

## Research Article

Basant S. M. Moustafa\*

# A comparative corpus stylistic analysis of thematization and characterization in Gordimer's *My Son's Story* and Coetzee's *Disgrace*

<https://doi.org/10.1515/opli-2020-0183>

received January 19, 2021; accepted December 22, 2021

**Abstract:** *My Son's Story* and *Disgrace* are two novels written by two South African Noble laureates – Nadine Gordimer and J. M. Coetzee, respectively. The current study attempts to comparatively uncover the thematic foci as well as characterization aspects and interpersonal relations between characters in the two novels by dint of corpus linguistic tools, i.e., keywords and key clusters. Thus, using Sketch Engine online interface, the study presents a comparative thematic categorization of the two novels – a categorization which proves congruent with the thematization provided by previous critical literary studies of each novel. Both novels are found to revolve around racial tensions and illicit relationships in South Africa. However, although *My Son's Story* is set during the apartheid and *Disgrace* is set after the apartheid is supposedly over, comparative corpus-driven investigation of the two novels reveals that the South Africa promised in *My Son's Story* is betrayed in *Disgrace*, since racial violence and cross-race sexual dissipation persist. Quantitatively, using the notion of local textual functions, the associations between the use of key clusters, their repercussions on characters' depiction, and their interpersonal relationships are uncovered. Clear affinities between the two protagonists, Sonny and David, are discerned, e.g., their love for music and literature, their sexual whims, their unsettled relationships with their children, etc. Empirically, the father–son/daughter turbulences, husband–wife vicissitudes, landlord–farmer intricacies, and love affairs' intricacies are fathomed.

**Keywords:** characterization, thematization, corpus stylistics, local textual functions, race, illicit love, Noble Laureates, South Africa, apartheid

## 1 Introduction

Ho (2011) and Fischer-Starcke (2010), among many others, define corpus stylistics as the use of corpus linguistics tools in analyzing literary texts. McIntyre and Walker (2019, 14–15) add that the application of corpus tools to the analysis of texts enhances the insights attained through traditional stylistics and contends that corpus stylistics is the use of stylistics frameworks, theories, and models in corpus analysis. Basically, an examination of a writer's individual style is partially based on examining the degree of the writer's inclination to use “particular ways of putting things.” To scrutinize the stylistic features of a

---

\* **Corresponding author: Basant S. M. Moustafa**, Department of English Language and Literature, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Kingdom of Saudi Arabia; Department of English Language and Literature, Faculty of Arts, Tanta University, Tanta 31511, Egypt, e-mail: b.moustafa@psau.edu.sa, basant.moustafa@art.tanta.edu.eg  
ORCID: Basant S. M. Moustafa 0000-0002-9910-3450

particular writer, preferably a comparison is to be made with the literary work of another as is the case in literary stylistics where “distinctions can be made with regard to the style of a particular work, a particular author, a period, and so on” (Mahlberg 2007a, 220). In this respect, corpus analysis has become a valuable tool that allows for a comparison between individual works, collections, or a specific language variety (McEnery and Wilson 2001, 117).

The present study aims at scrutinizing two outstanding literary works: *My Son's Story* (1990) by Nadine Gordimer and *Disgrace* by J. M. Coetzee. The study attempts to answer the following questions: (1) what thematic foci in *My Son's Story* and *Disgrace* (1999) can be identified by means of corpus techniques? (2) How can the concept of local textual functions in the two novels be examined using corpus tools? (3) What aspects of characterization and interpersonal relations in the two novels are revealed through corpus-driven analysis? To answer these questions, the study adopts a corpus stylistics approach whereby a thematization process is comparatively worked out through keyword analysis (Mahlberg and McIntyre 2011), and local textual functions of key clusters in the two novels are specified and investigated as a way into recognizing enriched characterization and interpersonal relations between characters (Mahlberg 2007a, 2007b). The present research argues that a quantitative approach represented by corpus linguistic tools can readily yield interesting insights into thematization and characterization when investigating two literary works in comparative terms, capturing potential aspects of comparison and contrast. These insights can empirically sustain literary interpretations and contribute toward a rather objective means of analyzing a fictional work.

## 2 Literature review

Recently, the use of corpus tools to analyze literary texts has been evidently increasing, e.g., Adolphs and Carter 2002, Hori 2004, Semino and Short 2004, Stubbs 2005, Culpeper 2009, McIntyre and Walker 2010, Mahlberg et al. 2019. Quantitatively, corpus linguistics “allows for decoding meanings of literary texts that cannot be detected either by intuitive techniques as in literary studies or with the necessary restriction to short texts or text extracts as in traditional stylistics” (Fischer-Starcke 2010, 1). However, though a stylistic analysis can be enabled through statistical means, it is never constituted by them (Leech and Short 2007). Most corpus stylistic research concentrates on word distributions with the aim of recognizing textual characteristics that are peculiar to a specific author, text, or fictional character. To this purpose, three methodological corpus tools are used: keyword analysis, extended phrases, and collocational analysis (Biber 2011, 16).

### 2.1 Keywords

Keywords are words that are significantly higher in frequency in a corpus than in a reference corpus (Mahlberg and McIntyre 2011, 207). Keyness is based on statistically comparing each word in the corpus under investigation against a reference corpus (Baker 2014, 13). The quantitative results obtained through comparison with a reference corpus are relevant to the foregrounding effects described by stylisticians (Mahlberg and McIntyre 2011, 211). Scott and Tribble (2006, 5) explain that the keywords of a given text reveal lexical patterns which elucidate the text content. Moreover, keywords of a text provide lexical items that can be good starting points for further examination, i.e., they act as “signposts,” providing the analyst with a “way in” to the corpus (Baker 2006, 125) or “signals for the building of fictional worlds as well as triggers for thematic concerns of the novel” (Mahlberg and McIntyre 2011, 207), i.e., keywords serve as pointers to literary meanings.

In this sense, Culpeper (2002) uses keywords to characterize the protagonists of Shakespeare's *Romeo and Juliet* by comparing the play with a corpus comprising all of Shakespeare's plays. He finds that the use of first-person singular and plural pronouns gives insights into the social status and personalities of

protagonists. Culpeper (2009) develops his analysis of *Romeo and Juliet* by analyzing the semantic categories of keywords and parts of speech concerning the speeches of individual characters. To this end, he uses WMatrix software package (Rayson 2008), which depends on tagging each word in the corpus according to its semantic domain.

In her analysis of Austen's novel *Pride and Prejudice*, Fischer-Starcke (2009) quantitatively analyzes keywords and frequent phrases to conclude that collocation and colligation of keywords and phrases can be used to interpret literary meaning of family relations in the novel. Mahlberg and McIntyre (2011) also probe keywords and key semantic domains in Fleming's *Casino Royale* using WMatrix software package (Rayson 2008). They categorize resultant keywords into two groups: fictional world and thematic signals. These groups in turn fall into two broad categories marked as text-centered and reader-centered. They also compare the manually classified keyword groups with the automatically recognized key semantic domains and conclude that keywords that do not fit into a semantic domain are regarded as reader-centered keywords with particular textual functions.

Ho (2011) uses a corpus stylistic approach to compare the 1966 version of John Fowler's *The Magnus* to the 1977 version of the same novel. To this end, she utilizes four different corpus tools: TESAS/Crouch and WCopyfind for measuring text similarity, Wmatrix for identifying semantic difference and different patterns in the narrative presentation of point of view, and WordSmith Tools for recognizing patterns of figurative language in the two versions. Focusing on salience and foregrounding in David Pearce's novel *1974*, McIntyre and Walker (2019) use WMatrix to extract and analyze keywords according to the concept of foregrounding. In their analysis of keywords, they ignored characters' names since they would always appear as key against the reference corpus. They conclude that keyness or statistical salience in itself does not constitute interpretive significance or stylistic foregrounding.

## 2.2 Clusters

Another concept that has been used in stylistic analyses is that of "cluster." A cluster is "a repeatedly occurring sequence of words" (Mahlberg 2014, 285, see also Biber et al. 2004, Partington and Morley 2004). Clusters are also called "*n*-grams" where *n* refers to the number of repeated occurrence of words. Corpus stylisticians stress that clusters can be regarded as pointers to textual encoded meanings and functions, and that readers possibly perceive the encoded meanings unconsciously. Therefore, a cluster analysis helps to uncover these meanings in a text. Mahlberg (2007a, 225) argues that "the occurrence of clusters reflects the functional relevance of these sequences in texts" (see also Biber and Conrad 2005, Mahlberg 2005). The concept of "lexical bundles" is akin to clusters (Biber et al. 1999). They are "the most frequent recurring lexical sequences in a register" (Biber et al. 2004, 376). So, besides referring to repeated sequences of words, lexical bundles also concentrate on those sequences which have a high frequency of occurring. As argued by Mahlberg (2014, 285), "[t]he frequency with which lexical bundles occur is a reflection of the textual functions that they fulfil in texts."

Analyzing Jane Austen's *Persuasion*, Fischer-Starcke (2006) examines the most frequent 3-grams (or 3-word clusters) in the novel. She also studies 3-frames which are "3-grams with a variable slot indicated by the wild card \*, as in the 3-frame *the \* of*" (Mahlberg 2014, 285). Fischer-Starcke's conclusion relates to patterns associated with the relationships amongst characters and the novel's atmosphere. Fischer-Starcke (2010) uses the software *kfNgram* to extract the most frequent 4-grams and 4-frames of three corpora: Jane Austen's *Northanger Abbey*, *Austen* (a corpus which consists of six novels by Austen), and *ContempLit* (a 4,370,000 word corpus which consists of literary works published between 1740 and 1859). By identifying the lexical and grammatical patterns in the corpora, insights into the encoded meanings, the information structures, and the textual organization of the texts are gained.

Mahlberg et al. (2019) explore the "fictional speech-bundles" using three corpora of nineteenth-century fiction provided by the corpus stylistic web application CLiC (Corpus Linguistics in Context), which allows the identification of "quotes" and "non-quotes" subsets of the corpora. Mahlberg et al. focus on text

between quotation marks, which represent the direct speech in the corpora. Hence, by comparing the quotes across the corpora with the spoken subset of BNC 1994, fictional speech-bundles are described “on a continuum from lexical bundles in real spoken language to repeated sequences of words that are specific to individual fictional characters” (Mahlberg et al. 2019, 326).

## 2.3 Local textual functions

Another concept that relates to the use of corpus techniques in literary text analysis is “local textual functions.” Local textual functions refer to “the meanings of lexical items in texts.” Their locality stems from the fact that the functions “do not claim to capture general functions, but functions specific to a (group of) text(s) and/or specific to a (group of) lexical item(s).” They are “textual” because “they describe the meanings of lexical items in texts” (Mahlberg 2007b, 193, see also Mahlberg 2005, 2007b, 2012, 2014).

Mahlberg (2007a) examines a corpus of 23 novels by Dickens using Wordsmith tools to extract 8-word clusters. Then, focusing on *Bleak House*, she scrutinizes the local textual functions of 5-word clusters. Comparing 5-word clusters which occur minimally five times in Dickens corpus with 5-word clusters in a reference corpus composed of 29 novels by 18 nineteenth-century writers, five functional groups of clusters are found to be particular to Dickens’ style, that is, labels, speech clusters, body part clusters, *as if* clusters, and time and place clusters. These groups correspond to five broad functions: references to characters, interaction between characters, description of characters’ body movements, construction of a textual world through comparison, and temporal and spatial references. Typically, when the research focuses on individual texts, local functions of clusters are investigated in detail. On the other hand, when two texts are compared, clusters are probed to sketch the functional differences between the two texts.

Mahlberg (2012) looks at local textual functions in Dickens’s fiction to demonstrate that clusters can be interpreted as textual building blocks. The local textual functions identified contribute to the interpretation of meanings in literary worlds, i.e., by examining the 5-word clusters in Dickens, Mahlberg relates the functionality of lexical bundles to the construction of characters and themes, the relationship between characters, character’s body language, narrator’s voice, and temporal and spatial references. The clusters are first used to recognize the functional relevant groups of words, then a concordance analysis of these clusters helps to feather the patterns related to local textual functions.

## 2.4 *My Son's Story* and *Disgrace*

The present study aims at probing two prominent novels: Gordimer’s *My Son's Story* and Coetzee’s *Disgrace*. *My Son's Story*, henceforth *MSS*, was published in 1990 toward the end of the apartheid, i.e., the racial segregation system that prevailed in South Africa from 1948 till the early 1990s, while Coetzee’s *Disgrace*, hereafter *DS*, was published in 1999 in the post-apartheid South Africa. Although there is a lack of any corpus-informed research on *MSS* and *DS*, whether separately or in comparison, there are several literary studies which tackle multivarious aspects of the two novels. To mention a few, Sakamoto (2002) focuses on the notion of colored identity as a locus of cultural transformation in *MSS*. He stresses Will’s deep self-awareness of his colored identity and Sonny’s sense of “unbelonging” since he is of “mixed” race and, thus, feels “colored” rather than “black.” The family’s sense of “unbelonging,” “in-betweenness,” and “ambivalence” is their way into cultural and political transformation. Will’s realization of Sonny’s affair with Hannah represents an evolvment in the story around the notion of deceptive appearance. Sakamoto also argues for a parallel between sexual commitment and political struggle, i.e., Sonny and Hannah’s sexual relationship is concomitantly associated with their political strife, and Sonny’s status as a political leader is correlated with his sexual power and virility. Gender role transformation is also explored by Sakamoto, as Aila transforms from the traditional gender role of a subordinate wife and a loving mother to a rebellious struggling figure.

Golden (2019) argues that for Gordimer, literary works can be used to create new mediums of justice against racial violence, i.e., through what Golden calls “zero chronotope,” *MSS* is used by Gordimer to highlight the potential of fiction to “force the law,” regardless of any ideological, temporal, and spatial confinements. According to Golden, the distinction between legal language and literary language is false since such a distinction would imply that literary language cannot be conveyed to political thinking. Möller (1999) explores how the characters, events, situations, the political, the moral, and the final arrangement of events in *MSS*, rather than their sequential arrangement, are connected together as interactive constituents of ironic discourse. In the process, Möller captures the Shakespearean underpinning of the novel and its relationship with *King Lear*, *Hamlet*, *Othello*, etc.

As for *DS*, Rinzler (2013) discusses the notion of Coetzee as a Beckettian author, arguing that a neutral, brief, and concise style can be more powerful than an ornated, fanciful, and elaborated one. The neutrality of Coetzee, according to Rinzler, enables him to erase his traces in the novel and manage conveying his intentions masterfully. Coetzee, thus, manages to tell more by telling less. He uses different linguistic devices to impersonalize his presence in the novel; this is what Deleuze terms “the fourth person.” These devices include short passive, lexical neutrality, lexical repetition, grammatical repetition, and adjectival reformulation for qualifying. Attridge (2002) sheds light on the thematic concerns of *DS* and relates them to the socio-political scene in South Africa after the apartheid. Oriaku (2016) argues that David’s masculinist behavior is a reminiscent of the apartheid era while Lucy’s overcoming of victimhood signals her willingness to make a necessary compromise for a post-apartheid South Africa. The multiracial identity of her unborn, rape-result baby is also portrayed as an outcome of a post-apartheid era. In so far as it deals with morality and ethics, *DS* is abundant with allegories and rich in ambiguities, according to Oriaku.

### 3 Research methodology

Corpus stylistics does not claim inclusiveness of all linguistic characteristics in the text. Rather, it aims at endowing insights into the text by focusing on selected linguistic patterns in the text (Fischer-Starcke 2010, 197). Its value lies in identifying the connection between quantitative results and qualitative analysis (Mahlberg 2010, 295). Moreover, it comes as a solution for the long-standing difficulty of stylistic analysis, that is, the issue of length and that of selecting examples and extracts to examine (Mahlberg and McIntyre 2011, 205, Leech and Short 2007, 2). The current study follows two corpus stylistic models: Mahlberg and McIntyre (2011) and Mahlberg (2007a, 2007b). These two particular models are necessitated by the nature of research questions, i.e., the two models enable the exploration of thematization and categorization in literary works by means of corpus tools. Hence, the rationale behind this methodological selection unfolds: the research questions fit perfectly into the models specified. The following two subsections shed light on the corpus under investigation and the procedure followed in analysis.

#### 3.1 Data

The current analysis draws on quantitative and qualitative methods as a way to comparatively analyze the two appointed corpora: *MSS* and *DS*. The former is composed of 8,232 types and 86,164 tokens, while the latter encompasses 7,452 types and 66,142 tokens. The rationale behind the selection of the two specified novels is their conspicuous comparability as to the following three respects: (1) socio-culturally, Gordimer and Coetzee are, most outstandingly, the only South African novelists who were awarded the Nobel Prize in 1991 and 2003, respectively, and the two novels were explicitly mentioned in the award press release by the Swedish Academy, (2) one novel deals with South Africa during the apartheid and the other after it; the fact that allows for a scope for comparability, and (3) the approach that the current study draws upon, i.e., corpus stylistics, is inherently comparative in nature (Mahlberg 2012).

## 3.2 Procedure

### 3.2.1 Keyword analysis

A keyword analysis can function as a starting point for a literary interpretation of a text (Fischer-Starcke 2010, 65, Mahlberg 2010, 295, Mahlberg and McIntyre 2011, 207, Mahlberg 2012, 2). Therefore, the first step of analysis is set to be the keyword analysis. Sketch Engine (SkE) (Kilgariff 2009), a leading corpus linguistics online interface, is used to set each corpus as a reference corpus for the other in order to generate a keyword list for each corpus. The score for keyword extraction is set as 1, marked by SkE as “rare,” which is particularly useful for research. The attribute for keywords is lemmatized, which means that the tool will treat different forms of the same lemma as the same thing. It is worth mentioning that the keyness score in SkE is based on simple math method (Kilgariff 2009). Subsequently, the concordance tool is used to examine the extracted keywords in their linguistic environment. A categorization of keywords into their suitable groups is, then, provided as a means to identify the thematic interests of each corpus, and work toward constructing the fictional worlds of both novels (Mahlberg and McIntyre 2011).

The close examination of keywords in their linguistic environment is meant to group words with the same meanings together. This approach is akin to Sinclair's (2003, 2004), Fischer-Starcke's (2010), and Baker et al.'s (2013) approaches for thematization. In the process of classification, the groups are continually modified and adjusted to fit each keyword properly into a group. To achieve this purpose, one method is to use *ad hoc* categories particular to a specific text to group them based on their similarities. Another method is to fit the keywords into categories provided by a particular theory (Mahlberg and McIntyre 2011, 207). The current study follows the latter.

Following Mahlberg and McIntyre (2011, 209), keywords are grouped into two broad groups: “fictional world signals” and “thematic signals.” The identification of a keyword as belonging to a particular category is based on “the interpretative processes of individual readers.” These interpretative capacities are formed through reading experience and schematic knowledge. In this respect, different parameters need to be considered: the textual context of the keyword, how the reader perceives it, and how it contributes to building meaning in the whole novel (Mahlberg and McIntyre 2011, 210).

“Thematic signals” keywords are more abstract than “fictional world” keywords; therefore, they are prone to more interpretations. Words with concrete meanings tend to be identified as “fictional world” keywords since they are perceived as world-building elements. On the other hand, “thematic signal” keywords, besides their concrete sense, contribute to the thematic interests of the novel. The analysis of concordance lines in which keywords occur is an indispensable step in the process of identification. The identification of fictional world keywords depends mainly on recognizing their connection to characters, places, and objects. The thematic signal keywords, on the other hand, relate largely to evaluative, abstract, and metaphorical meanings, i.e., occasionally, their ambiguous status is a result of being polysemous. Therefore, the latter category tends to be reader-centered since it pertains to the effects of the text on the reader at a higher abstract level. In contrast, the former category can be perceived as text-centered since it relates to the textual and contextual milieu of the text (Mahlberg and McIntyre 2011, 212).

### 3.2.2 Key cluster analysis

The second step in the procedure of the current study is cluster analysis. Cluster analysis is based on a prime claim in corpus linguistics: “units of meaning are not equivalent to single words” (Mahlberg 2007a, 224, see also Stubbs 2007, Fischer-Starcke, 2010). Thus, words alone cannot be considered units of meaning in a text; instead, extended units of meaning are used for that purpose. Clusters or lexical bundles as a phraseological feature and the textual co-occurrence of words are ways to achieve the same end. In this respect, it is essential to refer to the “individual qualities” of each text. Leech and Short (2007, 60) stress that the prominent linguistic features in one text may not be equally prominent in another since “[t]here is

no infallible technique of selecting what is significant. We have to make ourselves newly aware, for each text, of the artistic effect of the whole, and the way linguistic details fit into this whole.”

SkE is used to extract key clusters of each corpus using *DS* as a reference corpus to *MSS* and *vice versa*. In this sense, the extracted clusters are not the result of isolated internal calculations of each corpus; rather, their identification is based on the comparative values, which render the clusters significant. The keyness score is set to 1 and the minimum frequency to 5. Previous studies on cluster analysis in different texts suggest that a valid cluster length normally ranges between three to five or six words (Mahlberg 2007a, 225, see also Biber et al. 1999). Thus, the range of clusters in the current study is set to four words. The extracted clusters are then closely investigated in their concordance lines and categorized into their fitting textual functions groups (Mahlberg 2007a).

Exploring concordance lines of key clusters does not only aid in highlighting particular local textual functions of each novel but also gives insights into characterization and interpersonal relations between different fictional characters. Thus, all occurrences of each key cluster are manually checked to decide on their relevance to character revelation and their tenor correlations. In this sense, key clusters are employed as role indicators of the interpersonal relations in each novel. It should be noted that concordance analysis predominantly focuses on label clusters and speech clusters since they are the two cluster types that yield illuminating potential in describing characters and interpersonal relations.

## 4 Data analysis

### 4.1 Keywords as topicalizers of thematic salience

Corpus linguistics is based on the assumption that the frequency of a pattern in a corpus is significant either for the content of the corpus or its schematic organization (Teubert 2005). The quantitative aspect of frequency then affects the qualitative analysis of a corpus (Fischer-Starcke 2010, 195). In Section 4, the assumption mentioned above is put to practice as the two corpora under scrutiny are analyzed in terms of

**Table 1:** Top 20 keywords in *MSS* (characters' names excluded)

| Rank | Keyword       | Frequency | Keyness score |
|------|---------------|-----------|---------------|
| 1    | Prison        | 87        | 886.84        |
| 2    | Cottage       | 34        | 347.19        |
| 3    | Cinema        | 28        | 286.10        |
| 4    | Kid           | 22        | 225.01        |
| 5    | Ghetto        | 22        | 225.01        |
| 6    | Township      | 20        | 204.64        |
| 7    | Leadership    | 19        | 194.46        |
| 8    | Government    | 19        | 194.46        |
| 9    | Organization  | 18        | 184.28        |
| 10   | Benoni        | 17        | 174.09        |
| 11   | Meaning       | 16        | 163.91        |
| 12   | Detention     | 16        | 163.91        |
| 13   | Revolutionary | 15        | 153.73        |
| 14   | Principal     | 15        | 153.73        |
| 15   | Liberation    | 15        | 153.73        |
| 16   | Carryall      | 15        | 153.73        |
| 17   | Gathering     | 14        | 143.55        |
| 18   | Security      | 14        | 143.55        |
| 19   | Executive     | 13        | 133.37        |
| 20   | Discussion    | 13        | 133.37        |

keywords and key clusters. Table 1 displays the keywords of *MSS* with their rank, frequency, and keyness score as provided by SkE. Table 2 offers the same type of information for *DS*.

Fischer-Starcke (2010, 91) makes it explicit that the analysis of keywords' concordance lines (1) leads to the identification of significant topics in the text and (2) gives insight into the characterization of protagonists. To work toward a categorization of keywords, concordance lines of each keyword are examined using SkE concordance tool, which aids in analyzing text-specific meanings throughout each text. In this phase, contextual and functional aspects are considered instead of the quantitative aspect of word frequencies (Mahlberg 2014, 283–284).

The process of classification is by no means simple or straightforward. First of all, in *MSS*, some keywords which could possibly be intuitively classified as belonging to the “fictional world” category, e.g., *cottage* and *cinema* due to their reference to concrete places, are not actually so. Concordance line inspection reveals that they are perspicuously associated with the love affair, i.e., in 20 out of 34 occurrences *cottage* is directly correlated with Sonny and Hannah's illicit love, e.g., *they had Hannah's cottage, Hannah's bed, which was unlike any other bed Sonny had ever known, not only because of what passed between her and him there but because it was not a bed at all*. In the rest of the cases, their intimate relationship is the background of fictional incidents, as when Sonny is given the key to the cottage but laments the fact that he cannot enter it, e.g., *How could he account for himself, approaching this cottage, key in his pocket. He had better go away* (Figure 1). *Cinema* occurs 28 times; 17 out of which are explicitly related to Sonny and Hannah's love affair as well, since it is through the cinema that Will discovers their intimacy, e.g., *he did not need to explain anything to me, since the cinema, when he'd told me what to see and made clear what I was not to have seen*.

In the same way, *kid* seems at first glance to be an outsider of the “places” group. However, concordance analysis proves it fits in the aforementioned group since, in most of its occurrences, it refers to the kids of the ghetto, e.g., *in the ghetto where our kids are brought up*. Therefore, its collocation with *ghetto* guarantees its position in the “places” group. *Detention*, *liberation*, and *revolutionary* mark another thematic interest of *MSS*, that is, racial discrimination and the struggle against political injustice. Though the fictional world sub-category “political activity” subsumes abstract keywords, such as *leadership*,

**Table 2:** Top 20 keywords in *DS* (characters' names excluded)

| Rank | Keyword     | Frequency | Keyness score |
|------|-------------|-----------|---------------|
| 1    | Byron       | 71        | 868.33        |
| 2    | Teresa      | 46        | 562.94        |
| 3    | Professor   | 41        | 501.86        |
| 4    | Mr          | 25        | 306.40        |
| 5    | Clinic      | 22        | 269.75        |
| 6    | Ms          | 21        | 257.54        |
| 7    | Rape        | 17        | 208.67        |
| 8    | Wordsworth  | 17        | 208.67        |
| 9    | Goat        | 16        | 196.46        |
| 10   | Market      | 13        | 159.81        |
| 11   | Guess       | 13        | 159.81        |
| 12   | Kombi       | 12        | 147.59        |
| 13   | Grahamstown | 12        | 147.59        |
| 14   | Solid       | 11        | 135.38        |
| 15   | Bitch       | 11        | 135.38        |
| 16   | George      | 11        | 135.38        |
| 17   | Poet        | 10        | 123.16        |
| 18   | Flame       | 10        | 123.16        |
| 19   | Banjo       | 10        | 123.16        |
| 20   | Monday      | 10        | 123.16        |

is and forests were only post-coital reveries, they had Hannah's **cottage**, Hannah's bed, which was unlike any other bed Sonny had ever themselves off; to his son. </s><s> The instinct was to go to her **cottage** and hide. </s><s> Sonny was silent, his eyes concealed from Hannah began to make more little excursions outside her bed and the **cottage**. </s><s> She had a friend who was an absentee farmer. </s><s> a big house but that's not where he goes. </s><s> She lives in a **cottage** behind trees at the end of the garden. </s><s> Maybe there's even a table. </s><s> But this isn't like a house at all; well, all right, a **cottage**, but not even any kind of place where you'd expect a white woman to be on the waking level of speech. </s><s> When he arrived at her **cottage** his mood transformed: the loss of his daughter -- her totally unexpected question was of alliances; he and Hannah were sitting outside her **cottage** in the garden, which they had to themselves because the people were. </s><s> He pulled himself out of the chair and went into the **cottage**; to that one room. </s><s> Now there are things he doesn't know as she had gone because her grandfather had died. </s><s> The **cottage** was locked up. </s><s> She'd left him the key. </s><s> Of course it's a sleeper -- some meths drinker must have found the deserted **cottage** convenient to camp against -- then, seeing that the sleeping bag was there. </s><s> How could he account for himself, approaching this **cottage**, key in his pocket. </s><s> He had better go away. </s><s> Cornered where this intruder was waiting -- waiting for him, clearly -- her **cottage**, their room, to which he would return again and again, unable to be aware, without seeing, when she took him into her body in this **cottage**. </s><s> He, like all his kind, educated in political struggle, knew that it was good in everything'. -- The young man was living in the **cottage** while he was back at the house with his son, Will. </s><s> It was Hannah asked of Sonny, her message: let this person stay in the **cottage**, give him the key. </s><s> The key? </s><s> He stood in a hard

Figure 1: The keyword *cottage* in concordance lines.

*organization*, *gathering*, *discussion*, and *meaning*, they together are part of setting the scene of the novel and participate in constructing the fictional world of the novel.

In exactly 50% of the occurrences of the keyword *meaning*, it refers to the significance of political struggle but could also possibly be an example of a “familial relations” sub-category since in the other 50% it relates to Sonny’s relation to Aila, Hannah, and Will. The same applies to *discussion*, which refers in most of its concordance lines to the endeavors made on the part of civil organizations to defend Sonny in his trial. The rest of the cases are dedicated to familial relations and the love affair. In some cases, the struggle and the illicit love are inextricable, as in *they turned from caresses to worried discussion about the action of police agent provocateurs among church groups in the rent boycotts*. *Security*, though an abstract word, is almost entirely part of the collocational group *Security Police* which makes it most likely to be taken as constructing the fictional world. Another interesting keyword is *detention* which fits into the fictional world category as a background of the political atmosphere of the novel, but also fits in the thematic signal category as it refers to apartheid and discrimination as dominant themes of *MSS*.

It should be noted that characters’ names can readily be classified as thematic signals as well. For example, *Aila*, *Sonny*, and *Hannah* are the three ribs of a love triangle and so the three keywords notably refer to the theme of illicit love. The peculiar absence of Will’s name as a keyword can also be accounted for through concordance lines. A substantial part of the novel is narrated by Will, the teen-aged son of Sonny. Hence, Will’s name is substituted by the first-person pronoun “I.”

In *DS*, the classification process is equally intriguing. Keywords which on surface value can be taken as characters’ names are proved by concordance line examination to be names of literary figures repeatedly evoked by Professor David Lurie, who teaches Romantic Poetry, e.g., Lord *Byron* (71 occurrences), and William *Wordsworth* (71 occurrences as well). Moreover, *Teresa*, which occurs 46 times, refers to Lord Byron’s beloved, whom David takes special interest in and fantasizes about excessively: “*Teresa is past honour. She pushes out her breasts to the sun; she plays the banjo in front of the servants and does not care if they smirk.*” The “literature” sub-category builds the fictional structure of the academic atmosphere, which eventually condemns the protagonist for his unethical practices. Keywords which appear in the form of verbs and adjectives are particularly problematic when it comes to thematic classification. For example, *guess* and *solid* are used in the context of various topics and instances, and are, accordingly, difficult to categorize. The same applies to the time-referring noun “Monday.”

*Bitch* and *goat* are repeatedly used to portray the type of life on the farm and work duties in the clinic, respectively. *Flame* has a double role which guarantees its position in two categories. First, it is mentioned

in relation to David's sexuality with Melanie and equally in reference to being fired up by Lucy's rapists. Second, it is collocationally used in respect to the dog's incineration. The first two instances directly relate to two thematic foci in the novel: David's immoral sexuality and sexual violence against women. In the latter, it aids building the fictional world of the novel. The structure of the farm life is maintained by *kombi* in which Lucy's products are loaded, while *banjo* relates to David's whimsical imagination about Teresa's music as well as Lucy's old musical instrument, which instigates his fantasies.

Both novels share the "place" sub-category. In *MSS*, the sub-category encompasses *ghetto*, *township*, and *Benoni*, where Sonny and his family live, besides *prison* where he spends 2-year sentence on political charges. In *DS*, the "place" sub-category includes *George*, hometown of Melanie Isaacs (the student with whom David is involved), *Grahamstown*, the town where Lucy lives, *market*, where Lucy's stall is, and *clinic* where David works after his dismissal. Perspicuously, the "place" sub-category contributes toward constructing the literary whereabouts of the two novels.

"Profession" is also a common sub-category for both novels. In *MSS*, *executive* is mainly associated with governmental officials, while *principal* refers to the headmaster of the school where Sonny used to work. In *DS*, *Professor* and *Mr* are essentially references to David, while *Ms* is a reference to Melanie (Figure 2). Interestingly, Sonny, a black schoolteacher and a political leader, is never given a last name or addressed by a title, while David, a white professor, is repeatedly referred to by his last name and with two titles. The "thematic signals" group includes *rape* and *flame*. The former refers to the main theme in *DS* – Lucy's rape. The latter, as explained earlier, metaphorically relates to David's sexual desire toward Melanie and the actual event of being burned by the rapists. This is in accordance with literary critics who argue that these particular two thematic foci are most significant of *DS* (Attridge 2002, 315).

The corpus-driven categorization provided in this section is verifiably capable of sketching the thematic interests of each corpus as identified by literary critics. For example, Golden (2019) argues that *MSS* revolves around racial violence. Sakamoto (2002) discusses political struggle and racial identity on the one hand, and sexual commitment on the other as two important pivotal axes of the novel. Möller (1999) investigates the ethical, the political, and the multiracial strife as essential components of the novel. Also, in agreement with the current corpus-driven analysis, *DS* is described by literary critics as "a story of a rape" (Stott 2009), and thus the theme of rape is paramount (Attridge 2002, Graham 2003, Mardorossian 2011). Coleman (2009) and Oriaku (2016) contend that "sexual transgression" is a focus point. Desire (Malloy 2004) and scandal (Easton 2007) are also recognized as important foci of *DS*. Animals, especially dogs, and their euthanasia are equally critically scrutinized in *DS* (Sewall 2013, Marais 2001). Language itself is identified as an important focal point in the novel, i.e., as a university professor of literature, David uses the topic of language repeatedly throughout the novel to offer his thoughts about imagination, muses, literature, etc. (Attridge 2002, 319). In this sense, Attridge's remark about language is congruent with the current study classification of the keyword *literature*. Furthermore, Marais (2006) investigates imagination as a vital component of the novel. Interestingly, it is argued that the political theme of the novel is impeded into the names of characters such as Pertus (Swales 2003).

of him, teaches in the Business School. 'The body here gathered, **Professor** Lurie,' says Mathabane, opening proceedings, 'has no powers. All  
which you all have copies. Do I need to summarize that statement?' **Professor** Lurie?' 'Do I understand, Mr Chairman, that Ms Isaacs will not be a  
rans across to Mathabane. Murmured words pass between them.' **Professor** Lurie,' says Hakim, 'I must repeat, this is a committee of inquiry. Its  
'what?' 'Of all that I am charged with.' 'You are taking us in circles, **Professor** Lurie.' 'Of everything Ms Isaacs avers, and of keeping false record:  
a Rassool intervenes. 'You say you accept Ms Isaacs's statement, **Professor** Lurie, but have you actually read it?' I do riot wish to read Ms Isaac  
nt.' Farodia Rassool sits back in her seat. 'This is all very quixotic, **Professor** Lurie, but can you afford it? It seems to me we may have a duty to  
proposed', says Mathabane, 'that the committee recess to discuss **Professor** Lurie's plea.' A round of nods. 'Professor Lurie, could I ask you to s  
mittee recess to discuss Professor Lurie's plea.' A round of nods.' **Professor** Lurie, could I ask you to step outside for a few minutes, you and M  
not good: sour, it seems to him. 'So,' says Mathabane, 'to resume: **Professor** Lurie, you say you accept the truth of the charges brought against  
to say?' 'Yes. I want to register an objection to these responses of **Professor** Lurie's, which I regard as fundamentally evasive. Professor Lurie s

Figure 2: The keyword *professor* in concordance lines.

## 4.2 Key clusters as pointers to local textual functions

Often the use of one analytic technique is not sufficient for interpreting literary meaning in a text; rather, the analyst needs a combination of techniques for this purpose (Fischer-Starcke 2010, 106). This section provides another method for approaching the two novels, that is, clusters. The present section aims to use key clusters as signals to local textual functions, which are probed for implications for characterization and interpersonal relationships. It is important to stress that the extraction of clusters alone cannot provide a proper stylistic analysis; rather, it is regarded as an initial step toward interpreting literary meanings and recognizing local textual functions (Mahlberg 2007a, 225). The notion of local textual functions is “relational” in the sense that the local perspective views one text with regard to other similar texts. Therefore, the analysis of clusters in *MSS* is most useful when it is related to observations in *DS* and *vice versa*. The local point of view also affects the type of clusters, i.e., the length of repeated sequences taken into account.

As for the two novels at hand, the close examination of each cluster in its concordance lines results in the classification of functional groups provided in Tables 3 and 4. The clusters in *MSS* fall into four categories: labels, speech clusters, *as if* clusters, and time and place clusters, whereas those of *DS* also fall into four categories: labels, speech clusters, body part clusters, and time and place clusters. The category “Other” is dedicated to clusters that do not fit in any of the specified categories. The groups identified contribute toward a description of the local textual functions specific to each novel.

In *MSS*, label and speech clusters are particularly significant for characterization and sketching interpersonal relations. As for the label group, *was the one who*, for example, in 50% of its occurrences relates to Will as in *I was the one who opened the door to her jailers* and *I was the one who opened the door to them* as Will holds himself responsible for the arrest of his mother. The cluster contributes to the description of Will’s character and his feel of guilt that he opened the door for the Security Police who took Aila away and, thus, functions as a building block in the fictional world of the novel. *The father and son* is particularly an interesting label cluster since, in all of its occurrences, it refers to Sonny and Will in their struggle out of the legal labyrinth in Aila’s trial, e.g., *It was a long wait and at first the father and son walked again and again along the four sides of the verandah, as people do while expecting to be summoned any minute*. The repeated use of *father and son* as composed of two family-related nouns linked by the coordinating conjunction *and* plays a vital role in delineating the interpersonal relation between Sonny and Will. This key cluster has the local textual function of presenting Sonny and Will allied, in spite of their earlier discrepancies, in the face of Aila’s legal ordeal. It can be argued that *father and son* acts as a building block that marks a different phase in the father–son relationship in the *MSS*.

**Table 3:** Comparative classification of fictional world and thematic signal keywords in *MSS* and *DS* with examples

| Category            | Example keywords                                                   |                                                               |
|---------------------|--------------------------------------------------------------------|---------------------------------------------------------------|
|                     | MSS                                                                | DS                                                            |
| 1. Fictional world  |                                                                    |                                                               |
| Characters          | Aila, Sonny, Hannah, Baby                                          | Lucy, Petrus, Bev, Shaw, Melanie, Lurie, Isaacs, Soraya, etc. |
| Places              | Prison, kid, ghetto, township, Benoni                              | Clinic, market, George, Grahamstown                           |
| Profession          | EXECUTIVE, principal                                               | Professor, Mr, Ms                                             |
| Political activity  | Leadership, organization, gathering, discussion, meaning, carryall | —                                                             |
| Authorities         | Government, detention, security                                    | —                                                             |
| Literature          | —                                                                  | Byron, Teresa, Wordsworth, poet                               |
| Animals             | —                                                                  | Goat, bitch, flame                                            |
| Tools               | —                                                                  | Kombi, banjo                                                  |
| 2. Thematic signals | Detention, cottage, cinema, liberation, revolutionary              | Rape, flame                                                   |

Table 4: Four-word clusters in MSS

| Rank | Clusters                     | Frequency | Keyness score | Function group |
|------|------------------------------|-----------|---------------|----------------|
| 1    | As if he were                | 9         | 92.6          | AI             |
| 2    | Didn't want to               | 8         | 82.5          | O              |
| 3    | To be in the                 | 7         | 72.3          | TP             |
| 4    | On the other side            | 7         | 72.3          | TP             |
| 5    | Was the one who              | 6         | 62.1          | L              |
| 6    | In one of the                | 6         | 62.1          | TP             |
| 7    | He was going to              | 6         | 62.1          | L              |
| 8    | Don't know why               | 6         | 62.1          | S              |
| 9    | At the same time             | 6         | 62.1          | TP             |
| 10   | To get rid of                | 5         | 51.9          | O              |
| 11   | They were going to           | 5         | 51.9          | O              |
| 12   | The other side of            | 5         | 51.9          | TP             |
| 13   | The father and son           | 5         | 51.9          | L              |
| 14   | Out of the house             | 5         | 51.9          | TP             |
| 15   | One of their own             | 5         | 51.9          | L              |
| 16   | N't know what to             | 5         | 51.9          | S              |
| 17   | It was as if                 | 5         | 51.9          | AI             |
| 18   | Didn't have to               | 5         | 51.9          | O              |
| 19   | Close to the earth           | 5         | 51.9          | TP             |
| 20   | Between Aila and him         | 5         | 51.9          | L              |
| 21   | It was as if                 | 5         | 51.9          | AI             |
| 22   | I didn't want                | 5         | 51.9          | S              |
| 23   | High commission for refugees | 5         | 51.9          | L              |
| 24   | For the first time           | 7         | 2.8           | TP             |
| 25   | To be able to                | 5         | 2.0           | O              |
| 26   | Don't know what              | 18        | 1.9           | S              |
| 27   | I didn't know                | 7         | 1.4           | S              |
| 28   | I don't know                 | 38        | 1.3           | S              |
| 29   | I don't want                 | 13        | 1.2           | S              |
| 30   | In the middle of             | 5         | 1             | TP             |

L = labels, S = speech, AI = as if, TP = time and place.

The cluster *One of their own* is exclusively used to refer to black individuals as in *No wonder parents wanted removed from the school a teacher, **one of their own** kind, who led their children over there* and, thus, can be classified as specifically labelling black members of society. The cluster also characterizes a particular local textual function, portraying the black community as one whole, which encourages the inclusion of all black members. *Between Aila and him* in all its occurrences refers to Sonny and Aila. The use of the preposition *between*, which indicates separation, is mainly used to characterize mental and spiritual separation between the couple, e.g., *There's no freedom in working for freedom. – He could say it to Hannah, and they both laughed. There was pride and scepticism in the laughter. You couldn't say such things to Aila; **between Aila and him** was the old habit of simple reverence for living useful lives.* The local textual function represented by the aforementioned key cluster relates to the interpersonal relation between Sonny and Aila, who are divided by Sonny's involvement in both political struggle and liaison with Hannah.

The speech cluster group both helps in character description, on the one hand, and contributes toward a better understanding of textual meaning, on the other. For example, all cases of the cluster *don't know why* are preceded by the first-person pronoun *I* as in *I couldn't resist the compulsion. **I don't know why**; I went again to the cottage. What did I think I was going to see there?* In 16 instances out of 18, *don't know what* is predicated on first or second-person pronouns as in *my father isn't the man to be scared off his political work because he's been jailed for it. Or he wasn't the man; now **I don't know what** he is.* It is likely then to categorize *don't know what* as a speech cluster. The same applies to *n't know what to*. Most of the cases of *I*

*don't know* and *I don't know what* are ascribed to Will. It is noteworthy that the two clusters are recognized as particularly frequent in current contemporary fiction as a whole (Stubbs and Barth 2003).

In four out of five cases of *I didn't want*, the referent of *I* is Will. The cluster has a particular local textual function since most of its cases are dedicated to one situation where Will is wary that his mother does not notice Sonny's affair, e.g., ***I didn't want to be in the room alone with her, either. But if I kept out of her way she would know there was something wrong, thinking in her innocence this would be something concerning me.*** As a building block, the aforementioned key cluster delineates Will as heedful to his mother's feelings. The cluster *I didn't know* is also mostly reserved for Will. In most of the cases, it functions textually as an expression of Will's bewilderment about the liaison, e.g., *While he spoke to me he drew back as if I might smell her on him. I didn't know.* Both *I didn't want* and *I didn't know* contribute to sketching the interpersonal relations between Will and Sonny, on the one hand, and Will and Aila, on the other.

The *as if* cluster, including 19 cases, occurs exclusively in MSS. Generally, this type of clusters allows for the creation of a textual world by comparison. *As if he were*, for example, is mostly used by Will to create a world in which Sonny is a different person, e.g., *He was there in his usual place, as if he were my father again, not the man with his blonde woman in the foyer of the cinema.* The interpersonal father-son relation is also revealed, i.e., the crisis created by disclosing the father's love affair and its repercussions led to Will's denial of Sonny's fatherhood. In the concordance, *I felt strange that she wouldn't look at me. It was as if she wanted to deny this new bond of intimacy between us*, a textual world where Aila's love for her son is concealed is also created.

In DS, the classification of clusters is particularly challenging. The cluster *he doesn't know*, for example, is marked as a labelling cluster since in seven out of eight occurrences, it collocates with David's life crisis and bewilderment about himself as in *There are days when he does not know what to do with himself.* Accordingly, the key cluster perspicuously acts as a local textual function that portrays David as lost and hopeless. This is in accordance with Rinzler (2013) who sees David as an anti-hero "struggling with life." In nearly half of its cases, *the two of them* refers to David and Lucy, indicating their estrangement feeling, e.g., *Is she calling Johannesburg, speaking to Helen? Is his presence here keeping the two of them apart?* The cluster also refers to David with other characters: David and Soraya, one of the rapists, and Pollux. The cluster, then, entirely illuminates David's interpersonal relationships with other personae in the novel (Table 5).

The cluster *he would like to* is almost entirely associated with David, which justifies its classification as a labelling cluster. Most conspicuously, the local textual function of most of its occurrences relates to David's abhorrence and resentment of Petrus and the dissociation between what "he would like" to do and what is actually done, since he is eventually incapable of achieving his wishes:

Yet at this moment **he would like to** take Petrus by the throat. If it had been your wife instead of my daughter, **he would like to** say to Petrus, you would not be tapping your pipe and weighing your words so judiciously. Violation: that is the word **he would like to** force out of Petrus. Yes, it was a violation, **he would like to** hear Petrus say; yes, it was an outrage.

In 50% of its instances, *is the one who* is dedicated to David who takes the lead in his immoral involvement with Melanie, who takes care of Lucy after the rape, and who is responsible for incinerating the animals after helping Bev Shaw putting them to death, e.g., *To the extent that they are together, if they are together, he is the one who leads, she the one who follows.* Thus, the cluster highlights the main three stages in the development of David's character. In parallel, the other 50% of cases are dedicated to Melanie who feels embarrassed while asking to sleep in David's apartment, to Petrus who takes care of the farm after the rape, and to Bev Shaw who is in charge of injecting the animals.

The cluster *there is no sign* is particularly interesting. It is categorized as a labelling cluster since in four out of five occurrences, it relates to Petrus, e.g., *Of Petrus there is no sign and there is no sign he has even heard.* In this case, it contributes to molding his character in two directions: first, Petrus disappears in situations where he is supposed to be of help to Lucy. Suspiciously, he also disappears just before the rape, which sheds doubts on both the sincerity of his intentions to help and his involvement in the rape crime itself. Second, he ignores David's comments about Lucy's being unsafe on the farm after the attack, and

Table 5: Four-word clusters in *DS*

| Rank | Clusters             | Frequency | Keyness score | Function group |
|------|----------------------|-----------|---------------|----------------|
| 1    | I am going to        | 10        | 123.2         | S              |
| 2    | He doesn't know      | 8         | 98.7          | L              |
| 3    | The two of them      | 7         | 86.5          | L              |
| 4    | He would like to     | 7         | 86.5          | L              |
| 5    | Is the one who       | 6         | 74.3          | L              |
| 6    | In the old days      | 6         | 74.3          | TP             |
| 7    | You are going to     | 5         | 62.1          | S              |
| 8    | This is not the      | 5         | 62.1          | L              |
| 9    | There is no sign     | 5         | 62.1          | L              |
| 10   | That is what they    | 5         | 62.1          | O              |
| 11   | That he does not     | 5         | 62.1          | L              |
| 12   | On the point of      | 5         | 62.1          | TP             |
| 13   | He thinks to himself | 5         | 62.1          | L              |
| 14   | Does not know what   | 5         | 62.1          | L              |
| 15   | Are not going to     | 5         | 62.1          | L              |
| 16   | There is a long      | 5         | 62.1          | L              |
| 17   | It is not a          | 5         | 62.1          | O              |
| 18   | I won't be           | 5         | 62.1          | S              |
| 19   | I am prepared to     | 5         | 62.1          | S              |
| 20   | He shakes his head   | 5         | 62.1          | BP             |
| 21   | The rest of the      | 6         | 3.5           | TP             |
| 22   | I don't have         | 6         | 3.5           | S              |
| 23   | I don't think        | 8         | 1.9           | S              |
| 24   | Don't want to        | 10        | 1.2           | S              |

L = labels, S = speech, BP = body parts, TP = time and place, O = other.

even refuses to help David in bringing the boy who is suspected to be part of the attack into justice. The key cluster, then, is a local textual function referring to Petrus' character as a deceitful and manipulative man.

*He thinks to himself* is entirely related to David; this association warrants its classification as a labelling cluster, e.g., *She opens the third cage and releases the two Dobermanns into it. A brave gesture, he thinks to himself; but is it wise?* Instances of this key cluster suggest that David internalizes his thoughts rather than spelling them out even in potentially hazardous situations like the one that led to Lucy's rape. Likewise, the cluster *does not know what* in most of its occurrences relates to David and his occasionally senseless life, and his denouncement of Lucy's detachment after the rape, e.g., *Do they think he does not know what rape is? Do they think he has not suffered with his daughter?*

In nearly half of its instances, *are not going to* refers to Bill and Bev Shaw as Lucy defends her choice of befriending whom she likes even if it goes contrary to her father's preferences, e.g., *You don't approve of friends like Bev and Bill Shaw because they are not going to lead me to a higher life.* Therefore, its local textual function represents Lucy's persistence on preserving her free will. The other half refers to the police inefficiency, e.g., *The best is, you save yourself, because the police are not going to save you, not any more, you can be sure.* The cluster *there is a long* is particularly significant since in four out of five instances, it collocates with *silence* as in *They will not be able to deny that. There is a long silence between them. Do you love him yet?* and *They are out again on bail. Anyway, it's not my car, so whoever was arrested can't be whoever took my car. There is a long silence. Does that follow, logically? she says.* The long silence between David and other fictional personae marks the distance and sometimes the tension between him and other characters, mainly Lucy.

Speech clusters are specifically interesting when probing interpersonal relations. For example, in two concordances *I am going to call the police* is uttered by David when discovering that one of the attackers is working for Petrus. The strained relation between David and Petrus is further problematized. The same cluster is used in David's conversations with Rosalind, Bev Shaw, Lucy, Melanie, and even himself. The

cluster *you are going to* again refers to the tension relation between David and Rosalind as she tells him when discussing his affair with Melanie: **You are going to end up as one of those sad old men who poke around in rubbish bins.** David's denouncing question *Are you telling me you are going to have the child?* again exposes the conflicting nature of his relation to Lucy. Significantly, the powerful/powerless relationship between David and Melanie is manifested in three consecutive instances of the current cluster: **you are going to have to give more time to your work. You are going to have to attend class more regularly. And you are going to have to make up the test you missed.** By dictating Melanie what she is going to do, David continues his immoral befuddlement and disempowerment of a student of his.

Three out of five occurrences of the cluster *I won't be* are dedicated to David. Initially, **I won't be bored** is to dismiss Lucy's concern about his potential long stay in the farm. Conversely, in the remaining two cases he refers to his decision not to move in with Lucy, **I won't be moving in with Lucy**, marking their disagreement. The cluster *I am prepared to* is particularly interesting in revealing the interpersonal relation between David and Lucy. In two instances, David is addressing Lucy to offer her fatherly help: **I am prepared to send you to Holland** and **I am prepared to give you whatever you need to set yourself up again somewhere safer than here.** Ironically, the same cluster is used by Lucy in the same encounter to drive him away from her life: **I am prepared to do anything, make any sacrifice, for the sake of peace.** In the remaining two cases, David uses the cluster in his inquest session at the university to plead guilty: *I plead guilty. That is as far as I am prepared to go.* The key cluster *I don't have* is mostly used by David, i.e., it sheds light on his relationship with Soraya whom he desperately strives to stay in contact with but faced with her refusal: **I don't have a number.** It also highlights his relationship with Bev Shaw whom he makes fun of: *Provided that I don't have to call her Bev. It's a silly name to go by. It reminds me of cattle.* His attitude toward Bev is argued to reflect his misogynistic nature (Sewlall 2013, 82).

## 5 Findings and discussion

The corpus stylistic analysis provided by the current study is based on: (1) utilizing a keyword analysis to work toward a comparative thematic categorization of the two novels at hand, and (2) using key cluster analysis to reveal aspects of characterization and characters' interpersonal relationships. First, considering the thematic focal points of each novel, the analysis shows that *MSS* presents a turbulent South Africa encumbered with political violence in the apartheid time, while *DS*, though the apartheid is allegedly over, represents racialized sexual violence manifested in Lucy's gang rape which is also considered "a political act" (Oriaku 2016). The immoral in *MSS* is manifested in the illicit love between Sonny and Hannah. In *DS*, it covers David's sexual fornication and depravities, and Lucy's rape. The rape crime suggests an intersection of gender dimension with racial considerations. As Lucy explains: *"I think they are rapists first and foremost. ... I think they do rape."*

The thematic classification of keywords in *MSS* captures the two axes: the racial, as in *detention*, *liberation*, and *revolutionary*, and the immoral, as in *cottage* and *cinema*. In *DS*, the immoral is manifested in *rape* and *flame*, while the racial is implicit. However, the racial can be discerned in characters' names, e.g., *Petrus*, which appears as a keyword as well. Thus, the keyword-based categorization reveals that *MSS* is equally about racial discrimination and political struggle as much as it is about the immoral. In *DS*, on the other hand, the racial is implied and most indirect that it is not captured by thematic signal keywords. Only through a meticulous concordance analysis can the racial be recognized. Hence, the keyword-based classification proves congruent with the thematic foci provided by literary critics. This corpus-driven categorization of keywords which aims at comparatively marking thematic interests is a step further in validating the use of corpus tools in the analysis of literary texts.

Second, key clusters as indicators of local textual functions is a tool of describing lexical items in terms of the functions they achieve in texts. In this respect, local textual functions proved also most beneficial in discerning aspects of characterization and interpersonal relationships in both novels. Label and speech clusters are equally particularly revealing in *MSS* and *DS*. Just to mention few examples, in *MSS* the label

key cluster *was the one who* marks an important aspect of Will's character since it functions locally and textually as an indicator of his guilt feeling toward Aila's imprisonment. The key cluster *the father and son* characterizes a phase of alliance in Sonny and Will's fluctuating relationship. The mental and spiritual detachment between Aila and Sonny is portrayed through the local textual function of *between Aila and him*. The speech key clusters *I don't know why*, *I don't know what*, *I didn't want*, and *I didn't know* are all revealing of different aspects in Will's character as well as his relationship with his mother, father, and even Hannah. The *as if* group is particular to *MSS*. Interpersonally, the key clusters *as if he were* and *it was as if* refer to two constructed worlds in which Sonny is a different father and husband, and in which Aila's love for her son is vanished.

In *DS*, the key cluster *he does not know* has a local textual function that suggests a salient aspect of David's character – his perturbation and perplexity about his behavior. Different aspects of David's character are mediated through various key clusters, e.g., *that he does not*, *that he does not know*, *he thinks to himself*, *this is not the*, etc. Local textual functions operationalized through the key cluster *there is no sign* have also a noticeable role in depicting Petrus' character as a disingenuous, mendacious, and unscrupulous man. The key cluster *there's a long* characterizes David's estrangement with Lucy and unfamiliarity with Bev. *The two of them* also marks his tensioned relationship with different characters, e.g., Lucy, Soraya, the rapist, etc. Speech key clusters in *DS* are specially enriching in elucidating interpersonal relationships between characters: his strained relationship with Lucy in *I am going to*, *you are going to*, *I won't be*, and *I am prepared to*; his immoral powerful/powerless relationship with Melanie in *you are going to*; his hostility with Rosalind in *I am going to* and *you are going to*; his disdain of Bev in *I don't have*, and his implicit feud with Petrus in *I am going to*.

The analysis enabled through corpus tools has variant implications on literary meaning. The thematic analysis, for example, reveals that both protagonists are portrayed through their love of literature: Sonny of Shakespeare and David of Byron. Interpersonally, for different reasons, both have unsettled relationships with their children. Key cluster analysis, however, sets the two personae apart in their immorality: David is "a serial womanizer" (Sewlall 2013) while Sonny has an affair with Hannah. From a socio-political point of view, it is clear that the two novels are set on a temporal continuum: *MSS* overtly tackles the racial and political struggles in South Africa during the apartheid while sexual illicit love between races is highlighted. *DS*, though the apartheid "is" over, is *still* surprisingly stranded in manifestations of racism and even counter-racism practiced through sexual encroachment between races.

It is worth mentioning here that language patterns that construe tenor-related analysis belong either to speakers, that is, characters contributing direct speech exchanges (characters are typically rigorously delineated by novelists to achieve the overall aesthetic as well as ideological goal of the novel), or to the narrative voice which varies according to the stylistic schematic effect aimed at by the novelist (Matthiessen 1993). From a stylonarrative point of view, according to key cluster analysis, the narrator in *MSS* is revealed as Will. Hence, most of the fictional events are mediated through Will's personal lens. *DS*, on the other hand, is marked by impersonality or neutral style since the narrator is a neutral omniscient voice (Rinzler 2013).

## 6 Conclusion

Corpus stylistic analysis is quite an objective methodological procedure led by a quite subjective process of interpretation, i.e., corpus stylistics relies naturally on quantitative measures and statistical significance; however, its application necessitates taking qualitative decisions and providing interpretive acts in view of corpus data and occasionally ahead of it (Carter 2010, 67; see also Myrdal 1970, Baker 2006, Baker et al. 2013). The present study shows that the use of corpus linguistics in the analysis of literary texts can help to comparatively reveal facets of literary meanings, themes, and characterization in a systematic way by offering an objective tool of identifying keywords and key clusters and locating instances of their use in a literary text. The processes of categorization enabled through corpus tools provide powerful means for a

more detailed analysis of common and different literary meanings as well as linguistic features. Label clusters are particularly illuminating in uncovering different aspects of characterization, while speech clusters are specifically enriching in the examination of interpersonal meanings as an indicator of the relationships between characters.

The present analysis attempted to demonstrate how corpus tools can be used to identify lexical items and related co-text to set for further qualitative analysis in comparative terms. In this regard, it posed three study questions concerning the thematic focal points, characterization, and interpersonal relations in the two novels at hand. To answer the first question, the thematic foci of the two novels were comparatively schematized by dint of keywords extracted through corpus techniques. *MSS* is discerned to revolve around South Africa during apartheid, the struggle against the white regime, and unlawful love. *DS* centers around the violence practiced by both “white” and “black” men, mostly against women, and immoral sexuality of men. In both novels, gender-related themes conspicuously permeate through a well-woven fabric of the socio-political reality of South Africa at the peak of a “racial” and “social” discrimination era in *MSS*, and the undertones of such a hideous time, when the apartheid is over, in *DS*. Therefore, the intersectionality of race and gender is evident in both novels, though to varying degrees. To answer the second and third questions, key clusters in both novels were extracted, classified, and analyzed to identify local textual functions which yielded insights into characterization and interpersonal relations in both novels. The two protagonists are depicted as literature lovers. Both have tensioned relationships with their children and with other female protagonists in their lives. Both indulge in immoral sexual practices, albeit to varying degrees. Although *MSS* is set during the apartheid and *DS* after the apartheid came to an end, it is rather striking that the South Africa drawn in *DS*, with the racism, counter-racism, and cross-race sexual violence, is far from the one promised in *MSS*.

**Conflict of interest:** Author states no conflict of interest.

## References

- Adolphs, Svenja and Ron Carter. 2002. “Corpus stylistics: Point of view and semantic prosodies in *To The Lighthouse*.” *Poetica* 58, 7–20.
- Attridge, Derek. 2002. “J. M. Coetzee’s disgrace: Introduction.” *Interventions* 4(3), 315–20. doi: 10.1080/1369801022000013752.
- Baker, Paul. 2006. *Using corpora in discourse analysis*. London: Continuum.
- Baker, Paul. 2014. *Using corpora to analyze gender*. London: Bloomsbury.
- Baker, Paul, Costas Gabrielatos and Tony McEnery. 2013. “Sketching muslims: A Corpus driven analysis of representations around the word ‘Muslim’ in the British Press 1998–2009.” *Applied Linguistics* 2013, 34(3), 255–78.
- Biber, Douglas, Susan Conrad and Viviana Cortes. 2004. “If you look at.: Lexical bundles in university teaching and textbooks.” *Applied Linguistics* 25(3), 371–405.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *The Longman grammar of spoken and written English*. London: Longman.
- Biber, Douglas and Susan Conrad. 2005. “Register variation: A corpus approach.” In *The handbook of discourse analysis*, edited by Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton, p. 175–95. US: Wiley-Blackwell.
- Biber, Douglas. 2011. “Corpus linguistics and the study of literature: Back to the future?” *Scientific Study of Literature* 1(1), 15–23.
- Carter, Ronald. 2010. “Methodologies for stylistic analysis: Practices and pedagogies.” In *Language and style*, edited by Beatrix Busse and Dan McIntyre, p. 55–68. Basingstoke: Palgrave.
- Coetzee, John Maxwell. 1999. *Disgrace*. New York: Penguin Books.
- Coleman, Donaldson. 2009. “The ‘dog-man’: Race, sex, species, and lineage in Coetzee’s disgrace.” *Twentieth Century Literature* 55(4). 597–617. <http://www.jstor.org/stable/25733433> (29-11-2020).
- Culpeper, Johnathan. 2002. “Computers, language and characterisation: An analysis of six characters in *Romeo and Juliet*.” In *Conversation in life and in literature: Papers from the ASLA Symposium Uppsala*, edited by Ulla Melander-Marttala, C. Ostman and M. Kyto, 8–9 November. Uppsala: Universitetsstryckeriet.

- Culpeper, Jonathan. 2009. "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.
- Easton, Kai. 2007. "J. M. Coetzee's disgrace: Reading race/reading scandal." In *Scandalous fictions*, edited by Jago Morrison and Susan Watkins. London: Palgrave Macmillan. doi: 10.1057/9780230287846\_11.
- Fischer-Starcke, Bettina. 2006. "The phraseology of Jane Austen's *Persuasion*: Phraseological units as carriers of meaning." *ICAME Journal* 30, 87–104.
- Fischer-Starcke, Bettina. 2009. "Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*. A corpus stylistic analysis." *International Journal of Corpus Linguistics* 14(4), 492–523.
- Fischer-Starcke, Bettina. 2010. *Corpus linguistics in literary analysis. Jane Austen and her contemporaries*. London, New York: Continuum.
- Golden, Audrey J. 2019. "Nadine Gordimer and the force of law: Revisiting *My Son's Story*." *Commonwealth Essays and Studies* 41(2), Spring 2019, 75–89. <https://search.informit.com.au/documentSummary;dn=733343554159902;res=IELLCC> > ISSN: 0395-6989. (2-1-2021).
- Gordimer, Nadine. 1990. *My Son's Story*. London: Bloomsbury.
- Graham, Lucy Valerie. 2003. "Reading the unspeakable: Rape in J. M. Coetzee's disgrace." *Journal of Southern African Studies* 29(2), 433–44. doi: 10.1080/03057070306207. (10-12-2020).
- Ho, Yufang. 2011. *Corpus stylistics in principles and practice: A stylistic exploration of John Fowles' the magus (advances in stylistics)*. London: Continuum.
- Hori, Masahiro. 2004. *Investigating Dickens' style. A collocational analysis*. Basingstoke: Palgrave Macmillan.
- Kilgariff, Adam. 2009. "Simple math for keywords." In *Proceedings of corpus linguistics Conference CL2009*, edited by Michaela Mahlberg, Victorina González-Díaz and Catherine Smith. UK: University of Liverpool.
- Leech, Geoffrey and Mick Short. 2007. *Style in fiction*. 2nd ed. London, New York: Longman.
- Mahlberg, Michaela and Dan McIntyre. 2011. "A case for corpus stylistics: Ian Fleming's Casino Royale." *English Text Construction* 4(2), 204–27.
- Mahlberg, Michaela, Viola Wiegand, Peter Stockwell and Anthony Hennessey. 2019. "Speech-bundles in the 19th-century English novel." *Language and Literature* 28(4), 326–53.
- Mahlberg, Michaela. 2005. *English general nouns: A corpus theoretical approach*. Amsterdam: John Benjamins.
- Mahlberg, Michaela. 2007a. "Corpus stylistics: bridging the gap between linguistic and literary studies." In *Text, discourse and corpora: Theory and analysis*, edited by Michael Hoey, Michaela Mahlberg, Michael Stubbs, and Wolfgang Teubert, p. 219–46. London: Continuum.
- Mahlberg, Michaela. 2007b. "Lexical items in discourse: identifying local textual functions of sustainable development." In *Text, discourse and corpora: Theory and analysis*, edited by Michael Hoey, Michaela Mahlberg, Michael Stubbs, and Wolfgang Teubert, p. 191–218. London: Continuum.
- Mahlberg, Michaela. 2010. "Corpus linguistics and the study of nineteenth century fiction." *Journal of Victorian Culture* 15(2), 292–8.
- Mahlberg, Michaela. 2012. "Corpus analysis of literary texts." In *The encyclopedia of applied linguistics*, edited by Carol A. Chapelle. Blackwell. doi: 10.1002/9781405198431.wbeal0249. (13-6-2020).
- Mahlberg, Michaela. 2014. "Corpus stylistics." In *The Routledge handbook of stylistics*, edited by Michael Burke, p. 378–92. London: Routledge.
- Malloy, Catherine. 2004. "Reading J. M. Coetzee's disgrace: In search of the good." In *Mystery in its passions: Literary explorations. Analecta Husserliana (The yearbook of phenomenological research)* 82, edited by Anna-Teresa Tymieniecka. Dordrecht: Springer. doi: 10.1007/978-94-007-1017-7\_10.
- Marais, Mike. 2006. "J. M. Coetzee's 'disgrace' and the task of the imagination." *Journal of Modern Literature* 29(2), 75–93. <http://www.jstor.org/stable/3831793> (1-1-2021).
- Marais, Mike. 2001. "Impossible possibilities: Ethics and choice in J. M. Coetzee's *The Lives of Animals* and *Disgrace*." *English Academy Review: Southern African Journal of English Studies* 18(1), 1–20. doi: 10.1080/10131750185310031.
- Mardorossian, Carine. 2011. "Rape and the violence of representation in J. M. Coetzee's disgrace." *Research in African Literatures* 42(4), 72–83. doi: 10.2979/reseafillite.42.4.72.
- Matthiessen, Christian. 1993. "Register in the round: diversity in a unified theory of register analysis." In *Register analysis: Theory and practice*, edited by Mohsen Ghadessy, p. 221–91. London: Pinter.
- McEnery, Tony and Andrew Wilson. 2001. *Corpus linguistics: An introduction*. 2nd ed. Edinburgh: Edinburgh University Press.
- McIntyre, Dan and Brian Walker. 2010. "How can corpora be used to explore the language of poetry and drama?" In *The Routledge handbook of corpus linguistics*, edited by Michael McCarthy and Anne O'Keefe, p. 516–30. Abingdon: Routledge.
- McIntyre, Dan and Brian Walker. 2019. *Corpus stylistics: Theory and practice*. 1st ed. Edinburgh University Press.
- Möller, Karin. 1999. "'Bird out there': Traversing the middle space in Nadine Gordimer's *My Son's Story*, critique." *Studies in Contemporary Fiction* 40(2), 161–71. doi: 10.1080/00111619909601571.
- Myrdal, Gunnar. 1970. *Objectivity in social research*. London: Gerald Duckworth & Company Ltd.
- Oriaku, Remy. 2016. "J. M. Coetzee's disgrace as an allegory of the pain, frustration, and disorder of post-apartheid South Africa." *Matatu* 48(1), 145–60.

- Partington, Alan and John Morley. 2004. "At the heart of ideology: Word and cluster/bundle frequency in political debate." In *PALC 2003: Practical applications in language corpora*, edited by Barbara Lewandowska-Tomaszczyk, p. 179–92. Frankfurt: Peter Lang.
- Rayson, Paul. 2008. "From key words to key semantic domains." *International Journal of Corpus Linguistics* 13(4), 519–49.
- Rinzler, Simone. 2013. "Coetzee's style in disgrace." *Etudes de Stylistique Anglaise* 4, 191–203. doi: 10.4000/esa.1534.
- Sakamoto, Toshiko. 2002. "Coloured's identity and cultural transformation in Nadine Gordimer's *My Son's Story*." [www.ritsumei.ac.jp/~RitsILCS\\_14.1pp.313-330SAKAMOTO.pd](http://www.ritsumei.ac.jp/~RitsILCS_14.1pp.313-330SAKAMOTO.pd). (1-1-2021).
- Scott, Mike and Christopher Tribble. 2006. *Textual patterns. Key words and corpus analysis in language education*. Amsterdam: John Benjamins.
- Semino, Elena and Mick Short. 2004. *Corpus stylistics. Speech, writing and thought presentation in a corpus of English writing*. London: Routledge.
- Sewall, Harry. 2013. "Contested epistemological and ethical spaces: The place of non-humans in Milan Kundera's the unbearable lightness of being and J. M. Coetzee's *Disgrace*." *English Academic Review: Southern African Journal of English Studies* 30(1), 76–91. doi: 10.1080/10131752.2013.783392.
- Sinclair, John. 2003. *Reading concordances. An introduction*. Harlow: Pearson Education.
- Sinclair, John. 2004. *Trust the text. Language, corpus and discourse*. London: Routledge.
- Stott, Graham St. John. 2009. "Rape and silence in J. M. Coetzee's *disgrace*." *Philosophical Papers* 38(3), 347–62. doi: 10.1080/05568640903420905.
- Stubbs, Michael and Isabel Barth. 2003. "Using recurrent phrases as text-type discriminators: a quantitative method and some findings." *Functions of Language* 10 (1), 65–108.
- Stubbs, Michael. 2005. "Conrad in the computer: Examples of quantitative stylistic methods." *Language and Literature* 14(1), 5–24.
- Stubbs, Michael. 2007. "Quantitative data on multi-word sequences in English: The case of the word *world*." In *Text, discourse and corpora. Theory and analysis*, edited by Michael Hoey, Michaela Mahlberg, Michael Stubbs, and Wolfgang Teubert, p. 163–89. London: Continuum.
- Swales, John M. 2003. "Sex, shame and guilt: Reflections on Bernhard Schlink's *Der Vorleser* (*The Reader*) and J. M. Coetzee's *disgrace*." *Journal of European Studies* 33(1), 7–22. doi: 10.1177/0047244103033001002.
- Teubert, Wolfgang. 2005. "My version of corpus linguistics." *International Journal of Corpus Linguistics* 10(1), 1–13.