

# Utilizing corpus stylistics to facilitate literary analysis: An assessment of the effectiveness of semantic domains in identifying major literary themes in a selection of Charles Dickens novels

Wasem Mohamed Abdulkhalek Ibrahim

Department of Basic Sciences, Applied College,  
Princess Nourah bint Abdulrahman University, Riyadh, SAUDI ARABIA

WMIbrahim@pnu.edu.sa

**Published:** 30 June 2022

**To cite this article (APA):** Ibrahim, W. M. A. (2022). Utilizing corpus stylistics to facilitate literary analysis: An assessment of the effectiveness of semantic domains in identifying major literary themes in a selection of Charles Dickens novels. *AJELP: Asian Journal of English Language and Pedagogy*, 10(1), 114-138. <https://doi.org/10.37134/ajelp.vol10.1.9.2022>

**To link to this article:** <https://doi.org/10.37134/ajelp.vol10.1.9.2022>

**Abstract:** This paper argues that corpus linguistic procedures can be considered an addition to the analytical inventory of traditional stylistics. It aims to explore how corpus linguistic procedures, particularly semantic domains, can be effective in detecting major literary themes in fiction. In order to do so, five corpora have been compiled: a corpus for each of the four novels of Charles Dickens' selected (i.e., *Oliver Twist*, *David Copperfield*, *Great Expectations* and *Our Mutual Friend*) and a compiled corpus combining all four novels. Wmatrix 5, with the BNC Sampler-Written as a reference corpus, is used to extract the key semantic domains in each corpus respectively. The literature on the selected novels is consulted to identify the major themes; it is then verified whether these themes are reflected in the corpus analysis, and, finally, the extent to which the procedure is effective in reflecting the major literary themes is explored. The findings confirm the effectiveness of the procedure of analysing semantic domains in studying literary texts, particularly in relation to their themes.

**Keywords:** Charles Dickens, corpus stylistics, semantic domains, WMatrix5

## INTRODUCTION

Corpus linguistics “approaches the study of language in use through corpora [...]. A corpus is a large, principled collection of naturally occurring examples of language stored electronically” (Bennett, 2010, p. 2). It can be defined as “dealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions” (McEnery & Hardie, 2012, p. 1). Its goal is to describe “the patterns of language use in the target textual domain” (Biber, 2011, p. 15). It “serves to answer two fundamental research questions: 1. What particular patterns are associated with lexical or grammatical features? 2. How do these patterns differ within varieties and registers?” (Bennett, 2010, p. 2). According to Leech (1992), “corpus linguistics ... defines not just a newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject” (Leech, 1992, p. 106). However, there is a debate among scholars whether corpus linguistics is a methodology (see Meyer, 2002;

McEnery & Hardie, 2012; Taylor, 2008) or a theory in its own right (see Gilquin, 2010). McEnery and Gabrielatos (2006, p. 34) refer to this debate stating that while some scholars “assert that corpus linguistics is not a branch of linguistics, nor a linguistic theory, but a methodology”, others argue that it “is more than just a methodology”.

“In the effort to bring corpus linguistic tools to bear on literary linguistic analysis, many recent publications have begun to map out new possibilities” (Toolan, 2006, p. 181). The use of computer-assisted textual analysis in the field of stylistics has given rise to a new field termed “corpus stylistics” (Sinclair, 2007; Mahlberg, 2007b). McIntyre (2015) highlights the difference between corpus stylistics and corpus linguistics stating that “corpus stylistics is simply corpus linguistics with a different object of study (literature as opposed to non-literary language)”, and that it makes itself unique by using qualitative tools and techniques of stylistics to analyze texts with the help of computational methods (McIntyre, 2015, p. 60). Mahlberg (2014, p. 378-380) also draws attention to the fact that corpus stylistics links principles from corpus linguistics and literary stylistics. A corpus-based work is highly quantitative in nature, but when combined with qualitative analysis, both help us to reach a higher degree of understanding of literary works or any other selected data (Semino & Short, 2004, p. 7).

Mahlberg and Smith (2010, p. 450) describe corpus stylistics as “an emerging field that aims to combine questions from literary stylistics with approaches from corpus linguistics”. It constitutes an “interface of corpus linguistics and literary stylistics”, which “employs methods and approaches of corpus linguistics and links them with concerns in literary stylistics and literary criticism” (Mahlberg, 2010, p. 295). Corpus Stylistics allows “a combination of quantitative and qualitative analysis of literary texts” (Mahlberg, 2013, p. 355). It is generally defined as employing “corpus linguistic methods to support the analysis of textual meanings and the interpretation of texts” (Mahlberg, 2013, p. 346). It allows “for the quantification of recurring linguistic features to substantiate qualitative insights and vice versa” (O’Keeffe, 2006, p. 50).

“[T]he value of applying corpus methods” in studying literature, which is the core of corpus stylistics, “is defined through the links that can be made between quantitative findings and qualitative analysis” (Mahlberg, 2010, p. 295). O’Halloran (2007a, p. 241) adds that “[t]he value of the corpus stylistic approach ... is that it can help reveal such subtleties that even some close readings may not detect”. “Corpus stylistics provides new insights into narrative texture and demonstrates the importance of recurrent linguistic features in shaping meaning” (Hunt & Carter, 2012, p. 27). It aims to “typically explore interpretations of literary texts identified in conventional readings but which may be hard to substantiate by traditional methods” (Hunt & Carter, 2012, p. 30). According to McIntyre (2010, p. 180), corpus stylistics is an approach that “employs the service or the evidence elicited from corpus linguistics to support literary analysis, transforming it from a basic, subjective claim-and-quote strategy ... into evidence-based objectivity that encompasses linguistic features”. Thus, it “enables us to address what has long been an issue with the analysis of prose fiction ... the problem of length and the fact that most prose texts are simply too long for the stylistician to deal with” (Mahlberg & McIntyre, 2011, p. 205). “[T]he application of the methods of corpus linguistics to text analysis has made it possible to place the selection and description of stylistically significant features on a firmer empirical footing” (Semino, 2011, p. 543). When “both quantitative and qualitative techniques of analysis are employed”, [t]he consequent strengths of such an analysis lie both in its scope and reliability” (Murphy, 2007, p. 67). “[E]mpirical corpus evidence can usefully provide substantiation of such initial evaluations of literary works” (O’Halloran, 2007b, p. 33), or in other words, it “can provide textual substantiation to impressionistic interpretation” (Widdowson 2008, p. 294). Indeed, corpus stylistics has enabled analysts “to test empirically claims about the language of

literature, to search for and provide evidence from texts, to establish the norms of literary and non-literary style, and to have in-depth insights into the texts' structures and meanings'' (Keshabyan-Ivanova, 2014, p. 59).

Since its inception, scholars have been interested in using corpus stylistic techniques. Examples of corpus stylistic studies include Adolphs (2006), Adolphs and Carter (2002), Biber (2011), Culpeper (2002, 2009), Fischer-Starcke (2010), Hoover (1999, 2002), Hori (2004), Lawson (2000), Louw (1993), Mahlberg (2007a, 2007b), McEnery et al. (2006), McIntyre (2008), McIntyre and Walker (2010), O'Halloran (2007a, 2007b), Romaine (2010), Scott (2006), Scott and Tribble (2006), Semino and Short (2004), Sinclair (2004), Starcke (2006), Stubbs (2001, 2005), and Toolan (2006).

Semino and Short (2004) investigated speech, thought and writing presentation in a corpus of late 20<sup>th</sup> century fictional and non-fictional works. Stubbs (2005) looked at clusters in Conrad's *Heart of Darkness*. Starcke (2006) concentrated on clusters in Austen's *Persuasion*. Mahlberg (2007) studied clusters in a corpus of Dickens' works. Culpeper (2009) investigated characterization in Shakespeare's *Romeo and Juliet* addresses the usefulness of Wmatrix's ability to compare texts at the grammatical and semantic level. McIntyre's (2010) analysis of *Reservoir Dogs* utilized the semantic functionality of Wmatrix to study characterization by examining keywords and key semantic categories. Archer and Bousfield (2010) also used Wmatrix to explore the main characters in Shakespeare's *King Lear*. Leech (2008, p. 162-78) used Wmatrix to analyse Virginia Woolf's *The Mark on The Wall*. Ho (2011) used a number of corpus tools (including Wmatrix) in her analysis of Fowles' *The Magus*. Walker (2010) examined keywords and semantic domains in Julian Barnes's novel *Talking it Over*. Mahlberg and McIntyre (2011) demonstrated the interplay of qualitative and quantitative methods in their corpus stylistic analysis of Ian Fleming's novel *Casino Royale*.

This paper supports the argument that integrating corpus procedures in the toolkit of traditional stylistics or manual literary analysis would enhance analysis and enable analysts to deal with large amounts of data. In addition, it assesses the corpus procedure of semantic domains and reveals its effectiveness in identifying major literary themes. In order to do so, it investigates to what extent semantic fields can identify the literary themes in four of Charles Dickens' novels, namely: *Oliver Twist*, *David Copperfield*, *Great Expectations* and *Our Mutual Friend*.

The "limits and pitfalls" of corpus stylistics are acknowledged by its practitioners (Mahlberg 2007b, p. 222). The most common problem is difficulty in obtaining either digital copies of literary texts or the right to use the available ones. Fortunately, the texts of Charles Dickens' works are all in the public domain, having been made available in digital form via Project Gutenberg<sup>1</sup>. For the purposes of this paper, the four selected works of Charles Dickens have been downloaded from the Project Gutenberg website and compiled into five separate text corpora (one for each novel and one combined corpus).

### **Wmatrix: What it offers and how it works**

Corpus stylistics is not only based on the availability of large collections of machine-readable texts, but also on computer software programs, such as, for example, Wmatrix (Rayson, 2003, 2008, 2021), which is used primarily in this paper. Wmatrix<sup>2</sup> is a software tool for

---

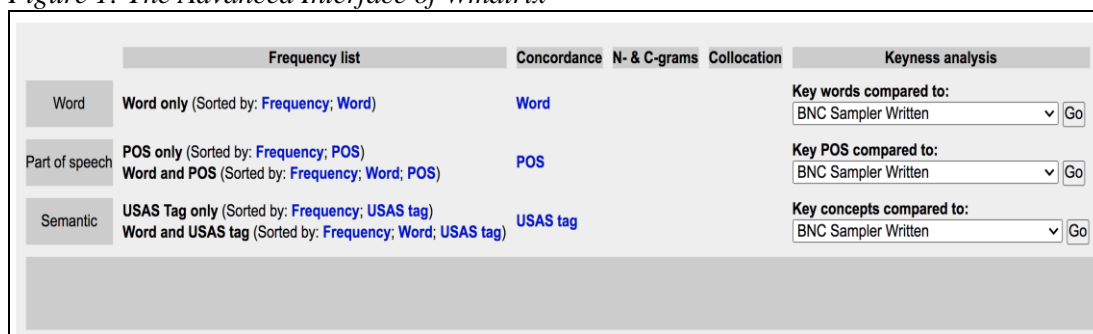
<sup>1</sup> Project Gutenberg is an online book catalogue that offers over 45,000 ebooks. They are all accessible and free to download. For further details see <http://www.gutenberg.org/>.

<sup>2</sup> See <http://ucrel.lancs.ac.uk/wmatrix.html> for further details

corpus analysis and comparison which provides a web interface to the CLAWS<sup>3</sup> (Constituent Likelihood Automatic Word-tagging System) and USAS<sup>4</sup> (UCREL<sup>5</sup> Semantic Analysis System) corpus annotation tools. It allows researchers to investigate the frequency and distribution of word forms and lemmas, keywords, grammatical word classes and semantic domains. Researchers can upload their own corpora to Wmatrix and click on the tag wizard which will automatically tag/annotate the uploaded files/corpora grammatically by CLAWS and semantically by USAS. Once the tagging process is complete, each word in the corpus will be separated, numbered and assigned different kinds of tag codes (i.e., Frequency, Semantic tags, and POS tags).

Wmatrix has two interfaces: simple and advanced. The simple interface consists of four functions: extracting wordlists, identifying frequencies, searching for a particular word, and extracting keyword clouds, which display the significance of each keyword by allocating it an appropriate font size. The advanced interface offers more tags, particularly semantic and part of speech tags.

*Figure 1: The Advanced Interface of Wmatrix*



The semantic tagger in Wmatrix is used in this study. In the semantic tagging/annotation process, USAS assigns a semantic tag to each word in the corpus. According to Rayson (2003, p. 66), the semantic tags are composed of:

1. An upper-case letter indicating general discourse field
2. A digit indicating a first subdivision of the field
3. A decimal point followed by a further digit to indicate a finer subdivision (optional)
4. One or more 'pluses' or 'minuses' to indicate a positive or negative position on a semantic scale (optional)

USAS includes 21 major semantic domains, as shown in Table 1 below.

*Table 1: Semantic fields of USAS – Adapted from Rayson (2003)*

1.	A	GENERAL & ABSTRACT TERMS
2.	B	THE BODY & THE INDIVIDUAL
3.	C	ARTS & CRAFTS
4.	E	EMOTION
5.	F	FOOD & FARMING
6.	G	GOVERNMENT & PUBLIC
7.	H	ARCHITECTURE, BUILDINGS, HOUSING & THE HOME
8.	I	MONEY & COMMERCE

<sup>3</sup> See <http://www.comp.lancs.ac.uk/ucrel/claws/> for further information.

<sup>4</sup> See <http://www.comp.lancs.ac.uk/ucrel/usas/> for further information.

<sup>5</sup> University Centre for Computer Corpus Research on Language.

9.	K	ENTERTAINMENT, SPORTS & GAMES
10.	L	LIFE & LIVING THINGS
11.	M	MOVEMENT, LOCATION, TRAVEL & TRANSPORT
12.	N	NUMBERS & MEASUREMENT
13.	O	SUBSTANCES, MATERIALS, OBJECTS & EQUIPMENT
14.	P	EDUCATION
15.	Q	LANGUAGE & COMMUNICATION
16.	S	SOCIAL ACTIONS, STATES & PROCESSES
17.	T	TIME
18.	W	THE WORLD & ENVIRONMENT
19.	X	PSYCHOLOGICAL ACTIONS, STATES & PROCESSES
20.	Y	SCIENCE & TECHNOLOGY
21.	Z	NAMES & GRAMMATICAL WORDS

These major domains are refined and expanded into 232 semantic categories<sup>6</sup>. Figure 2 below provides an example showing how the major domains of money & commerce (I), entertainment, sports, & games (K), and life & living things (L) are expanded into subcategories.

*Figure 2.: An illustration of the expansion of three semantic domains*

<b>I MONEY &amp; COMMERCE</b>	
I1	Money generally
I1.1	Money: Affluence
I1.2	Money: Debts
I1.3	Money: Price
I2	Business
I2.1	Business: Generally
I2.2	Business: Selling
I3	Work and employment
I3.1	Work and employment: Generally
I3.2	Work and employment: Professionalism
I4	Industry
<b>K ENTERTAINMENT, SPORTS &amp; GAMES</b>	
K1	Entertainment generally
K2	Music and related activities
K3	Recorded sound etc.
K4	Drama, the theatre & show business
K5	Sports and games generally
K5.1	Sports
K5.2	Games
K6	Children's games and toys
<b>L LIFE &amp; LIVING THINGS</b>	
L1	Life and living things
L2	Living creatures generally
L3	Plants

The investigation of semantic domains is particularly useful in corpus stylistics. Many studies have employed the USAS component of Wmatrix (see, for example, McIntyre & Walker, 2010; Murphy, 2007; and Walker, 2010, 2012). It should be noted, however, that, to

<sup>6</sup> See <https://github.com/UCREL/Multilingual-USAS/blob/master/English/USASSemanticTagset.pdf> for the complete Semantic Tagset.

my knowledge, no other study has used Wmatrix to study the effectiveness of the semantic domains procedure in identifying the major literary themes in Charles Dickens' works.

## METHODOLOGY

This paper aims to investigate the following research issues: (1) whether the semantic domains (provided by WMatrix5) relate to the major themes of literary text(s); and (2) the similarities and/or differences between the results of corpus analysis in relation to semantic domains and the traditional critical reviews of the selected novels. To achieve this aim, four novels of Charles Dickens were downloaded from the website of *Project Gutenberg* (See Table 2 below).

*Table 2: A list of the selected novels and their word counts*

No.	Novel	Word Count
1	<i>David Copperfield</i>	358644
2	<i>Great Expectations</i>	186487
3	<i>Oliver Twist</i>	158656
4	<i>Our Mutual Friend</i>	328329
<b>Total</b>		<b>1032116</b>

Each novel was saved as a plain text file. Then, the four novels were compiled into one large corpus also saved as a plain text file. Using the BNC Sampler-Written (a subset of the BNC - British National Corpus - consisting of about one million words) as a reference corpus, the key semantic domains of the selected novels and the compiled corpus were identified by the WMatrix5 semantic tagger. Finally, the categories of the semantic fields were examined and compared to the identified literary themes suggested in the literature.

## RESULTS

Literary critics argue that Charles Dickens' novels offer deep insights into the spirit of the Victorian age through addressing themes such as social class, poverty, child labor, morality, education, self-actualization, and so on<sup>7</sup>. Dickens was

“one of the first popular writers who brought pictures of what is called common life into fashion. It is he who has been mainly instrumental in leading the present generation of authors to disregard to a great extent the pictorial advantages of life on the upper levels of society, and to find a counter picturesqueness in the experiences of the poor” (Oliphant, 1862, in Wall, 1970, p. 158).

Charles Dickens was born in 1812 into a bourgeois family as his father, John Dickens, worked in the Naval Pay Office and earned a good salary (Johnson, 1977, p. 13). However, due to his merrymaking, Dickens' father went bankrupt and was sent to the debtors' prison, and so the family became poor. As the eldest son, Dickens, who was 11 years old, had to work as a child laborer in a blacking warehouse. This painful experience led Dickens to deeply sympathize with poor and oppressed people, especially children. After his father's

---

<sup>7</sup> Analysis of Charles Dickens's Novels – Literary Theory and Criticism (literariness.org)  
<https://literariness.org/2019/04/02/analysis-of-charles-dickenss-novels/>

release from prison, Dickens was sent to study at the Wellington House Academy for two years. Then, he had to drop school and work again as his father faced financial problems for the second time. He successively managed to work in many jobs: a junior clerk at a law office, a freelance reporter in a newspaper office and a stenographer in the court. Dickens' early work experience and suffering helped him to understand the hardships of the people, especially the heavy burden on the poor, and inspired many of his novels (Johnson, 1977).

“[N]o other writer of the English language except Shakespeare has left so many types of characters as Dickens has done” (Trollope, 1870, in Wall, 1970, p. 180).

“Dickens' world was richly populated with a variety of characters drawn from different social structures and often his moral vision is engaged through the mechanism of caricature. [...] Dickens' novels also show, at the same time, a remarkable awareness of contemporary social issues” (Choudhury, 2009, p. 5).

Several critics have identified Dickens as a writer for “the people” because his characters are ordinary people: lawyers, clerks, merchants, hotel owners, craftsmen, criminals, servants and workers (Ackroyd, 1990, p. xiv). Indeed, Dickens' characters embody people from all occupations and social classes in 19<sup>th</sup> century Britain.

“His range is very varied. He has attempted to describe every kind of scene in English life, from quite the lowest to almost the highest...Mr. Dickens's novels aim to delineate nearly all that part of our national life which can be delineated” (Bagehot, 1858, in Wall, 1970, p. 124).

One of the major themes in Dickens' novels is child labor and suffering. “No one, at any rate no English writer, has written better about childhood than Dickens. [...] no novelist has shown the same power of entering into the child's point of view” (Orwell, 1968, p. 423-424). In Dickens' works, “the figure of the child and the topic of childhood have always loomed large” (Merchant & Waters, 2015, p. 1). He portrayed “how children were perceived by the Victorians and how the phenomenon of increasing child labor did fit into this particular perception” (Schuster, 2014, p. 7). Dickens' amazing talent for portraying the world of children is evident in his vivid child characters in novels like *Oliver Twist*, *David Copperfield*, and *Great Expectations*. Dickens' miserable childhood had an impact on his portrayal of the darkness of the social reality faced by poor children. Ackroyd (1990) explained this idea as follows:

“There is something in Dickens's infancy, something which cannot now be recovered or understood, some primal fear which left him casting about for images with which to express it and which gave him as a novelist that sensitivity to the adult world which is most often to be found in the eyes of a frightened child” (Ackroyd, 1990, p. 54).

Dickens also criticized the social system. As George Orwell (1968) puts it:

“Fasten upon this or that minor abuse, expose it, drag it into the open, bring it before a British jury, and all will be well – that is how he sees it...There is no clear sign that he wants the existing order to be overthrown, or that he believes it would make very much difference if it were overthrown” (Orwell, 1968, p. 416).

It seems that Dickens' criticism of the social system is almost exclusively moral. His main target of criticism is not the social system but human nature which he was much concerned to improve. Therefore, Dickens stresses the need for reformation, through focusing on morality (Orwell, 1968, p. 427).

“It seems that in every attack Dickens makes upon society he is always pointing to a change of spirit rather than a change of structure .... His approach is always along the moral plane... Useless to change institutions without a ‘change of heart’ – that, essentially, is what he is always saying” (Orwell, 1968, p. 427).

Morality is emphasized in Dickens' works. Dickens' focus on morality might be attributed to his Christian puritanism. He advocated what he thought to be typically Christian moral values: love, mercy, benevolence and so on. According to Walder (1981, p. 208), Dickens' “knowledge of the Bible ... as well as the Book of Common Prayer, is testified to by frequent, accurate and often surprisingly relevant allusion throughout his works.” Dickens attempted to convey charity as a kind of Christian virtue. Charity “means more than the simple human virtue of benevolence, or giving alms to the poor; it implies the more general motive of Christian love, expressed as a love of God and one's neighbor” (Walder, 1981, p. 45). As George Orwell (1968, p. 417) explains, Dickens' “whole ‘message’ is one that at first glance looks like an enormous platitude: If men would behave decently the world would be decent.” Indeed, Dickensian humanitarianism is reflected in his calling for mercy, universal love and forgiveness, and opposing all kinds of evil, violence and oppression in his novels (Orwell, 1968). “[T]he novels of Dickens can all be reduced to one phrase, to wit: Be good, and love” (Taine, 1856, in Wall, 1970, p. 103). Dickens strove to create a sense of “‘human solidarity’” through shared experience and emotions to evoke in his readers a sense of collective responsibility for social problems (Bell, 2000, p. 126-127).

The German philosopher Karl Marx valued the critical realism of Victorian novelists, particularly Dickens.

The ... graphic and eloquent pages have issued to the world more political and social truths than have been uttered by all the professional politicians, publicists, and moralists put together... (Raina, 1986, p. 137)

This is true of Charles Dickens, whose works offer deep and enlightening insights into the spirit of the 19<sup>th</sup> century.

### **A Corpus-assisted Analysis of Dickens' Selected Works**

In this section, the corpus-assisted analysis is conducted as follows: (1) the semantic domains are identified in each selected work using WMatrix 5; and (2) the extracted semantic domains are compared to the literary themes suggested in critical reviews of Dickens' works (Churchill, 1975; Wall, 1970).

#### ***Oliver Twist***

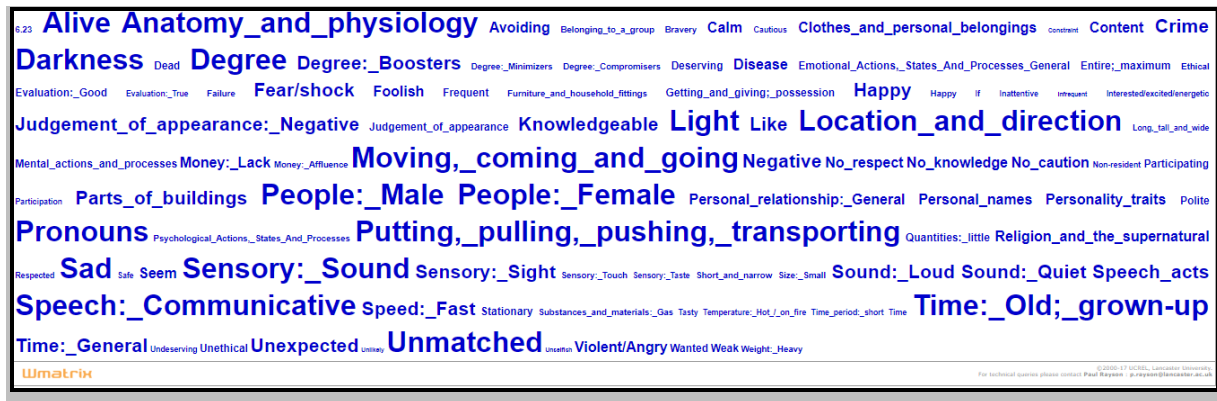
*Oliver Twist* is set in a workhouse and discloses the cruel, greedy and egotistic nature of Victorian industrial society. It tells the story of the miserable childhood of an abandoned orphan, Oliver Twist. As a child, he was mistreated, beaten, and forced to join the diabolic underworld of criminal life. He ran away and always preserved his positive outlook despite his suffering. Morality is the overarching theme of this novel since it depicts a conflict between good and evil, exposes and punishes vice, and advocates public morality. It revealed the poor conditions of workhouses and orphanages, and reflected how they were devoid of



compassion. It also reflected on the financial discrepancies between the social classes, portraying the lives of both the lower classes who inhabit the London underworld, and of the rich aristocrats. It also tackles the themes of thought and belief in criticizing the Church of England for not providing enough care for the poor who sought their help.

The semantic categories/fields extracted from the text of Oliver Twist are shown in Figure 3 and listed in Table 3.

Figure 3: Semantic clouds in Oliver Twist



As shown in Figure 3, the bigger the font size of a semantic field, the more significant it is in the text. This makes it easier for anyone to get a general idea about the major components of the story: the setting, inhabitants of the story world, events and so on.

Table 3: Wmatrix Semantic Tagging of Charles Dickens' Oliver Twist

Semantic Field	LL	Semantic Field	LL	Semantic Field	LL	Semantic Field	LL
Pronouns	+3259.46	<b>Moving, Coming &amp; Going</b>	+362.84	Negatives	+233.69	<b>Judgement of Appearance : Negative</b>	+145.92
Anatomy & Physiology	+1472.67	<b>Putting, Pulling, Pushing, Transporting</b>	+360.04	Sensory: Sight	+222.97	<b>Crime</b>	+145.81
Unmatched	+1070.20	Darkness	+356.20	Like	+197.40	Knowledgeable	+130.58
<b>People: Male</b>	+979.37	Sensory: Sound	+335.70	Speech Acts	+196.08	<b>Fear/Shock</b>	+129.10
Speech: Communicative	+747.32	Alive	+265.37	Sound: Loud	+179.04	<b>Disease</b>	+128.71
Light	+605.55	<b>Time: Old; Grown-up</b>	+257.77	Time: General	+166.26	Content	+127.37
Degree	+494.73	Speed: Fast	+256.49	<b>Happy</b>	+164.56	Seem	+115.55

Location & Direction	+481.90	Degree: Boosters	+252.3 6	Unexpected	+153.44	<b>Money: Lack</b>	+108.79
<b>Sad</b>	+466.61	Parts of Buildings	+245.3 5	Sound: Quiet	+152.28	<b>Violent/Angry</b>	+104.75

Table 3 lists the semantic categories according to their statistical significance. The semantic categories which are relevant to the major literary themes in the story are highlighted in the table. Table 4 provides a list of the significant words used in the text to reflect each category.

*Table 4: Oliver Twist: Significant words in each semantic category*

Oliver Twist, Charles Dickens	
Semantic tag	Word
People: Male	man – gentleman – boy – men - fellow - gentlemen - lad - Mr – male - fellows - bachelor - chaps - mister - lads - chap
Sad	cry - desperate - misery - grief – suffered - miserable - sad - sorrow - weeping - suffering - melancholy - pity - rueful - lamentations - pitied – sobbed – wept - unhappy – grim - desolation - repent - sob - cheerless – tragic – sullenly – lamented
Moving, Coming & Going	come - returned - left - go - passed – walked - went – followed - fell - entered - reached – nodded – got - ran – return - arrived – rising - crossed - darted - sunk - journey – emerged - staggered
Putting, Pulling, Pushing, Transporting	turned - put - lay - brought – raised – held - shook - carried – turning - hung - placed - set - threw - raising - thrust - putting – moved – send - hold - bring – dropped - pushed - delivered - thrusting - throwing - dragged - pulling - carriage
Alive	life - alive - lives - live - bloodless
Time: Old; Grown-up	old - venerable - growing up - grown up - adult- middle aged - grow up - mature
Happy	happy - laugh - smile - merry - joy - grin - delight - humour - relief - cheerful - merriment - jolly - chuckled - jokes - sanguine – mirth – funny – amused - fun - lol – playful - gaiety - grinned
Judgement of Appearance: Negative	dirty - duff – wretched – dismal - hideous - ugly - ghastly - awful - repulsive - horrible - frightful - filthy - ragged - horrid - soiled - unpleasant - vulgar - squalid - filth – unwashed
Crime	thief - robber– stole - villain - guilty - crime - ruffian - offence - burglary - fraud - plundering - kidnapped – injustice - conspiracy - offender
Knowledgeable	know – awakened - information - acquainted - remember - experience - recollection - news - conscious - wisdom - recognised – recollect - recall
Fear/Shock	fear - afraid - alarm - terror – frightened - horror - dread - startled - frighten - menacing - coward – startling – cowardly - cowered - fearing - scared
Disease	ill – fever – pain - mad – sick – cold – cripples – painful – fit - hurt - exhausted – wounded – weary - sickness - symptoms - crazy – wearied - bruised - sickly - fatigue - injured - cough - wounds – limped
Money: Lack	poor – paupers – poverty – destitute - in need - beggar - pauper - debased
Violent/Angry	violent - force - violence - rage - angry – fierce - wrath - agitation – threatening - hit - indignation - angrily - furious - brute - cruel - fury - threats - ferocity - anger

The semantic category **People: Male** (+979.37) relates to the inhabitants of the story world and shows that it was dominated by males. The significance of the category **Time: Old; Grown-up** (+257.77) can be accounted for by the fact that the story is told from the perception of a child who is controlled by grown-ups and that the story depicts his life

journey as he grows up. The semantic categories **Sad** (+466.61), **Judgement of Appearance: Negative** (+145.92), **Fear/Shock** (+129.10), **Disease** (+128.71), and **Money: Lack** (+108.79) relate to the miserable life and work conditions of the protagonist, and hence, to the themes of child labor and poverty. The categories **Crime** (+145.81) and **Violent/Angry** (+104.75) relate to the mistreatment and cruelty the protagonist suffered from and also the life of crime he was forced to join. The category **Happy** (+164.56) is related to the protagonist's positive outlook despite his suffering. The categories **Moving, Coming & Going** (+362.84) and **Putting, Pulling, Pushing, Transporting** (+360.04) relate to the actions and, hence the main events, in the story. Hence, the semantic categories reflect plot elements and relate to the themes of morality, criminality, child labor and poverty.

**David Copperfield**

Dickens' *David Copperfield* is a bildungsroman which depicts the hardworking and the moral and spiritual development of its protagonist, David Copperfield. The novel highlights the themes of injustice in the educational system, marriage, morality, financial discrepancies between social classes, and the miserable conditions of the working class.

Figure 4 and Table 5 provide the extracted semantic categories in the text. Table 6 provides a list of the words reflecting each semantic category in the text.

Figure 4: Semantic clouds in David Copperfield



Table 5: Wmatrix Semantic Tagging of Charles Dickens' David Copperfield

Semantic Field	LL	Semantic Field	LL	Semantic Field	LL	Semantic Field	LL
Pronouns	+19642.29	<b>Thought, Belief</b>	+652.76	<b>Polite</b>	+211.58	Discourse Bin	+137.93
Unmatched	+1915.23	<b>Kin</b>	+606.87	Degree: Compromisers	+202.30	Parts of Buildings	+135.16
Anatomy and Physiology	+1624.12	Negative	+584.23	Size: Small	+192.63	Sensory: Touch	+134.36
Speech: Communicative	+1527.33	<b>Moving, Coming &amp; Going</b>	+506.78	<b>Respected</b>	+187.21	Entire; Maximum	+124.64
Degree: Boosters	+981.67	Alive	+487.99	Personality traits	+173.50	<b>Putting, Pulling, Pushing, Transporting</b>	+122.71
Like	+913.72	Content	+481.77	<b>Calm</b>	+173.24	Frequent	+114.48

<b>Knowledge able</b>	+842.24	Existing	+353.16	Frequent	+163.73	Sound: Quiet	+114.04
Degree	+830.04	<b>Happy</b>	+338.16	Time	+159.77	Seem	+107.23
Time: General	+774.47	Stationary	+257.79	<b>Relationship: Intimacy and Sex</b>	+154.26	No Knowledge	+104.29
Sensory: Sight	+758.59	Unexpected	+244.68	Degree: Minimizers	+144.25	Expected	+102.72
Light	+730.97	Darkness	+232.95	<b>Time: Old; Grown-up</b>	+138.31	<b>Psychological Actions, States and Processes</b>	+101.11

*Table 6: David Copperfield: Significant words in each semantic category*

David Copperfield, Charles Dickens	
Semantic tag	Word
Knowledgeable	know- remember- recollect- knowledge- remembrance- informed- aware- experience- conscious- recall- information- recollection- looking back- famous- news- memorable- recalled- recognized- wisdom- recognition- recollections- inform- recalling- informing- experiences- spies- identify- informs- recognizing- expert
Thought, Belief	think- felt- believe- opinion- considered- wonder- suppose- supposed- impression- regarded- belief- thinks- suspicion- conviction- suspect- musing- conceived- attitude- views- opinions
Kin	aunt- mother- family- father- wife- sister- married- brother- son- papa- daughter- husband- engaged- mama- uncle- marriage- niece- cousin- marry- nephew- widow- Mrs- the old man- marrying- families- parents- wedding- parent- godmother- brothers- mothers- grandfather – father-in-law - widower- twin- husbands- offspring- bride- son in law- great aunt- brother in law- great grandfather- sister in law- fathers
Moving, Coming & Going	returned- went- go- come- left- walked- passed- trot- return- entered- journey- rising- followed- nodded- come back- went away- went in- departure- steps- advanced- voyage
Happy	happy- smile- laughing- delighted- cheerful- joy- delight- relief- smiled- cheerfully- merry- laughter- humour- amused- cheer up- happily- gaily- jest- cheerfulness- joyfully- joke- grin
Polite	thanked- grateful- compliments- gratitude- gracious- compliment- polite- gentlemanly- indebted- propriety- decently- complimentary- politely- civilities- courtly- pleasantry
Respected	respect- admiration- dignity- praise- honoured- honour- admire- homage- respecting- esteem- glorified- reverence- deference- admirable
Calm	rest- softly- gentle- calm- peace- resting- patience- rested- placid- cordiality- soothing- gentility- soothed- rests- serenity- gentler- reposed- soothe- serene
Relationship: Intimacy & Sex	love- kiss- in love- embrace- romantic- intimacy- hug- fell in love- intercourse- hugged- hugging- sweetheart- fall in love- sweethearts- suitor- sleeps with- love making- live together
Time: Old; Grown-up	old- grown up- grew up- matured- grow up- grown- growing up- dated- middle aged- old age- superannuated- mature- established- patrician- mature

Putting, Pulling, Pushing, Transporting	put- turned- brought- lay- shook- held- bear- set- raised- turning- sent- carried- moved- thrown- dropped- put down- send- shed- seated- lifted up- squeeze- fetch- hanging- pushed
Psychological Actions, States and Processes	mind- state of mind- instinctively- instinct- trance

The semantic categories **Knowledgeable** (+842.24) and **Thought, Belief** (+652.76) relate to the themes of the protagonist’s self-development and education. **Kin** (+606.87) and **Relationship: Intimacy & Sex** (+154.26) reflect the themes of family and marriage. The protagonist’s psychological state, positive outlook and maturity are reflected in the categories **Happy** (+338.16), **Calm** (+173.24), and **Psychological Actions, States & Processes** (+101.11). The themes of morality and social class are conveyed by the categories **Polite** (+211.58) and **Respected** (+187.21). The protagonist’s progress into maturity is reflected by the semantic category **Time: Old; Grown-up** (+138.31). This category may also refer to older inhabitants of the story world. The categories **Moving, Coming & Going** (+506.78) and **Putting, Pulling, Pushing, Transporting** (+122.71) reflect the actions which propel the plot. As such, the semantic categories show that *David Copperfield* discusses the themes of education, marriage/family, morality, and self-development to maturity.

**Great Expectations**

Dickens’ *Great Expectations* portrays the journey of its protagonist, Philip Pirrip (Pip) from childhood into youth. The major themes conveyed by the novel include morality, self-development, the social class system, and crime. Figure 5 and Table 7 provide the extracted semantic categories in the text. Table 8 provides a list of the words reflecting each semantic category in the text.

Figure 5. Semantic clouds in Great Expectations

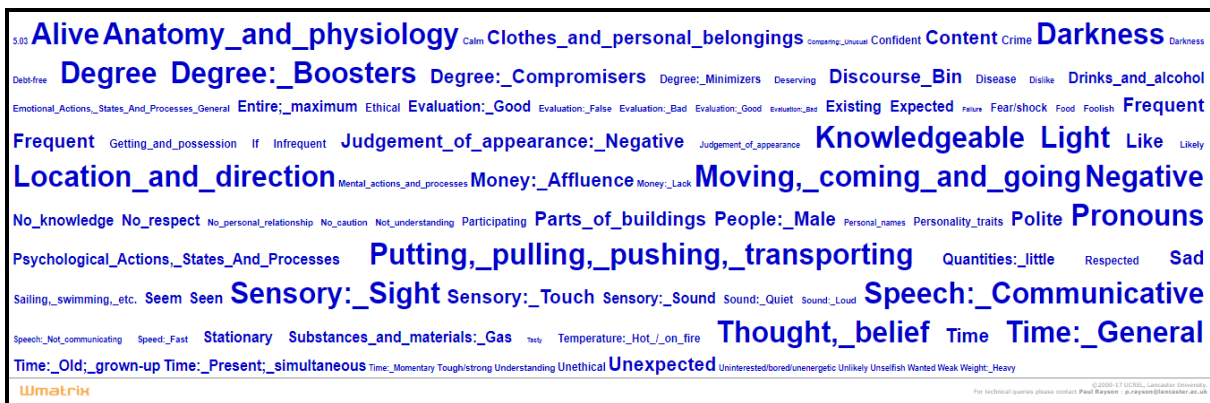


Table 7: WMatrix Semantic Tagging of Charles Dickens’ Great Expectations

Semantic Field	LL	Semantic Field	LL	Semantic Field	LL	Semantic Field	LL
Pronouns	+1295 1.68	Time: General	+423.2 1	<b>People: Male</b>	+209.9 3	<b>Polite</b>	+134.03
Anatomy & Physiology	+1122. 50	Location & Direction	+384.7 7	Frequent	+189.0 8	Time	+127.51

Light	+794.5 9	<b>Moving, Coming &amp; Going</b>	+379.7 0	Parts of the Buildings	+182.1 5	<b>Money: Affluence</b>	+126.51
Speech: Communicative	+671.7 8	<b>Putting, Pulling, Pushing &amp; Transporting</b>	+324.5 2	Discourse Bin	+181.3 1	<b>Clothes &amp; Personal Belongings</b>	+123.59
Sensory: Sight	+647.6 4	Degree: Boosters	+315.2 3	Like	+181.1 0	Content	+112.19
Degree	+488.6 9	Darkness	+283.5 2	<b>Sad</b>	+161.9 9	Sensory: Touch	+92.83
Negative	+446.0 0	<b>Thought, Belief</b>	+233.8 2	Degree: Compromises	+148.1 7		
<b>Knowledgeable</b>	+441.8 2	Alive	+226.6 2	Unexpected	+146.0 9		

Table 8: *Great Expectations*: Significant words in each semantic category

Great Expectations, Charles Dickens	
Semantic tag	Word
<b>Knowledgeable</b>	know - remember - knowledge - remembrance - information - aware - informed - recognized - acquainted - experience - recalled - remembered - news - memorable - recognition - recall - recollect - identify - wisdom
<b>Moving, Coming &amp; Going</b>	go - come - returned - left - passed - walk - leave - fell - coming - entered - rising - crossed - nodding - tumbling - movement - steps - arrival - journey - staggered - rolled - approached - scattered - ascended - creep - drifting
<b>Putting, Pulling, Pushing, Transporting</b>	put - held - lay - turned - bring - shook - threw - seated - placed - pulled - hanging - pushed - removed - dropping - carriage - shipping - dragged
<b>Thought, Belief</b>	thought - felt - believe - considered - opinion - supposed - suspicion - impression - meaning - conscience - judgement - attitude
<b>People: Male</b>	man - boy - gentleman - fellow - chap - Mr - bachelor - lad - male - manhood - manly
<b>Sad</b>	cried - miserable - suffered - melancholy - desperate - sulky - despair - mourning - regret - dejected - unhappy - remorse - grim - suffering - wept - sorrowful - sad - miserably - gloomily
<b>Polite</b>	thanked - grateful - compliments - polite - gracious - gratitude - apologetically - compliment - thanks - peasantry - indebted
<b>Money: Affluence</b>	fortune - fortunes - treasure
<b>Clothes &amp; Personal Belongings</b>	pocket - dress - clothes - boots - hat - wore - coat - put on - sleeve - suit - gloves - shoes - bag - purse - cloak - apron - waistcoat - trousers - veil - jewels - shroud - crown - skirts - brooch - knitting - tailor

The semantic categories **Knowledgeable** (+441.82) and **Thought, Belief** (+233.82) reflect the protagonist's self-development. The high statistical significance of the categories **Moving, Coming & Going** (+379.70) and **Putting, Pulling, Pushing & Transporting** (+324.52) convey an idea about the nature of the plot as full of events and actions. The category **People: Male** (+209.93) informs us that most of the inhabitants of the story world

are males. The category **Sad** (+161.99) reflects the miserable conditions of the inhabitants of the story world and may relate to the theme of social class. The category **Money: Affluence** (+126.51) reflects the theme of social class and may potentially relate also to the theme of criminality. The category **Polite** (+134.03) relates to the theme of morality. Finally, the category **Clothes and Personal Belongings** (+123.59) may relate to the theme of criminality particularly stealing, which is evident through the use of words such as ‘pocket’, ‘purse’, ‘jewels’, and ‘brooch’. It may also relate to the female inhabitants of the story world through references to items of clothes and actions stereotypically linked with women such as ‘dress’, ‘apron’, ‘veil’, ‘skirts’, ‘brooch’, and ‘knitting’.

### Our Mutual Friend

*Our Mutual Friend* criticizes moral and social problems, particularly, the wide gaps between the social classes, crime, and corruption. It also tackles the theme of family and marriage in Victorian society. Figure 6 and Table 9 provide the extracted semantic categories in the text. Table 10 provides a list of the words reflecting each semantic category in the text.

Figure 6: Semantic clouds in *Our Mutual Friend*



Table 9: WMatrix Semantic Tagging of Charles Dickens’ *Our Mutual Friend*

Semantic Field	LL	Semantic Field	LL	Semantic Field	LL	Semantic Field	LL
Pronouns	+10782.94	Darkness	+380.19	<b>Money: Lack</b>	+205.50	<b>Respected</b>	+144.58
Unmatched	+3931.05	Sensory: Sight	+372.29	Content	+193.98	<b>No Respect</b>	+142.58
Anatomy & Physiology	+1864.49	Alive	+355.12	Speech Acts	+183.88	<b>Unethical</b>	+139.31
Speech: Communicative	+1059.46	Location & Direction	+343.22	Unexpected	+179.82	Sensory: Touch	+131.57
Negative	+1002.99	<b>People: Male</b>	+335.99	<b>Ethical</b>	+169.53	<b>Personal Relationship: General</b>	+120.95
Light	+824.68	<b>Putting, Pulling, Pushing &amp; Transporting</b>	+299.11	<b>Kin</b>	+167.67	Deserving	+118.46
Like	+785.20	<b>Moving,</b>	+285.6	Frequent	+164	Degree:	+110

		<b>Coming &amp; Going</b>	4		.84	Compromises	.86
Degree	+620.60	<b>Thought, Belief</b>	+278.10	<b>Judgement of Appearance : Positive</b>	+154.53	Sensory: Sound	+103.97
Time: General	+431.95	<b>Sad</b>	+238.01	Foolish	+150.13	Quantities: Little	+102.44
<b>Knowledgeable</b>	+414.19	<b>Money: Affluence</b>	+220.64	Size: Small	+148.74	<b>Confident</b>	+100.60
<b>Work &amp; Employment: Professionalism</b>	+395.13	<b>Polite</b>	+212.20	Seem	+147.25	Calm	+96.41

*Table 10: Our Mutual Friend: Significant words in each semantic category*

Our Mutual Friend, Charles Dickens	
Semantic tag	Wordlist
Knowledgeable	know – remember – knowledge – aware - information - acquainted-remembrance – news – recall – experience – recollect - wisdom - recognized – awakened - recollection
Work & Employment: Professionalism	secretary- reputation- secretaries- reputations- gatekeeper- colleagues
People: Male	man- gentleman- boy- Mr- men- fellow- gentlemen- chap- fellows- boys- male- manly- lad- bachelor- Mister- lads- chaps
Putting, Pulling, Pushing, Transporting	put- turned- brought- held- set- lay- shaking- carriage- turn- raised- removed- seated- borne- thrown- delivered- hung- pick up- took down- twisted- stuck- deposited- put into- tossed- turns- stretched out.
Moving, Coming & Going	returned- come- go- left- passed- walk- sat down- followed- comes- pass- return- steps- arrived- move- nod- come home- wandering- rolled- laid down- departed- floating- fallen- advancing- slipped
Thought, Belief	think- feel- believe- opinion- suppose- wonder- suspicion- regarded- attitude- meaning- trust- impression- distrust- musing- guess- find- meditating- reason- assumed- deemed- deliberation
Sad	cried- melancholy- miserable- glum- mourning- wept- sad- grim- desperate- weeping- unhappy- gloomily- suffering- despair- suffered- miseries- cries- regret- downcast- burst into tears- embarrassing- misery- sobbing
Money: Affluence	fortune- fortunes- riches- treasure- richest- fortune hunter
Polite	thankful- gracious- grateful- compliments- polite- thanked- compliment- thank- thanks- politely- politeness- gratitude- apologetic- complimentary- graciously- courteous- apologetically- thanking- complimented- indebted- thankfulness- refined- good manners- decently- well bred- courtesy
Money: Lack	poor- beggars- poverty- insolvent- pauper- beggar- destitution- in need- needy- hard up- broke- beggar- penniless
Ethical	honour- fair- honourable- noble- respectable- forgive- moral- gallant- virtue- justified- virtuous- virtues- forgiveness- justly- unimpeachable- honours- justification- excusable- chivalrous- integrity- honourably
Kin	father- sister- husband- wife- daughter- family- brother- mother- married- godmother- marry- son- marriage- parent- bride- mum- marrying- wedding- daughters- fathers- aunt- bridegroom- mamma
Judgement of	pleasant- pretty- lovely- comfortable- beautiful- charming- handsome-



Appearance: Positive	delicate- grand- neat- clean- nice- majestic- attracted- beauty- grace- radiant- impressive- adorable- charmed- fashionable- gorgeous
Respected	respect- dignity- admiration- honoured- deference- dignified- homage- submissively- estimable- flattered- praises- honorary
No Respect	disgrace- contempt- scorn- disdain- degraded- irreverent- disgraced- humiliation - derision- profane- disdained
Unethical	rogue- mercenary- wicked- betrayed- rascal- cheat- reproachful- disgraceful- fiddle- swindlers- sin- traitor- corruption- vices
Personal Relationship: General	friend- partner- meet- acquaintance- companion- comrade- escort- friendship- mate- playmate
Confident	confidence- trust- emphatic- peace of mind- confident- reassuring- trustfulness

The semantic categories **Knowledgeable** (+414.19) and **Thought, Belief** (+278.10) relate to the themes of self-development and education. The theme of the gap between the social classes is quite significant and is reflected in the categories **Work & Employment: Professionalism** (+395.13), **Sad** (+238.01), **Money: Affluence** (+220.64), **Money: Lack** (+205.50), and **Respected** (+144.58). The category **People: Male** (+335.99) reflects the fact that the majority of the inhabitants in the story world are males. The categories **Putting, Pulling, Pushing & Transporting** (+299.11) and **Moving, Coming & Going** (+285.64) reflect the amount of action in the story. The theme of morality is highlighted in the story through the categories **Polite** (+212.20), **Ethical** (+169.53), **No Respect** (+142.58), **Unethical** (+139.31), and **Confident** (+100.60). The theme of family is conveyed by the categories **Kin** (+167.67) and **Personal Relationship: General** (+120.95). Finally, the category **Judgement of Appearance: Positive** (+154.53) relates to the theme of family/relationships. It may, however, reflect the theme of social class, which is evident in the use of these words: ‘grand’, ‘majestic’, ‘grace’, and ‘fashionable’.

### The Combined Charles Dickens’ Corpus

Figure 7 and Table 11 provide the extracted semantic categories in the combined corpus of the four texts. Table 12 provides a list of the words within each of these semantic categories.

Figure 7: Semantic clouds in Charles Dickens’ combined corpus

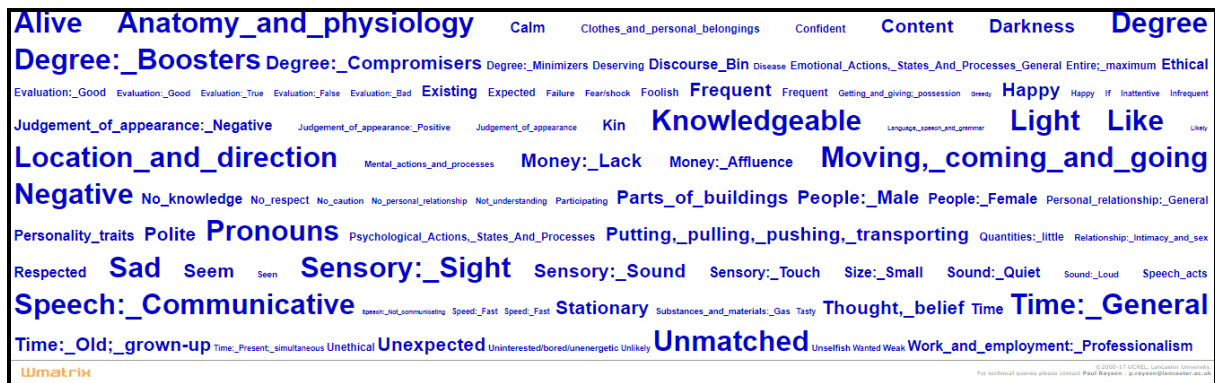


Table 11: Wmatrix Semantic Tagging of Charles Dickens’ Combined Corpus

Semantic Field	LL	Semantic Field	LL	Semantic Field	LL	Semantic Field	LL
Pronouns	+26807.	<b>People:</b>	+639	Sensory:	+218.0	Foolish	+169

*Utilizing corpus stylistics to facilitate literary analysis: An assessment of the effectiveness of semantic domains in identifying major literary themes in a selection of Charles Dickens novels*

	89	<b>Male</b>	.91	Touch	0		.36
Anatomy & Physiology	+3448.65	<b>Putting, Pulling, Pushing, Transporting</b>	+592.96	Calm	+217.66	Frequent	+167.85
Unmatched	+2839.41	Darkness	+528.89	<b>Money: Affluence</b>	+216.27	<b>No Respect</b>	+164.21
Speech: Communicative	+2358.95	Content	+509.90	Size: Small	+215.98	Degree: Minimizers	+147.57
Negative	+1348.69	Unexpected	+379.78	Existing	+214.43	Expected	+145.29
Light	+1269.87	<b>Polite</b>	+334.99	<b>People: Female</b>	+213.87	Quantities: Little	+139.60
Degree: Boosters	+1207.57	<b>Happy</b>	+293.08	Discourse Bin	+211.34	Deserving	+129.85
Like	+1191.28	Frequent	+291.43	Sound: Quiet	+201.20	Clothes & Personal Belongings	+129.41
Sensory: Sight	+1131.01	Sensory: Sound	+285.35	Time	+191.91	Speech Acts	+126.94
Degree	+1071.37	Degree: Compromises	+274.36	<b>Personality Traits</b>	+191.88	<b>Emotional Actions, States &amp; Processes General</b>	+125.33
<b>Knowledgeable</b>	+1050.76	Stationary	+269.36	No Knowledge	+188.42	<b>Confident</b>	+122.35
<b>Moving, Coming &amp; Going</b>	+900.01	Parts of Buildings	+257.95	<b>Ethical</b>	+179.48	Sound: Loud	+121.68
Time: General	+887.25	Seem	+256.60	<b>Work &amp; Employment: Professionalism</b>	+177.52	<b>Fear/ Shock</b>	+114.99
<b>Sad</b>	+788.80	<b>Money: Lack</b>	+227.19	<b>Judgement of Appearance: Negative</b>	+176.45	If	+114.26
Location & Direction	+767.73	<b>Time: Old; Grown-up</b>	+225.00	Evaluation: Good	+172.57	<b>Mental Actions &amp; Processes</b>	+105.97
Alive	+683.19	<b>Kin</b>	+219.55	<b>Unethical</b>	+171.73	Unselfish	+101.60
<b>Thought, Belief</b>	+659.67	<b>Respected</b>	+218.96	Psychological Actions, States & Processes	+171.25	Seen	+100.79

Table 12: The Combined Corpus: Significant words in each semantic category

Charles Dickens Combined Corpus	
Semantic tag	Wordlist
Knowledgeable	know- remember- knowledge - aware- recollect- information- experience-

	heard of- recall- news- awakened- recognized- wisdom- looking back- recalled- recognition- remembrances- memorable- famous- identify- recalling- call to mind- experiences
Moving, Coming & Going	returned- go- come- passed- walked- leave- fell- rose- nodded- arrived- coming back- trot- fall- journey- steps- trembling- crossed- wandering- advanced- enter- arrival- rolled- crept- approached- departed- movement
Sad	cried- miserable - unhappy- melancholy- desperate- suffered- sorrow- sad- gravely- misery- weeping- mourning- despair- suffer- wept- sobbing- suffering- grim- sobbed- remorse- sorrowful- burst into tears
Thought, Belief	think- felt- opinion- considered- suppose- wonder- impression- suspicion- trust- meaning- attitude- belief- conviction- conscience- suspicious- imagine- found- assumed- regard- guess- musing- deemed
People: Male	man- boy- gentleman- men- fellow- Mr- boys- gentlemen- chap- lad- fellows- male- Mister- bachelor-manly- messrs- manhood- chaps- masculine- lads- manful- boy- nobs- males- man - manliest
Putting, Pulling, Pushing, Transporting	put- turned- brought- lay- held- laid- shook- set- bear- dropped- turn- carriage- placed- threw- seated- hung- raising- removed- delivered- pushed- shed- put into- withdrew- pull- thrust- hanging- stuck- bearing- tossed- dragged
Polite	grateful- thanked- thankful- gracious- compliments- thank- compliment- polite- gratitude- thanks- politeness- politely- complimentary- apologetic- apologetically- graciously- gentlemanly- indebted- pleasantry- thankfulness- propriety- complimented- thanking- courteous- decently- refined- civilities- courteously- civilly- cultivated- good manners- proprieties- tact- civility- courtly- well bred- courtesy- complimenting- pleasantries
Happy	happy- smile- laughed- smiling- cheerful- delighted- relief- joy- delight- humour- merry- gaily- grin- amused- playful- beaming- sanguine- cheer up- chuckled- comic- jest- fun- rejoiced- bliss- rapturous- merriment- mirth- overjoyed- festive
Money: Lack	poor- paupers- beggar- poverty- pauper- in need- beggars- destitute- insolvent- needy- destitution- beggared- poor laws- fruitlessly- beggar woman- debasing- underpaid- pauper- debased- beggar- looking poor- penniless
Time: Old; Grown-up	old- mature- old fashioned- grown up- pensioner- matured- old age- matronly- adult- maturer- looked old- maturity
Kin	aunt- mother- father- sister- family- wife- husband- married- brother- daughter- son- engaged- marriage- papa- uncle- marry- mama- godmother- cousin- parent- niece- bride- mum- marrying- parents- wedding- widow- twins- sisters- aunts
Respected	respect- admiration- dignity- admired- honoured- stately- praise- homage- dignified- esteem- deference- praised- reverence- deferential- praises- self-respect- estimable- glorified
Money: Affluence	fortune- fortunes- treasure- riches- richest- fortune hunter
People: Female	lady- girl- woman- miss- female- housewife- feminine- girlish- hag- womanhood- crones <sup>8</sup> - begum- lass- ladylike- crone- maiden
Personality Traits	kind- childish- unjust- temperament- human nature- kinder- unkind- childishly- stubborn- trait- stiff necked- pig headedness- unjustly
Ethical	honour- fair- forgive- noble- respectable- mercy- virtue- gallant- innocently- disinterestedness- scruple- acquitted- humane- fairest- ingenuous- blameless- chaste
Work & Employment:	secretary- reputation- carpenter- accountant- repute- practitioners- malpractices- colleague- bricklayer- carpentering- gatekeeper- colleagues

<sup>8</sup> Croness means ugly old women. Begum refers to a Muslim woman of high rank.

Professionalism	
Judgement of Appearance: Negative	wretched- awful- dismal- dirty- ragged- ugly- horrible- desolate- shabby- frightful- duff- hideous- unpleasant- unspeakable- withered- imperfectly- clumsy- wreck- abhorrence- nasty- harsh- imperfect- bleak- filthy- horrid- gaunt
Unethical	ashamed- evil- shame- wicked- rogue- mercenary- betrayed- tricks- mischief- rascal- naughty- reproachful- sordid- sin- vice- guilt- degradation- wickedness- devilish- corruption
No Respect	disgrace- contempt- scorn- disdain- impudent- insolent- affront- degraded- humiliation- dishonourable- dishonour- abase- derision- irreverent- undignified
Emotional Actions, States & Processes General	pity- temper- sentiment- emotion- compassion- flushed- wits- awe- touching- shudder- moody- hysterical- wit- tempers- feel- discomposed- rosy cheeked
Confident	confidence- trust- faith- emphatic- confident- emphatically- peace of mind- forcibly- trustfulness- reassuring- forcible- minds- reassured
Fear/Shock	fear- frightened- dread- terror- alarmed- horror- shock- timid- fearful- startled- shy- terrified- fright- shocked- timidly- scared- frighten- coward
Mental Actions & Processes	memory- mental- dreamed- intellect- dreaming- meditation- mentally- intellectual- dream- meditations- memories- intuitively

The furnishing of the story worlds of Dickens' selected novels is reflected by the semantic categories **People: Male** (+639.91) and **People: Female** (+213.87). The high statistical significance of the **People: Male** semantic category reflects the fact that the majority of the inhabitants of these story worlds are males. The category **Personality Traits** (+191.88) elaborates Dickens' characters even further. The categories **Moving, Coming & Going** (+900.01) and **Putting, Pulling, Pushing, Transporting** (+592.96) reflect the actions of the stories. The categories of **Knowledgeable** (+1050.76) and **Thought, Belief** (+659.67) relate to the theme of education. They may also relate to the theme of self-development and progress to maturity which is primarily conveyed by the categories **Time: Old; Grown-up** (+225.00), **Emotional Actions, States & Processes General** (+125.33), and **Mental Actions & Processes** (+105.97). The theme of discrepancies in the social class system, which is characterized by the presence of two extremes: affluence and poverty, is reflected by the semantic categories **Sad** (+788.80), **Money: lack** (+227.19), **Respected** (+218.96), **Money: Affluence** (+216.27), **Work & Employment: Professionalism** (+177.52), **Judgement of appearance: Negative** (+176.45), and **Fear/Shock** (+114.99). The theme of morality is reflected in the categories **Polite** (+334.99), **Ethical** (+179.48), **Unethical** (+171.73), **No Respect** (+164.21), and **Confident** (+122.35). The theme of family is conveyed by the category **Kin** (+219.55). Finally, the semantic category **Happy** (+293.08) relates to Dickens' positive attitude in the face of suffering and his tendency to have happy endings where evil is always punished and morality is always rewarded.

## CONCLUSION

The objective of this study was to address the effectiveness of semantic domains in detecting the major themes of literary texts. Through suggesting a methodology consisting of a synergy of quantitative and qualitative approaches, the study managed to contribute to the evidence supporting the effectiveness of corpus tools, particularly Wmatrix, in supplementing traditional stylistic analysis. The results of the corpus analysis indicate the effectiveness of Corpus Stylistics in detecting literary texts. This study recommends the use of corpus-based

analysis of literary texts as a milestone step to provide quantitative evidence for further qualitative assessments.

The implications of this study can extend to the field of pedagogy since the corpus-stylistic approach to studying literary themes, being an evidence-based approach, may facilitate the teaching of literary works objectively and, hence, their study for non-native readers. According to Semino (2011), the pedagogical implications for stylistic analysis have been examined by stylisticians in respect to native speaker, second language and foreign language contexts (e.g., Widdowson, 1975; Widdowson, 1992; Watson & Zyngier, 2007). In fact, the effectiveness of using stylistic evidence-based approaches in teaching literary texts in language classrooms is advocated in many studies (Carter & McRae, 1996; Carter & Simpson 1989; Cook, 1994; Fowler, 1996; Kramersch, 1993; Kramersch & Kramersch 2000; Short, 1996; Simpson, 1993; Toolan, 1998; Widdowson, 1975, 1992).

“Herein lies its educational value – for it offers an alternative to the traditional teaching of literature. Rather than being the passive recipients of the second hand interpretations of literary critics, students can be enabled (empowered even) to take the initiative and engage actively and directly with literary texts themselves” (Widdowson, 2008, p. 302).

Applying the procedure of semantic domains to literary works would, in essence, advance non-native readers’ comprehension of these works by, and probably also contribute further to, developing their level of English language proficiency through relating themes to linguistic triggers. Behnam (1996, p. 12) asserts that when non-native learners do not only depend on “some intuitive responses for which little evidence can be drawn”, they may activate “a general analytic skill which can serve as a first step towards [the] ultimate interpretative responses expected from readers”. Naciscione (2010, p. 205) adds that literary texts can be used as a significant medium for language acquisition and awareness raising, through focusing on “lexical and grammatical accuracy” and additionally enhancing “literary awareness and stylistic sensitivity”. Hernández (2011, p. 235-6) argues that using literary texts in language teaching facilitates the development of “literary comprehension and sensibility” in learners, and elevates their communicative competence through “the expression of ideas, opinions, and beliefs”. The “use of literary texts is often advocated as a means to enhance proficiency in reading, vocabulary growth and cultural knowledge, if not indeed, in more traditional systems” (Hall, 2007, p. 4).

The approach suggested in this paper seeks to achieve “two main objectives: firstly, to enable students to make meaningful interpretations of the text itself; secondly, to expand students’ knowledge and awareness of the language in general” (Lazar, 1993, p. 31). Hence, practicing this corpus-stylistic approach would likely extend teaching literature in the classroom beyond traditional literary criticism, which primarily depends on intuition, to encompass more systematic and objective practices, and eventually help develop non-native learners’ English language proficiency.

## REFERENCES

- Ackroyd, P. (1990). *Dickens*. London: Sinclair-Stevenson.
- Adolphs, S. (2006). *Introducing Electronic Text Analysis. A Practical Guide for Language and Literary Studies*. London: Routledge.
- Adolphs, S. and Carter, R. (2002). “Corpus stylistics: Point of view and semantic prosodies in Virginia Woolf’s *To the Lighthouse*”. *Poetica* 58: 7-20.

- Archer, D. and Bousfield, D. (2010) “‘See Better, Lear’? See Lear Better! A Corpus-Based Pragma Stylistic Investigation of Shakespeare’s *King Lear*”, in McIntyre, D. and Busse, B. (eds.) *Language and Style*, pp. 183-203. Basingstoke: Palgrave.
- Bagehot, W. (1858). “Charles Dickens”, *National Review*, vol.7, 1858, in Stephen Wall, (ed). 1970. *Charles Dickens: A Critical Anthology*. Middlesex: Penguin Books Ltd.
- Behnam, B. (1996). *A stylistic study of cohesive features in English prose fiction with some pedagogical implications for non-native contexts*. Sheffield: Department of English Language and Linguistics, University of Sheffield. PhD thesis.
- Bell, M. (2000). *Sentimentalism, Ethics and the Culture of Feeling*. Basingstoke: Palgrave.
- Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan: University of Michigan Press.
- Biber, D. (2011). “Corpus linguistics and the study of literature: Back to the future?” *Scientific Study of Literature* 1(1): 15-23.
- Carter, R. and McRae, J. (1996). *Language, literature and the learner: Creative classroom practice*. London: Routledge.
- Carter, R. and Simpson, P. (eds). (1989). *Language, discourse and literature: An introductory reader in discourse stylistics*. London: Routledge.
- Dickens, C. (2008). *David Copperfield*, ed. by Nina Burgis. Oxford: Oxford University Press.
- Dickens, C. (2008). *Oliver Twist*, ed. by Kathleen Tillotson. Oxford: Oxford University Press.
- Dickens, C. (2008). *Our Mutual Friend*, ed. by Michael Cotsell. Oxford: Oxford University Press.
- Dickens, C. (2008). *Great Expectations*, ed. by Margaret Cardwell. Oxford: Oxford University Press.
- Choudhury, B. (2009). *Charles Dickens: Great Expectations*. New Delhi: PHI Learning Private Limited.
- Churchill, R. C. (1975). *A Bibliography of Dickensian Criticism 1836-1975*. London, Basingstoke, and New York: Macmillan.
- Cook, G. (1994). *Discourse and literature: The interplay of form and mind*. Oxford: Oxford University Press.
- Culpeper, J. (2002). “Computers, language and characterisation: An analysis of six characters in *Romeo and Juliet*”. In U. Melander-Marttala, C. Östman and M. Kytö (eds) *Conversation in Life and in Literature*, pp. 11-30. Uppsala: Universitetstryckeriet.
- Culpeper, J. (2002). “Keyness: Words, Parts-of-Speech and Semantic Categories in the Character-talk of Shakespeare’s *Romeo and Juliet*”. *International Journal of Corpus Linguistics* 14 (1): 29-59
- Fischer-Starcke, B. (2010). *Corpus linguistics in literary analysis: Jane Austen and her contemporaries*. London: Continuum.
- Fowler, R. (1996). *Linguistic criticism*. Oxford: Oxford University Press.
- Gilquin, G. (2010). *Corpus, cognition and causative constructions*. Amsterdam: John Benjamins Publishing Company.
- Hall, G. (2007). “Stylistics in Second Language Contexts: A Critical Perspective”. In Greg Watson and Sonia Zyngier (eds.), *Literature and stylistics for language learners: theory and practice*, pp. 91–105. Basingstoke: Palgrave Macmillan.
- Hernández, P. S. (2011). “The potential of literacy texts in the language classroom: The study of linguistic functions”. *Odisea: Revista de Estudios Ingleses*(12), 233-244.
- Ho, Y. (2011). *Corpus Stylistics in Principles and Practice: A Stylistic Exploration of John Fowles' The Magus*. London: Continuum.

- Hoover, D. L. (1999). *Language and style in The Inheritors*. Lanham, Md: University Press of America.
- Hoover, D. L. (2002). "Frequent word sequences and statistical stylistics". *Literary and Linguistic Computing* 17(2): 157-180.
- Hori, M. (2004). *Investigating Dickens' Style. A Collocational Analysis*. Basingstoke: Palgrave Macmillan.
- Hunt, D. and Carter, R. (2012). "Seeing through the bell jar: Investigating linguistic patterns of psychological disorder". *Journal of Medical Humanities* 33(1), 27-39.
- Johnson, E. (1977). *Charles Dickens: His Tragedy and Triumph*. New York: The Viking Press.
- Keshabyan-Ivanova, I. (2014). "A computer-assisted analysis of literary texts: a sample study". *Computer modelling & new technologies* 18(6), 59-64.
- Kramersch, C. (1993). *Context and culture in language teaching*. Oxford: Oxford University Press.
- Kramersch, C. and Kramersch O. (2000). "The avatars of literature in language study". *The Modern Language Journal* 84(4), 553-573.
- Lawson, A. (2000). "Die schöne Geschichte': a corpus-based analysis of Thomas Mann's Joseph und seine Brüder". In B. Dodd (ed.) *Working with German Corpora*, pp. 161-80. Birmingham: University of Birmingham Press.
- Lazar, G. (1993). *Literature and language teaching: a guide for teachers and trainers*. Cambridge: Cambridge University Press.
- Leech, G. (1992). "Corpus linguistics and theories of linguistic performance". In Jan Svartvik (ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82*, Stockholm, 4-8 August 1991, 105-122. Berlin: Mouton de Gruyter.
- Leech, G. (2008). *Language in Literature: style and foregrounding* Harlow, UK: Pearson.
- Louw, B. (1993). "Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies". In M. Baker, G. Francis and E. Tognini-Bonelli (eds.), *Text and Technology: In honour of John Sinclair*, pp. 155-176. Amsterdam: John Benjamins Publishing.
- Mahlberg, M. (2007a). "Clusters, key clusters and local textual functions in Dickens". *Corpora*(2) 1, pp. 1-31.
- Mahlberg, M. (2007b). "Corpus stylistics: bridging the gap between linguistic and literary studies". In Michael Hoey, Michaela Mahlberg & Michael Stubbs (eds.), *Text, Discourse and Corpora: Theory and Analysis*. London: Continuum.
- Mahlberg, M. (2010). "Corpus Linguistics and the Study of Nineteenth-Century Fiction". *Journal of Victorian Culture* 15(2): 292-298.
- Mahlberg, M. (2013). *Corpus Stylistics and Dickens' Fiction*. New York: Routledge.
- Mahlberg, M. (2014). "Corpus stylistics". In M. Burke (ed.), *The Routledge Handbook of Stylistics*, pp. 378-392. London: Routledge.
- Mahlberg, M. and McIntyre, D. (2011). "A case for corpus stylistics: Ian Fleming's *Casino Royale*". *English Text Construction* 4(2): 204-227.
- Mahlberg, M. and Smith, C. (2010). "Corpus Approaches to Prose Fiction: Civility and Body Language in *Pride and Prejudice*". In D. McIntyre and B. Busse (eds.), *Language and Style*, pp. 449-467. Basingstoke: Palgrave Macmillan.
- McEnery, T. and Gabrielatos, C. (2006). "English corpus linguistics". In Bas Arts and April McMahon (eds.), *The handbook of English linguistics*, pp. 33-71. London: Blackwell Publishing.
- McEnery, T. and Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- McEnery, A., R. Xiao and Y. Tono. 2006. *Corpus Based Language Studies. An Advanced*

- Resource Book. London: Routledge.
- McIntyre, D. (2008). "Integrating multimodal analysis and the stylistics of drama: a multimodal perspective on Ian McKellen's *Richard III*". *Language and Literature* 17(4): 309-334.
- McIntyre, D. (2010). "Dialogue and Characterization in Quentin Tarantino's *Reservoir Dogs*: A Corpus Stylistic Analysis". In Dan McIntyre & Beatrix Busse (eds.), *Language and Style*, pp. 162–182. Basingstoke: Palgrave Macmillan.
- McIntyre, D. (2015). "Towards an integrated corpus stylistics". *Topics in Linguistics* (16)1: 59-68.
- McIntyre, D. and Walker, B. (2010). "How can corpora be used to explore the language of poetry and drama?" In A. O'Keeffe and M. McCarthy (eds), *The Routledge Handbook of Corpus Linguistics*, pp. 516–530. London: Routledge.
- Merchant, P. and Waters, C. (2015). *Dickens and the Imagined Child*. Ashgate Publishing, Ltd.
- Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Murphy, S. (2007). "Now I am alone: A corpus stylistic approach to Shakespearian soliloquies". In Costas Gabrielatos, Richard Slessor & J.W. Unger (eds.), *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching*, vol. 1, 67–85. Lancaster: Department of Linguistics and English Language, Lancaster University.
- Naciscione, A. (2010). *Stylistic use of phraseological units in discourse*. Amsterdam: John Benjamins Publishing Company.
- O'Halloran, K. (2007a). "The subconscious in James Joyce's 'Eveline': a corpus stylistic analysis that chews on the 'Fish hook'". *Language and Literature* 16(3): 227-244.
- O'Halloran, K. (2007b). "Corpus-assisted literary evaluation". *Corpora* 2(1): 33-63. doi:10.3366/cor.2007.2.1.33.
- O'Keeffe, A. (2006). *Investigating media discourse*. London: Routledge.
- Oliphant, Margaret. (1862). "Sensation Novels", *Blackwood's Magazine*, vol.91, 1862, in Stephen Wall, ed. 1970. *Charles Dickens: A Critical Anthology*, Middlesex: Penguin Books Ltd.
- Orwell, G. (1968). "Charles Dickens". In George Orwell, Sonia Orwell and Ian Angus (eds), *The Collected Essays, Journalism and Letters of George Orwell*. London: Secker and Warburg.
- Raina, Badri. (1986). *Dickens and Dialectic of Growth*. Madison, Wisconsin: The University of Wisconsin Press.
- Rayson, P. (2003). *Matrix: A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison*. Ph.D. Thesis. Lancaster University.
- Rayson, P. (2008). "From key words to key semantic domains". *International Journal of Corpus Linguistics* 13(4), 519–549.
- Rayson, P. (2021). *Wmatrix5: A Web-based Corpus Processing Environment*. Computing Department, Lancaster University. <https://ucrel-wmatrix5.lancaster.ac.uk/wmatrix5.html>
- Romaine, S. (2010). "19<sup>th</sup> Century key words, key semantic domains and affect: In the rich vocabulary of love 'most dearest' be a true superlative". *Studia Neophilologica* 82(1): 12-48.
- Schuster, S. (2014). *An Analysis of Childhood and Child Labour in Charles Dickens' Works: David Copperfield and Oliver Twist*. Hamburg: Anchor Academic Publishing.
- Scott, M. (2006). "Key words of individual texts". In M. Scott and C. Tribble (eds), *Textual Patterns: Key Words and Corpus Analysis in Language Education*, pp. 55–72.



- Amsterdam: John Benjamins.
- Scott, M. and Tribble, C. (2006). *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.
- Semino, E. (2011). "Stylistics". In James Simpson (ed.), *The Routledge handbook of applied linguistics*, 541–554. London: Routledge.
- Semino, E. and Short, M. (2004). *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English writing*. London and New York: Routledge.
- Short, M. (1996). *Exploring the language of poems, plays, and prose*. London: Longman.
- Simpson, P. (1993). *Language, ideology, and point of view*. London: Routledge.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.
- Sinclair, J. (2007). "Introduction". In Michael Hoey, Michaela Mahlberg and Michael Stubbs (eds.), *Text, discourse and corpora: Theory and analysis*. London: Continuum
- Starcke, B. (2006). "The phraseology of Jane Austen's *Persuasion*: phraseological units as carriers of meaning". *ICAME Journal* 30: 87-104.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- Stubbs, M. (2005). "Conrad in the Computer: Examples of Quantitative Stylistic Methods". *Language and Literature*, 14(1): 5-24.
- Taine, H. "Charles Dickens, son talent et ses oeuvres", *Revue des Deux Mondes* February 1856, (translated by H. Van Laun, 1871, in *History of English Literature*). In Stephen Wall, ed. 1970. *Charles Dickens: A Critical Anthology*. Middlesex: Penguin Books.
- Taylor, C. (2008). "What is corpus linguistics? What the data says". *ICAME Journal*(32), 179-200.
- Toolan, M. (1998). *Language in literature: An introduction to stylistics*. London: Routledge.
- Toolan, M. (2006). "Top keyword abridgements of short stories: A corpus linguistic resource?" *Journal of Literary Semantics* 35(2): 181-194.
- Trollope, A. (1870). "Charles Dickens", *St Paul's Magazine*, vol.6, 1870, in Stephen Wall, ed. 1970. *Charles Dickens: A Critical Anthology*. Middlesex: Penguin Books Ltd.
- Walder, D. (1981). *Dickens and Religion*. London: George Allen & Unwin Ltd.
- Walker, B. (2010). "Wmatrix, Key-Concepts and the Narrators in Julian Barnes' *Talking It Over*". In D. McIntyre and B. Busse (eds) *Language and Style*, pp. 364-387. Palgrave MacMillan, Basingstoke.
- Walker, B. (2012). *Character and Characterisation in Julian Barnes' Talking it over: A Corpus Stylistic Analysis*. Ph.D. Thesis. Lancaster University.
- Wall, S. (1970). *Charles Dickens: A Critical Anthology*. Middlesex: Penguin Books Ltd.
- Watson, G. and Zyngier, S. (eds). (2007). *Literature and stylistics for language learners: Theory and practice*. Basingstoke: Palgrave Macmillan.
- Widdowson, H. G. (1975). *Stylistics and the teaching of literature*. London: Routledge.
- Widdowson, H. G. (1992). *Practical stylistics: An approach to poetry*. Oxford: Oxford University Press.
- Widdowson, H. G. (2008). "The novel features of text: Corpus analysis and stylistics". In Andrea Gerbig & Oliver Mason (eds.), *Language, people, numbers, corpus linguistics and society*, pp. 293-304. Amsterdam: Rodopi.
- Wynne, M. (2006). "Stylistics: Corpus Approaches" in K. Brown (ed. in chief) *The Encyclopaedia of Language and Linguistics*, pp. 223–26. Oxford: Elsevier.