................................................................................14

3. Conclusion.....................................................................................................................15

4. Works Cited...................................................................................................................17
1. INTRODUCTION

Deriving from corpus linguistics, a rather innovative field known as corpus stylistics has provided a revolutionary approach to the traditional linguistic analysis. Corpus stylistics deals exclusively with the investigation of how an analysis based on corpus can be applied to literary works and, despite there being a scare number of studies, its contribution to literary criticism has been a major breakthrough.

In this paper I will look closely at how corpus stylistics can be valuable when analysing characterisation in Dickens’s *Great Expectations*. I will focus on two characters that act as doubles, Magwitch and Miss Havisham, and explore their role on the novel, paying close attention to their relationship with Pip, the main character.

2. FRAMEWORK OF THE STUDY

2.1. Corpus linguistics and literature

By defining corpus linguistics and corpus stylistics, this section aims to introduce a framework for the subsequent analysis based on corpus.

2.1.1. Defining corpus linguistics

Despite the fact that several attempts have been made to define corpus linguistics, there is still no consensus and linguists hold divergent views. In “What is *corpus linguistics*? What the data says” Charlotte Taylor distinguishes two main perspectives: whereas some linguists identify it as a methodology, others embrace a more theoretical approach, mainly depending on the ultimate goals they pursue in their analyses.

On the one hand, Leech defines it as a methodology for studying language, and refers to it as a tool. McEnery, Wilson and Meyer have come to a similar conclusion, although they highlight that the methodological practices for such purpose are various and they must be explained by taking into account theoretical considerations. On the
other hand, Stubbs, Sinclair and Teubert describe corpus linguistics as a discipline and, thus, as a part of theoretical linguistics.

Following the theoretical approach, Douglas Biber, Susan Conrad and Randi Reppen have focussed on four main features which are exclusive to corpus linguistics and, thus, helpful to describe it. Firstly, being under the framework of empirical research, it offers objective data. Secondly, it is based on the analysis of a “corpus”, which according to Hunston is an “electronically stored collection of samples of naturally occurring language” (1). Thirdly, it makes use of computers because corpora are stored electronically and computer programmes are necessary for the analysis. Finally, it can provide both quantitative and qualitative data.

2.1.2. Corpus stylistics

Corpus stylistics is the term used to designate research methods based on corpus analysis to study literature. Its importance lies especially in the fact that, although “the findings of corpus stylistics (comparative frequencies, distributions and the like) sometimes document more systematically what literary critics already know (and therefore add to methods of close reading) . . . they can also reveal otherwise invisible features of long texts” (Stubbs 22). Having been developed in the last fifty years, corpus stylistics is an emerging field and, as such, it is yet to be explored further in order to exploit its potential for literary criticism. The few studies which have been published so far demonstrate through practical examples how it can contribute to the analyses of literary works. The path of analysis that each paper follows is different, since “ultimately, the choice of method depends on the text under analysis and the overall aim of the study” (Mahlberg 298), which serves to illustrate the flexibility of this approach.

Among the analysts embracing this approach, Michaela Mahlberg is key in this paper, not only for her work showing the possibilities offered by corpus stylistics, but
also due to her book *Corpus Stylistics and Dickens’s Fiction*. Her studies of Dickens’s narrative include the analysis of keywords, clusters and suspensions. The notion of suspended quotation was created by Lambert, who defined it as “a protracted interruption by the narrator of a character’s speech. And here, ‘protracted’ means containing at least five words” (6). However, Mahlberg has acted as a pioneer in studying suspensions employing corpus-based methods. The reason for this is, as she explains, that “suspensions analyzed in the form of concordances can be useful places to study character information” (Mahlberg 64). In fact, the suspension is a place where the narrator can place information about the speaker or character, either about gestures, expression, tone, pitch, or loudness. The suspensions can easily be analysed by using a corpus tool known as CLiC, which is a collaborative project of the Nottingham and Birmingham universities and enables to search for “specified sections of texts” (52). Although this paper does not provide an analysis of suspensions because of length requirements, it could be improved by searching for repetitive patterns associated to the representational speech of the characters to be studied. In this way, examples such as “cried Miss Havisham, passionately striking her stick upon the floor” (Dickens 284) and “cried Miss Havisham, striking her stick upon the floor and flashing into wrath so suddenly that Estella glanced up at her in surprise” (Dickens 331) offer insights about the behaviour of the character.

Other examples of corpus stylistics are the analyses by Culpeper and Stubbs on *Romeo and Juliet* and *The Heart of Darkness*, respectively. In his paper “Computers, language and characterisation: An analysis of six characters in *Romeo and Juliet*”, Jonathan Culpeper has focussed on the study of syntactic features of characters’ speech because “the accumulative effect of lexical or grammatical features is decisive in shaping an impression of character” (11). Culpeper’s paper includes the analysis of the
most frequent words of six characters in *Romeo and Juliet*, followed by a detailed examination of the keywords. Michael Stubbs, a German linguist at the University of Trier, analysed Conrad’s *The Heart of Darkness*. In this study not only he searches for keywords, but also lemmatises the most frequent verbs and searches for word distribution and collocation.

The progress so far attained demonstrates that corpus stylistics can offer an alternative perspective for the analysis of literature. Besides, Biber suggests that the range of possible applications can be extended since “there seems to be great potential for new lines of research that integrate the statistical methods of earlier research with the more rhetorical concerns of recent studies” (Biber 21). In this sense, this paper not only aims to provide data on how Dickens constructed two characters in *Great Expectations*, but also to explore the way in which these findings reveal their relationship with the main character.

2.2. Corpus stylistics in Dickens’s *Great Expectations*: characterisation of Magwitch and Miss Havisham

This section begins by providing background information on Dickensian characterisation and continues with the study of two characters in *Great Expectations*. Eventually, the paper provides an analysis based on corpus.

2.2.1. Describing Dickensian characters

Although many aspects have been analysed in Dickens’s novels, a large number of the studies, such as the ones of Mahlberg, Lambert, Çakir, Gillman and Patten, and Wentersdorf, have focussed on character analysis. The reason for this is that “one of the achievements that Dickens possesses in connection with the art of his novels is his obvious success in creating memorable characters in his substantial novels” (Çakir 565). Indeed, such is the complexity of the construction of Dickens’s characters that they
seem to disclose or reveal themselves to the reader. As stated in *The Life of Charles Dickens*, written by the friend and biographer of the Victorian novelist, John Forster, Dickens made “characters real existences, not by describing them but by letting them describe themselves” (Forster 274).

Dickens employed various techniques in order to achieve such mastery. Dickensian characters are described either by means of thoughts, actions and interactions with other characters, or also by placing them in a particular setting that confers them a unique aura. In addition, some characters are made even more remarkable by using the *Doppelgänger* technique, which consists on exposing “the conflicting elements in the make-up of a Faust-like character through the introduction of psychological and even physical doubles” (Wentersdorf 203). This narrative style is complicated by further contrasts which give way to endlessly proliferating doubles that throw light on the main character’s personality. Apart from this, linguistic features are also very important in Dickens; for instance, his characters usually have their own idiolect. As Çakir explains, “linguistic writers define idiolect as a definitive word denoting the specific and personal speech quality of an individual, which differentiates individuals from each other like finger prints” (567). Thus, the thorough examination of a character’s idiolect can help the reader become aware of the character’s “social position, regional and ethnic distinctiveness, educational level, and even their real personality, along with their differentiation through bold lines” (Çakir 567).

One example of how characters are constructed is the case of the doubles Magwitch and Miss Havisham.

2.2.2. The construction of Magwitch and Miss Havisham in Dickens’s *Great Expectations*
Both Magwitch and Miss Havisham can be considered doubles. They are, indeed, “tropological doubles” since “they act out moral polarities” (Gillman and Patten 442). The two characters represent the dilemma between two forces: money along with isolation, or humbleness along with love. As the narration moves forward, they guide the main character, Pip, in his process towards maturity and warn him against taking the wrong path by revealing their own mistakes.

It is remarkable how the social position of the characters differs: whereas Miss Havisham belongs to the upper class, Magwitch is a convict and a member of the lower class. However, as Gillman and Patten highlight, the quest that the characters are connected is “a perception others must come to by going beyond dissimilarities of feature or social station to underlying convergences of moral or psychological situation” (445). To begin with, the roles of both characters are mixed because although at first Miss Havisham is supposed to be Pip’s benefactor, in the end readers discover that it is actually Magwitch who performs this function. The link between the characters is prolonged by the fact that both characters “have been denied love, and both attempt to exact revenge because of it” (Selby 48). Whereas Magwitch desires to become a gentleman or someone who is not considered worthless by society, Miss Havisham seeks revenge on men by denying the possibility of love to her adoptive daughter, Estella. They both feel alienated from society and suffer from isolation because Magwitch has had to live in prisons and finally in Australia, while Miss Havisham has been cut off from the light of day ever since she was jilted on her wedding-day. In addition, the two characters share the same enemy: Compeyson. While he abandoned Miss Havisham in the altar, Magwitch got arrested because of him. The characters are also connected through Estella, whose biological father is Magwitch and her adoptive mother Miss Havisham.
The two characters are crucial regarding Pip’s process towards maturity and his choices in life. Indeed, *Great Expectations* being a novel of formation or a *Bildungsroman*, everything centres around the main character, including all the other characters. In this state of affairs, Magwitch and Miss Havisham become “ways of expressing the spectrum of possibilities for character and of enacting alternative futures for the protagonist” (Gillman and Patten 444).

2.2.3. Analysing Magwitch and Miss Havisham

The study of the characterisation of Magwitch and Miss Havisham has mainly followed a corpus-based analysis; in other words, corpus data has been used to explore a hypothesis in order to validate, refute or refine it. However, the corpus itself has also been sometimes used as a direct source by embracing a corpus-driven approach. Indeed, both methods can be complementary ways of obtaining information.

2.2.3.1. Corpus compilation

To begin with, I will briefly explain the steps that I have followed prior to the analysis, as part of the corpus compilation process. I first downloaded the novel from the website of Project Gutenberg, which is frequently accessed by linguists who aim to conduct corpus-based research on literature because it offers open access. Then, I converted the novel into plain text format in order that the software programs created for corpus analysis would easily provide data. Afterwards, I compiled two corpora that would enable me to obtain data about the two characters that I had chosen, namely the MagCorp corpus, with all the running words of Magwitch’s speech along the novel, and the HavCorp corpus, which contains all Miss Havisham’s speech along the novel. Once I had the corpora, I used AntConc, a concordance software program designed by Anthony Laurence, to carry out an analysis of the two characters. The analysis focussed on finding patterns in how they address Pip, which could shed light on their
relationships, by developing wordlists, searching for keywords and concordance, and exploring clusters.

In the next section, I shall present the results that I have obtained using the AntConc program.

2.2.3.2. Results

First of all, I created a wordlist of each corpus to identify the frequency ranking of the words found to occur. The findings showed that the most repeated words in both Magwitch and Miss Havisham’s speech are first and second person pronouns. On the one hand, the MacCorp corpus consists of 1,224 different words (word types) that altogether occur 7,057 times (word tokens). The wordlist of this corpus indicates that, focusing on rank order and with 321 occurrences, the most frequent word is the first person singular pronoun “I”. The log likelihood test, which is used to calculate the type-token ratio of the text, shows that it makes up 4.55% of the character’s whole speech. On the other hand, the HavCorp corpus consists of 516 word types and 2,081 tokens. The first person singular pronoun “I” is the second most common word in Miss Havisham’s speech and it makes up 3.8% of the character’s whole speech. However, attention must be drawn to the fact that, as opposed to Magwitch, the most frequent word by Miss Havisham is the second singular pronoun “you”, with a log likelihood of 5.05%. Although in Magwitch’s speech the use of the second singular pronoun is noticeable as well, “you” ranks fifth with 176 occurrences and a log likelihood of 2.49%. As stated by Biber, “it has been suggested that first and second pronouns are features of interaction” (qtd. in Culpeper 17) and, thus, it is desirable to search for the pronouns in their co-text (i.e. the surrounding words) by using the concordance tool in order to find out how they are meaningful.
The concordance tool provides a contextualised view of the selected item in the corresponding corpus and it is very useful, since “we can only fully understand what is expressed by a word when we see it in combination with other words” (Römer 89). Despite the findings not providing valuable data about the pronouns, from a theoretical perspective, the first person singular pronoun “I” indicates self-referentiality and, according to Selby, its dominance might be related to the deep sense of alienation from society that both characters experience (48). On the other hand, both in Magwitch and Miss Havisham’s speech, the second singular pronoun “you” refers usually to Pip. This proves that they both exert great influence on the boy and also confirms that they are essential in the making of Pip as a character. Despite these findings not being surprising, as has been suggested by Biber, “it is perhaps better not to find anything too new but to confirm findings from many years of traditional study” (3).

The analysis can be taken further by comparing a wordlist with another wordlist derived from a reference corpus. This procedure generates keywords, that is to say, words that are either unusually frequent or unusually infrequent, since keyness is “a matter of being statistically unusual” (Culpeper 4). Keywords are crucial to show the features that can be typically associated to a character. Magwitch’s and Miss Havisham’s keywords were obtained by using Great Expectations as a reference. In this way, the most typical expressions and meanings of each character were identified. On the one hand, the findings obtained from comparing the MagCorp corpus with the reference corpus excluding the character’s speech showed that “boy”, “dear” and “gentleman” were unusually frequent words. To understand the implications of such uncommon frequency rates I used the collocates tool, which finds words that occur near the searchword, usually in a contextual span of four or five words to the left and to the right of it. Among the collocates of “boy”, the most frequent one was “dear”, which is a
“My” and “comrade” were also common. The three of them refer to Pip, the “boy”, and have a positive stance. The adjective “dear” and the first person singular pronoun “my” show an emotional bond between Pip and Magwitch. The collocate “comrade” refers to Pip’s friend, to whom Magwitch calls “Pip’s comrade” rather than by his name, and it underlines the centrality of the main character of the novel. However, the attitude of Magwitch towards Pip changes considerably along the novel and this is shown by the concordance plot tool, which offers a visualisation of the density of the occurrences along the text. When Magwitch and Pip meet for the first time, the convict contemptuously refers to the boy as “young dog” and “you little devil”. Regarding the keyword “gentleman”, it is usually followed or preceded by the collocate “you”, which refers to Pip. Here, it is important to notice that sometimes keywords “indicate dominant topics or themes of a text” (Fischer-Starcke 496) and, thus, they are closely related to the aboutness of the text. In fact, the emphasis that Magwitch puts on making Pip a gentleman is central to the whole novel, since the plot is based on the conflict between love and money, suggesting the impossibility of achieving the two of them. This way, in the process of becoming a gentleman, Pip rejects his own family and feels ashamed of his roots. On the other hand, the results obtained by comparing the HavCorp corpus with the reference corpus excluding the character’s speech were as follows: the second person singular pronoun “you”, the verbs “love” and “play”, and the third person singular pronoun “her” were the most salient words. “You” refers once more to Pip and, indeed, the meaning of the collocates used to refer to Pip are akin to Magwitch’s: “ingrate” and “the sister of this boy” are negatively employed in the beginning, whereas “dear boy” appears by the end. Apart from this, keywords such as “love”, “play” and “her” can easily be related to the aboutness of the novel since they all indicate Miss Havisham’s eagerness for revenge by using the
beautiful Estella to lure and dominate Pip. Here, the dilemma between love and money is visible once again, since she makes Pip think he will not be fit for Estella’s love until he becomes rich.

Another way to obtain data from a corpus is to search for clusters; in other words, to search for sequences of words that are recurrent in the corpus. The clusters tool of AntConc not only shows sequences based on the search condition, but also scans the entire corpus for sequences of different length thanks to the N-grams tool. The findings confirm conclusions on prevailing topics of the novel drawn earlier from the keywords. Indeed, the analysis of the Magcorp corpus reveals that “dear boy” and “a gentleman” are the most repeated clusters, emphasising the strong bond between Magwitch and Pip and supporting the theory that the word gentleman is associated with the conflict of the plot. The text of Miss Havisham shows the prevalence of “love her”, along with several other clusters such as “look at her” and “call Estella”. Here, unlike in the case of Magwitch, Miss Havisham makes use of imperative forms which, in turn, possess a dramatic poetic quality because they appear in the text as repetitive rhythmic patterns such as the so-called “three-fold syntactic rhythms” (Golding 174) and “four-square structures of double time” (174). The sequence “Love her, love her, love her!” (Dickens 228), for instance, shows how Miss Havisham plays with Pip’s emotions, whereas “Let her call me mad, let her call me mad!” (284) and “So proud, so proud!” (285) refer to her adoptive daughter Estella, by whom Miss Havisham exerts control over Pip’s life.

2.2.3.3. Concluding remarks

The analysis based on corpus has proved that despite the many differences between Magwitch and Miss Havisham, both possess a close resemblance in their
behaviour towards Pip. The two characters are firmly related to Pip, as shown by the high frequency of the second singular pronoun “you” that refers to Pip. It is in order to compensate for their own personal lacks that they turn to him, thus acting similarly in this sense. Indeed, the use of the first person singular pronoun “I” hints at the isolation and lack of love they suffer from, revealing that they both yearn for love. However, they disregard their feelings and they encourage Pip to aim for social rank and wealth either by making him dream of being a gentleman, or urging him to become rich in order to be fit for Estella’s love, as shown by the dominant topics revealed by the keywords and clusters. Overall, these features underline and help Dickens construct the conflict of money and love that is central to the novel. In the end, Pip redeems his soul from the dreams of wealth and status and emerges triumphant.

3. CONCLUSION

In this paper I have introduced the concept of corpus stylistics in order to describe the framework within which this study has been developed. Dickensian characterisation techniques have been described and a corpus-based analysis of the characters of Magwitch and Miss Havisham has been presented. The purpose of the analysis was to discover the patterns that provide information about the links between the two characters and Pip, as well as their influence on him. In order to do this, three features have been described, namely wordlists, keywords and their concordance, and clusters, showing how the speech of the characters sheds light on their relationship with Pip.

Due to length restrictions, the paper has not developed further the analysis on characterisation. Research on the use of suspensions, or the thorough examination of the meanings constructed through the use of the most frequent lexical verbs in each character would offer a deeper insight on the construction of the characters analysed.
These and other aspects on the construction of characters remain for further studies on this topic from the perspective of corpus linguistics.
Works Cited


Software.


