

# The corpus stylistic analysis of fiction – or the fiction of corpus stylistics?

Michaela Mahlberg

University of Nottingham

## Abstract

*This paper argues that corpus stylistics can contribute methodologies and concepts to support the investigation of character information in fiction. Focusing on Charles Dickens, the paper looks at lexicogrammatical patterns as well as places in the literary text. It suggests that clusters, i.e. repeated sequences of words, and suspensions, i.e. interruptions of characters' speech by the narrator, can serve as textual cues in the process of characterization. These concepts are illustrated with examples for the characters Bucket and Tulkinghorn in Bleak House. The analysis of the examples leads to an outline of challenges for corpus stylistics that result from the need to interpret features on the textual surface in relation to the effects they might have on the processing of the text by readers.<sup>1</sup>*

## 1. Introduction

The number of studies that approach literary texts with corpus linguistic methods is growing and the term 'corpus stylistics' is becoming increasingly popular. Corpus stylistics has started to be recognized as an emerging field or branch by both corpus linguists (cf. Chapelle 2011; Lindquist 2009; O'Keeffe and McCarthy 2010) and stylisticians (cf. Jeffries and McIntyre 2010; Leech and Short 2007; McIntyre and Busse 2010). It is set in the wider context of the Digital Humanities where new technologies enable new ways of conducting research and both support and challenge traditional approaches in the humanities. Part of the challenge is to assess the usefulness of innovative technologies. Any level of sophistication in the storage, presentation and manipulation of data cannot replace the need for human analysis and the identification of relevant research questions. Against this background it seems worthwhile to consider what kind of contributions corpus stylistics can make to the study of literary texts. What makes a text 'literary' is to some extent determined by its reception through the reader, who approaches the text "in a literary way" (cf. Carter 2004: 69). The enthusiasm about the application of corpus methods to the study of literary texts holds the danger of overinterpreting findings and viewing their implications as far more important than they might be in a wider critical context. Corpus methods help us to focus on salient features on the textual surface, but corpus methods alone cannot assess the relevance of those features for the process of reading a text and creating a textual world in the reader's mind. In this article I want to argue that the strengths of corpus stylistics are most apparent when it is linked to literary

criticism (or stylistic approaches in a wider sense) and when findings are not presented in isolation but as part of a bigger picture, or as Carter (2010: 67) puts it: “Corpus stylistic analysis is a relatively objective methodological procedure that at its best is guided by a relatively subjective process of interpretation”.

To explore the usefulness of corpus stylistic methodology, I want to focus on characterization in Dickens and in particular look at two characters from *Bleak House*: Inspector Bucket and the lawyer Tulkinghorn. I argue that clusters and suspensions provide two options of looking for character information. It will be shown that the character information found with the help of those methods can be linked to literary criticism on characterization in Dickens. Thus corpus stylistics is seen as complementing other approaches. In the final section of this article I will highlight some of the challenges that such a corpus stylistic approach has to face.

## 2. Extending the checklist of textual cues for character information

Culpeper (2001) views characterization as a process in which impressions of characters are formed in the reader’s mind. These impressions result from the interplay of top-down and bottom-up processes in the reading of texts. The bottom-up processes draw on cues that are found in the text. Culpeper suggests a number of textual cues that can be useful starting-points for the analyst. In his list, he includes, for instance, conversational structure, conversational implicature, terms of address, dialect, syntactic features, etc. The list illustrates how stylistics can draw on a range of linguistic concepts and approaches. Corpus linguistics adds to the development of linguistic concepts. The impact of such developments is also seen in Culpeper’s (2001) approach when he includes key words in his checklist. As Culpeper (2001) mainly focuses on plays, key words are suggested as a useful approach to compare the speech of different characters. Following Culpeper (2001), McIntyre (2010) illustrates how key words and key semantic domains can contribute to the study of characters in a screenplay. I want to suggest two further concepts that seem helpful additions to a checklist for character information: clusters (also briefly referred to in McIntyre 2010) and suspensions. The motivation for the inclusion of these concepts is a combination of arguments from corpus linguists and literary criticism. In the following I begin with an overview of the concepts which will then be discussed in more detail when applied to the characterization of Bucket and Tulkinghorn.

### 2.1 Patterns: Clusters

Clusters are repeated sequences of words, e.g. *I don’t know what*. The fact that such sequences are used repeatedly suggests that they are associated with textual functions. Focusing on five-word clusters in Dickens, the following functions can be identified: clusters label or refer to characters and themes, are part of characters’ speech, describe body language, introduce comments by the narrator

and function in reference to time and place. Examples (1) to (5) illustrate these functions respectively, the five word clusters are in italics. As the examples show, a cluster can also be part of a larger expression with the respective function, but for the sake of brevity, I refer to the cluster as having the function.

- (1) [...] and the *young lady with the black eyes* and the fur round the boots, whispered something in Emily's ear, and then [...]  
(*The Pickwick Papers*)
- (2) [...] Thank you, Mr. Bumble, sir, *I am very much obliged* to you, I'm sure.'  
(*Oliver Twist*)
- (3) [...] said Riderhood, when his visitor sat down, resting his chin on his hand, *with his eyes on the ground*.  
(*Our Mutual Friend*)
- (4) Exceedingly red-eyed and grim, *as if he had been* up all night at a party which had taken anything but a convivial turn, Jerry Cruncher worried his breakfast rather than ate it, [...]  
(*A Tale of Two Cities*)
- (5) Kit stood *in the middle of the road* [...].  
(*The Old Curiosity Shop*)

Because of the functions they fulfil, clusters can be regarded as building blocks of fictional worlds. They point to settings in space and time, indicate the presence of a narrator, highlight themes, refer to characters and provide information on characters through their speech and body language. These observations are mainly based on the clusters in Dickens and would need further study for other novels. For Dickens the most important function of clusters seems to be the provision of character information which can be achieved by four out of the five groups: labels, speech, body part clusters, but also *as if* clusters as illustrated in example (4). This fact might be related to the importance of characters in Dickens. In Dickens characters seem to be almost more important than plot (cf. also Kucich 1994: 403). More importantly, however, observations on clusters can be linked to the externalization of characters as will be explained below. A question about the applicability of the cluster approach to the analysis of specific texts is concerned with the frequency of the clusters. How many 'useful' clusters can be found for any one novel to draw some meaningful conclusions? Or does the usefulness of clusters lie in the cumulative picture that they provide for Dickens's novels as a whole? I will return to these questions, when I discuss two characters from the novel *Bleak House*. It will become clear that clusters alone provide only limited information and their main value lies in their consideration alongside other concepts and approaches. One such concept is the suspension.

## 2.2 Places: Suspensions

A ‘suspension’ is a span of five or more words of (narrator) text which interrupts a span of quoted speech (or thought, or writing). A suspension occurs in the same orthographic sentence as the speech it interrupts (Mahlberg and Smith 2010: 461). This definition is motivated by Lambert’s (1981: 6) concept of the suspended quotation, which he describes as “a protracted interruption by the narrator of a character’s speech”, with ‘protracted’ meaning “containing at least five words”. Lambert’s (1981) main point about suspensions is that they reflect a hostile attitude of Dickens towards the characters he created. Leaving aside this very specific interpretation as well as technical detail as to differences between Lambert’s definition and the one used here, suspensions seem to be a useful place to look for character information and specifically body language. The following is an example of a suspension (the suspension is highlighted in italics).

- (6) “Once had a friend and brother serjeant of the same name,” *says Mr. Bucket, offering his hand*, “and consequently feel a liking for it. Mrs. Chadband, no doubt?”

*(Bleak House)*

While the suspension interrupts the character’s speech it also creates the impression of immediacy between the speech and body language. This is an important effect, as the description of body language in literature can seem lengthy compared to the time that would be taken up by the body language in the real world (cf. also Korte 1997). Also, in reality body language often accompanies speech. Thus, suspensions seem to be a useful place to create an effect of simultaneous representation. In addition to presenting body language, suspensions can also have other functions that will be illustrated in Section 3.2 below. Suspensions are not limited to Dickens, but also occur in the works of other authors. For a discussion of suspensions in *Pride and Prejudice* see Mahlberg and Smith (2010). Findings for suspensions also link in with observations by Busse (2010) on thought presentation and narrative reference to the thinker’s eye. A corpus approach can aid the study of suspensions through the annotation of the text, so that concordance searches, for instance, can focus on the text in suspensions.

## 2.3 Interpreting the cues: Externalization of characters

John (2001) argues that characters in Dickens are externalized. She seeks to explain Dickens’s approach to characterization through his views on culture and the role of entertainment. With reference to Dickens’s own writings on ‘The Amusements of the People’, for instance, John (2001) argues that Dickens believed in popular forms of entertainments as those best serve cultural inclusivity. John (2001) points out that for Dickens’s characters the focus is not

on hidden inner lives, but feelings and emotions that are displayed. Thus John (2001) draws parallels to the modes of melodrama where the ostension of character is prevalent. Although John (2001) does not explicitly discuss approaches to the representation of body language in literature, Korte's (1997) study provides support for John's argument (2001). Korte (1997) suggests a modal-functional approach to the description of body language, drawing on research on body language in the real world and linking her modal-functional categories to their functions in literary texts, such as characterization, dramatization or authentication. Korte's (1997) approach is based on the analysis of samples of texts from four different periods to identify tendencies as to the main categories of body language in each period. Because of the importance of Dickens for the 19th century, examples from his texts are naturally included to illustrate the various types of body language that Korte (1997) discusses. John's claims about Dickens's preferences for transparent characters is supported by Korte's (1997) observation on the importance of 'externalizers' in the 19th century, i.e. forms of body language that function to display relatively stable features of characters.

The interpretation of Dickens's approach to characterization as the externalization of characters is to some extent based on the observation that Dickens does not present the inner life of characters. Corpus stylistics is less suited for the discussion of what is not there than for an analysis of what is there. Unless the description of inner lives is linked to precise features that can be quantified, corpus findings cannot be interpreted as evidence of the absence of something. Additionally, Culpeper (2001) points out that while potentially any information could be relevant to the construction of character, contextual factors have an impact on how 'characteristic' behaviour is perceived to be or to what extent it is seen as being determined by its context. In the following section I want to address the question of what kind of character information clusters and suspensions can provide and how this information is to be interpreted in the textual context of the novel and the interpretative context of the externalization of character. This analysis will lead to an outline of some of the challenges that corpus stylistics is faced with.

### **3. Bucket and Tulkinghorn**

In the following two sections I want to discuss the kind of character information that clusters and suspensions provide for Bucket and Tulkinghorn. Section 3.1 compares the two characters with regard to body language clusters that are found for them in *Bleak House*. Section 3.2 shows how concordance searches of the characters' names in suspensions reveal character information.

### 3.1 Clusters

The five groups of clusters illustrated above were identified by investigating clusters in a corpus of 23 texts by Charles Dickens (cf. Mahlberg 2007a). I have previously shown how viewing clusters in a single text can highlight thematic features and links between characters (cf. Mahlberg 2007b, 2007c) and how a general picture provided by clusters can serve as background for the analysis of a specific text extract (Mahlberg 2009). In the following, I want to focus on two characters in the novel *Bleak House* to examine what kind of character information clusters might provide and how this information relates to other character cues in the text. *Bleak House* contains 97 five-word clusters that occur at least five times. Most of the clusters are labels, i.e. they contain names or are otherwise specific to *Bleak House*. The three most frequent clusters are *not to put too fine*, *to put too fine a*, *put too fine a point* each occurring 15 times. They are classed as a label as they only appear in *Bleak House* (when looking at the Dickens Corpus and 19C). They are in fact part of Mr Snagsby's favourite phrase *not to put too fine a point (up)on it*. It does not seem necessary to employ the computer to find such phrases, as a major point about their existence in the novel is to be striking and noticeable. Such phrases have been described as 'tags' that can serve to identify characters and make the reader remember language or behaviour specific to a character.

Based on an initial overview of the five-word clusters in *Bleak House*, Bucket and Tulkinghorn appear as different kinds of characters. Bucket is associated with clusters such as *Sir Leicester Dedlock Baronet and*, *Sir Leicester Dedlock Baronet I*, *by Sir Leicester Dedlock Baronet*, *now Sir Leicester Dedlock Baronet* which show his fondness for using titles and also portray him in a comic way. Also his own title, *Inspector Bucket of the Detective*, appears among the clusters. For Tulkinghorn in contrast, there is no five-word cluster that is so strikingly associated with him. Tulkinghorn and Bucket still have sufficient similarities to enable a comparison between the two. Both are men of authority, and by virtue of their profession are associated with the law, Bucket as an inspector and Tulkinghorn as a lawyer and both are involved in discovering Lady Dedlock's secret.

If we focus on body language as character information, the five-word clusters provide different kinds of information for the two characters. Table 1 contains all five-word clusters in *Bleak House* that contain body parts or can be regarded as body language clusters because of the information they provide (*leaning back in his chair*). In addition to the frequencies of the clusters in *Bleak House*, Table 1 provides information on how often the clusters refer to Tulkinghorn (column 'T') and Bucket (column 'B'). Three of the clusters have occurrences that refer to Tulkinghorn. Bucket, in contrast does not have any body language clusters. What this information tells us is not that there is no description of body language for Bucket. It only tells us that there are no descriptions of body language that appear in repeated sequences of at least five words. Body language is in fact important for the description of Bucket, as one of his characteristic features is his use of his

forefinger that is repeatedly referred to and described very pointedly in example (7):

- (7) Mr. Bucket and his fat *forefinger* are much in consultation together under existing circumstances. When Mr. Bucket has a matter of this pressing interest under his consideration, the fat *forefinger* seems to rise, to the dignity of a familiar demon. He puts it to his ears, and it whispers information; he puts it to his lips, and it enjoins him to secrecy; he rubs it over his nose, and it sharpens his scent; he shakes it before a guilty man, and it charms him to his destruction. The Augurs of the Detective Temple invariably predict that when Mr. Bucket and that *finger* are in much conference, a terrible avenger will be heard of before long.

(Chapter 53)

Table 1. Body language clusters in *Bleak House*

Cluster	BH	B	T	D	D texts	19C	19C texts
his head against the wall	8	0	0	12	5	0	0
with his head against the	7	0	0	10	4	0	0
with his hands behind him	8	0	6	20	9	6	3
his hands in his pockets	5	0	2	90	20	13	8
with his back to the	5	0	2	43	14	22	11
leaning back in his chair	5	0	0	38	13	8	4

The body language of Tulkinghorn that is described by the clusters in Table 1 is not highlighted in the same way that Bucket's forefinger is. Instead, its meanings are defined through the contexts in which it appears. Table 1 gives some initial pointers to those contexts. The table shows the frequency of the clusters in *Bleak House*, their frequency in the whole Dickens Corpus and in a reference corpus made up of nineteenth century novels by authors other than Dickens. As the Dickens corpus and reference corpus are of similar size (about 4.5 million words each), numbers of occurrences in the corpus are not normalized. The table also contains the number of texts in the Dickens Corpus and the reference corpus in which the clusters occur. All clusters that refer to Tulkinghorn appear both in Dickens and in the reference corpus, which underlines the fact that the clusters can have different functions in different contexts. The cluster *his hands in his pockets*, for instance, which is the most frequent five-word body language cluster in Dickens can occur in contexts where it highlights exaggerated or comic behaviour as well as in contexts where it appears to add to the authenticity of a description (cf. Mahlberg 2007a). In example (8) Tulkinghorn's having his hands in his pockets is depicted as reflecting indifference, in example (9) his body

language adds to show the calmness with which he conducts his conversation with Lady Dedlock.

- (8) Mr. Tulkinghorn rises *with his hands in his pockets* and walks into one of the window recesses. “Now! I have no time to waste.” In the midst of his perfect assumption of indifference, he directs a sharp look at the trooper, taking care to stand with his own back to the light and to have the other with his face towards it.

(Chapter 34)

- (9) “My experience teaches me,” says Mr. Tulkinghorn, who has by this time got *his hands in his pockets* and is going on in his business consideration of the matter like a machine. “My experience teaches me, [...]”

(Chapter 41)

The two cases of *with his back to the* refer to Tulkinghorn standing with his back to the fire/chimney-piece which was a common pose in the nineteenth century characteristic of men and associated with power as the fire was the most prominent place of the room (cf. Korte 1997: 212). The fact that Tulkinghorn is repeatedly depicted *with his hands behind him* points to him being calm, distanced and passionless. This becomes particularly apparent in example (10), where he and Krook find Nemo dead. Tulkinghorn does not show any interest when the doctor examines Nemo and discusses his cause of death with Krook. Tulkinghorn is described as standing “aloof”, being “removed”, his face is “imperturbable” and “inexpressive”, all that he shows is his “shell”.

- (10) During this dialogue Mr. Tulkinghorn has stood aloof by the old portmanteau, *with his hands behind him*, equally removed, to all appearance, from all three kinds of interest exhibited near the bed – from the young surgeon’s professional interest in death, noticeable as being quite apart from his remarks on the deceased as an individual; from the old man’s unctious; and the little crazy woman’s awe. His imperturbable face has been as inexpressive as his rusty clothes. One could not even say he has been thinking all this while. He has shown neither patience nor impatience, nor attention nor abstraction. He has shown nothing but his shell. As easily might the tone of a delicate musical instrument be inferred from its case, as the tone of Mr. Tulkinghorn from his case.

(Chapter 11)

As the figures in Table 1 are relatively low, they are best interpreted as tendencies. Overall, they suggest that body language can occur with a range of different functions, appearing natural or unobtrusive or having more specific effects. This point is further supported by the distribution of the clusters *his head against the wall/with his head against the* which is neither related to Bucket nor Tulkinghorn. All instances of the two clusters in *Bleak House* refer to the same

character (Mr Jellyby) and the body language identifies him uniquely. These clusters do not occur in the reference corpus, and their overall occurrence in the Dickens corpus as well as the distribution across chapters is lower than for the other clusters. This seems to reflect the unique function of these clusters in *Bleak House* and shows another option of highlighting character information: the body language functions in a way similar to Snagby's favourite phrase or the clusters reflecting Bucket's fondness of titles as exemplified above.

### 3.2 Suspensions

The concordance in Figure 1 shows the occurrences of *Bucket* in suspensions. Suspensions fall into three groups with regard to the character information that they provide. They can contain body language, the narrator's interpretation of speech or direct characterization, as illustrated by examples (11) to (13) respectively. Example (11) describes body language so that it is left to the reader to draw conclusions from this observation. In example (12) in contrast, the reader only gets the narrator's interpretation of the way in which Bucket speaks, but there is no information as to the extent to which the *very seriously* is reflected by the tone of Buckets voice, his facial expression or maybe even his posture. Example (13) is even less descriptive. In relation to Bucket's speech, the information in the suspension does not provide additional detail on how he speaks but comments on what he says. The narrator emphasizes that Buckets likes to use full titles, which the reader may have already noticed because of Bucket's repeated use of *Sir Leicester Dedlock, Baronet*. However, by making this point in a suspension and presenting it as a character trait it receives even more prominence.

- (11) "[...] And Lord," says Mr. Bucket, opening his arms, "here's children too! [...]"
- (12) "Sir Leicester Dedlock, Baronet," returns Mr. Bucket very seriously, "I hope it may [...]"
- (13) "Sir Leicester Dedlock, Baronet," proceeds Mr. Bucket, who delights in a full title and does violence to himself when he dispenses with any fragment of it, "the last point in the case [...]"

The initial categorization of character information in suspensions into three types provides a simplified account, as categories also overlap. Example (14) illustrates body language that is not only described but also interpreted: Bucket does not only have his head on one side, but his head is *persuasively* on one side. Additionally, the reference to his forefinger has to be seen as part of a bigger picture. Bucket's use of his forefinger is presented as a character trait and is referred to repeatedly. Its significance is explained explicitly in example (7) above. In example (14) the presence of the forefinger points to the importance of the secret conversation between Bucket and Sir Leicester. In the concordance in

Figure 1, there are five occurrence of the (*fore*)finger visible in the given context (there are more if the suspensions are viewed in full) illustrating this characteristic feature.

1	other. "I don't mind telling YOU," says	Bucket	with an engaging appearance of frankness
2	stinctly. "Now, what YOU want," pursues	Bucket,	again tapping Mr. Snagsby on the breast
3	n't what you endeavour to do," says Mr.	Bucket,	shaking hands with him and blessing him
4	ight." "You see, Mr. Snagsby," says Mr.	Bucket,	accompanying him to the door and shaking
5	ugh workman. And you needn't," says Mr.	Bucket	in a considerate and private voice, "you
6	uld speak. "Now, George," continues Mr.	Bucket,	putting his hat upon the table with an a
7	you suspect ME?" "George," returns Mr.	Bucket,	keeping his forefinger going, "it is cer
8	uld you believe it, governor," says Mr.	Bucket,	struck by the coincidence, "that when I
9	der?" "Murder!" "Now, George," says Mr.	Bucket,	keeping his forefinger in an impressive
10	thunderstruck. "Now, George," says Mr.	Bucket,	urging a sensible view of the case upon
11	sides. "Now George, old boy," says Mr.	Bucket,	taking his arm at the shop-door, "come a
12	with you, governor? And Lord," says Mr.	Bucket,	opening his arms, "here's children too!
13	ted station of society, miss," says Mr.	Bucket,	quite reddening at another narrow escape
14	YOU know life, you know, sir," says Mr.	Bucket	with a complimentary twinkle of his eye
15	rpose. "Why you see, miss," returns Mr.	Bucket,	bringing the finger into persuasive acti
16	Leicester Dedlock, Baronet," returns Mr.	Bucket	very seriously, "I hope it may at one an
17	twenty-four hours. "And this," says Mr.	Bucket,	spreading it out on the table, "is in th
18	fference. Now I think of it," says Mr.	Bucket,	warming his hands and looking pleasantly
19	s a beautiful case, you see, miss," Mr.	Bucket	goes on, glancing gravely at Sir Leicest
20	Leicester Dedlock, Baronet," proceeds Mr.	Bucket,	who delights in a full title and does vi
21	Leicester Dedlock, Baronet," returns Mr.	Bucket	with his head persuasively on one side a
22	es!" "Now, when you're ready," says Mr.	Bucket	after awaiting his recovery, "to come to
23	r Leicester Dedlock, Baronet," says Mr.	Bucket,	and from this time forth the finger neve
24	othing( this bit of business," says Mr.	Bucket	--Sir Leicester mechanically bows his he
25	y, of course you wanted to get in," Mr.	Bucket	asserts with cheerfulness; "but for a ol
26	a menace. "Now, mademoiselle," says Mr.	Bucket	in a cool determined way, "you go and si
27	man of business, you are," returns Mr.	Bucket,	very attentive, "and consequently you're
28	ded is now in this house," proceeds Mr.	Bucket,	putting up his watch with a steady hand
29	w, Sir Leicester Dedlock, Baronet," Mr.	Bucket	begins, standing over him with one hand
30	m limb." "Bless you, darling," says Mr.	Bucket	with the greatest composure, "I'm fully
31	Don't you think any more," returns Mr.	Bucket	with admonitory finger, "of throwing you
32	er serjeant of the same name," says Mr.	Bucket,	offering his hand, "and consequently fee
33	nods. "Now, mademoiselle," repeats Mr.	Bucket,	making no demonstration except with the
34	y man." "Now I tell you what," says Mr.	Bucket,	instantaneously altering his manner, com
35	None was charged against him," said Mr.	Bucket,	coolly lifting off his hat, "but I suppo
36	out of employment, I believe," said Mr.	Bucket	apologetically for Michael Jackson, "and
37	y--" "Who has been here, you know," Mr.	Bucket	struck in, addressing the whole group wi
38	o with great speed. "My dear," said Mr.	Bucket,	jumping to his seat and looking in again
39	t's possible, Miss Summerson," said Mr.	Bucket,	pondering on it, "that her ladyship sent
40	the only amends you can make," said Mr.	Bucket,	rapidly glancing at it, "is to let me sp
41	And to hold the candle," pursued Mr.	Bucket	without correcting himself, "or hold her
42	arrow street. "Mr. Woodcourt," said Mr.	Bucket,	who had eyed him closely as we came alon
43	ce. "In consequence of which," said Mr.	Bucket,	dismissing his agreeable manner all at o
44	d so you chance to find, you know," Mr.	Bucket	went on, stooping over him with an air o
45	e. That's the drollery of it," said Mr.	Bucket	with the same lively air of recalling a

Figure 1. Concordance of *Bucket* in suspensions

- (14) “Why, Sir Leicester Dedlock, Baronet,” returns Mr. Bucket *with his head persuasively on one side* and his *forefinger* pendant at one ear like an earring, “we can’t be too private just at present. [...]”  
(Chapter 54)

The forefinger is one example of how information in suspensions has to be seen in relation to a wider textual context. There are also examples in which the textual relations are more immediate. In concordance line 2, the adverb *again* points to the relationship illustrated by examples (15a) and (15b) which are a few lines apart. Here Bucket repeatedly taps Snagsby on the breast and the repetition is highlighted in the suspension. Examples (16a) and (16b) also show related examples. This time both are suspensions, corresponding to concordance lines 44 and 45 respectively. Although the wording is not exactly repeated the similarity between the suspensions is underlined through the repetition of *same*: *the same lively air* and *the same crest-fallen appearance* which emphasizes Smallweed feeling very uncomfortable in the position Buckets put him in. Example (16) not only shows the extent a suspension can take on, it also illustrates that suspensions do not necessarily contain information that refers (only) to the speaker. In suspensions the narrator can also comment on the effects that the manner of speech and its accompanying actions can have on other characters. It seems that the more information is packed into a suspension the more it actually interrupts the flow of speech which in turn also gives prominence to the information in the suspension.

- (15) a. “Yes! And lookee here, Mr. Snagsby,” resumes Bucket, taking him aside by the arm, *tapping him familiarly on the breast*, and speaking in a confidential tone.  
b. “Now, what YOU want,” pursues Bucket, *again tapping Mr. Snagsby on the breast* in a comfortable and soothing manner, “is [...]”  
(Chapter 22)
- (16) a. “[...] so you chance to find, you know,” Mr. Bucket went on, stooping over him with an air of cheerful raillery which Mr. Smallweed by no means reciprocated, “and so you chance to find, you know, a paper with the signature of Jarndyce to it. Don’t you?”  
b. “[...] That’s the drollery of it,” said Mr. Bucket with the *same* lively air of recalling a joke for the enjoyment of Mr. Smallweed, who still had the *same* crest-fallen appearance of not enjoying it at all; “what do you find it to be but a will?”  
(Chapter 62)

Compared to example (16), example (11) contains relatively unobtrusive information: the suspension is much shorter and it does not contain narrator’s interpretation or comment. Instead of focusing on the interruption that a suspension entails to the flow of speech, in the case of example (11) the

suspension seems to suggest a simultaneousness of Bucket's body language with his speech. He has just entered the room and greeted all the adults before he now addresses the children and gives them a hug.

When a suspension contains explicit character information, as in example (13) above, or refers to character traits that occur repeatedly, such as Bucket's use of his forefinger, it appears straightforward to see them as textual cues for characterization. For examples such as (12), however, it is not immediately obvious whether Bucket's seriousness is part of his character or whether it results from the kind of situation he is faced with. This question is not specific to examples in suspensions but holds for all potential textual cues of character information. Culpeper (2001) draws on attribution theory to assess the relevance of textual cues. However, it seems that the analysis of the concordance in Figure 1 also gives some pointers as to how individual instances fit into a bigger picture. We find examples that present Bucket as pleasant or cheerful: line 1 "with an engaging appearance", line 18 "looking pleasantly at the blaze", line 25 "asserts with cheerfulness". But there are also examples showing him more serious and composed: line 19 "glancing gravely at Sir Leicester, line 26 "in a cool determined manner, line 30 "with the greatest composure. At the same time, the concordance also contains information that helps explain such variety in behaviour. Line 34 refers to an instance of Bucket "instantaneously altering his manner", or line 43 shows him "dismissing his agreeable manner all at once and becoming strictly business-like". He is able to adjust his behaviour and uses this ability strategically. We can see his versatility as a characteristic feature. Throughout the novel Bucket appears very friendly and cheerful, but also very serious and business-like. This feature is very apparent in Chapter 49 when he arrests his friend George. He does what he sees as his duty as an officer investigating a murder case, but he also wants to "make things as pleasant as is consistent with [his] duty" and offers George the option of more comfortable handcuffs. With the detail that the suspensions provide on Bucket, they appear as useful places to look for information on this character. It is important, however, to stress that the investigation of suspensions, as all corpus stylistics methods, is not a replacement for reading the novel. It is a method to complement the 'manual' analysis of the novel. For instance, one of the aspects that the suspension information links up with is Bucket's way of suggesting that his presence is incidental. Similar to adjusting his manner to serve his investigation, he also tries to make his presence appear natural. When he arrests George he appears to be joining in the celebration of Mrs Bagnet's birthday (cf. Chapter 49), or when he recounts how he saw Lady Dedlock in the garden at night he claims to have been on his way to visit his aunt (cf. Chapter 53).

By highlighting Bucket's variation in manner and attempts of being calculating and appearing controlled the opposite effect is created. He is not portrayed as a subtle person, but has a very strong presence, much like his active forefinger. Limits to his attempts at controlled behaviour are illustrated in examples (17a) and (17b) below (corresponding to concordance lines 15 and 13): Bucket finds it hard to avoid being overly friendly which is not only made clear

through the narrator's comment but also through Bucket's reddening – body language that he cannot control.

- (17) a. “Why you see, miss,” returns Mr. Bucket, bringing the finger into persuasive action – and such is his natural gallantry that he had almost said “my dear” – “it ain't easy to [...]”  
b. “[...] society, miss,” says Mr. Bucket, quite reddening at another narrow escape from “my dear.”

(Chapter 53)

Also his business-like manner appears not to fit the business he is in, as shown by example (18), corresponding to concordance line 6:

- (18) “Now, George,” continues Mr. Bucket, putting his hat upon the table with an air of business rather in the upholstery way than otherwise, “my [...]”  
(Chapter 49)

In contrast to inspector Bucket, the lawyer Tulkinghorn is portrayed as more reserved and calculating. The concordance in Figure 2 contains the occurrences of *Tulkinghorn* and *lawyer* in suspensions. Only those suspensions are listed that refer to Tulkinghorn's speech. I added *lawyer* to the search to obtain more examples. In line 1 Tulkinghorn is described as “a cautious man”, line 7 refers to “his methodical, subdued, uninterested way”, line 10 to “his jog- trot style”, line 17 to “his dry passionless manner”. The possessive *his* highlights these references to his manner as permanent features, thus examples such as line 11 “returns methodically” and 16 “with undisturbed calmness” fit into the picture of permanent features.

It is ironic that although Tulkinghorn appears to be cautious and systematic, he ends up being killed by Hortense, a minor character whom he thinks he can easily threaten. While Bucket in contrast appears less professional with his exaggerated gallantry or his business approach that resembles that of an upholsterer, he is successful in solving a murder case. Tulkinghorn's passionless manner not only relates to him being methodical, it also is a sign of him being aloof. Concordance lines 19 and 20 are given below in examples (19). Here Tulkinghorn threatens that he will have Hortense arrested if she comes to see him uninvited again. Although he apologizes for being “unpolite”, his aim is to threaten her and he takes no note of her reactions. The two suspensions highlight that he continues “without minding her”.

- (19) “In a word, mistress,” says Mr. Tulkinghorn, “I am sorry to be unpolite, but if you ever present yourself uninvited here – or there – again, I will give you over to the police. Their gallantry is great, but they carry troublesome people through the streets in an ignominious manner, strapping down on a board, my good wench.”

“I will prove you,” whispers mademoiselle, stretching out her hand, “I will try if you dare to do it!”

“And if,” *pursues the lawyer without minding her*, “I place you in that good condition of being locked up in jail, it will be some time before you find yourself at liberty again.”

“I will prove you,” repeats mademoiselle in her former whisper.

“And now,” *proceeds the lawyer, still without minding her*, “you had better go. Think twice before you come here again.”

(Chapter 42)

1	ceedings in a cause--cautious man Mr.	Tulkinghorn,	taking no more responsibility than neces
2	you say?" "I say I am afraid," says Mr.	Tulkinghorn,	who had risen hastily, "that Lady Dedloc
3	did." "There was one of them," says Mr.	Tulkinghorn,	carelessly feeling-- tight, unopenable o
4	it." "Hadn't you better see," says Mr.	Tulkinghorn	to Krook, "whether he had any papers tha
5	, my dear." "I supposed, sergeant," Mr.	Tulkinghorn	resumes as he leans on one side of his c
6	u can go. So you are the man," says Mr.	Tulkinghorn,	opening his door with the key, "in whose
7	es the air. "The question is," says Mr.	Tulkinghorn	in his methodical, subdued, uninterested
8	n--Mob. "It's the place, you know," Mr.	Tulkinghorn	goes on to say in the fast-increasing da
9	y. "And he was much assisted," says Mr.	Tulkinghorn	as a wind-up, "by his son." "By his son,
10	arable." "Go on!" "Therefore," says Mr.	Tulkinghorn,	pursuing his case in his jog- trot style
11	g more to be sald to-night?" "Why," Mr.	Tulkinghorn	returns methodically as he softly rubs h
12	wide open. "Now, let us see," says Mr.	Tulkinghorn,	tapping his chin with the key and lookin
13	dismissed by your lady, you know," Mr.	Tulkinghorn	observes, following her out upon the sta
14	r." "She seems after all," observes Mr.	Tulkinghorn,	loitering a little forward with his hand
15	concentrated. "This woman," thinks Mr.	Tulkinghorn,	standing on the hearth, again a dark obj
16	lly-- I was about to say," resumes the	lawyer	with undisturbed calmness, "that whether
17	have got it here, sir." "Sergeant," the	lawyer	proceeds in his dry passionless manner,
18	matters this much, mistress," says the	lawyer,	deliberately putting away his handkerchi
19	u dare to do it!" "And if," pursues the	lawyer	without minding her, "I place you in the
20	former whisper. "And now," proceeds the	lawyer,	still without minding her, "you had bett
21	nt, sir?" "Why, Lady Dedlock," says the	lawyer,	taking a chair at a little distance from

Figure 2. Concordance of *Tulkinghorn* in suspensions

While Tulkinghorn only threatens Hortense, it is in fact Bucket who arrests her. Bucket talks to Hortense “in a cool determined way” (cf. line 26 in Figure 1), but he still stresses that he wants to be polite if possible and he insists that she sits down on the sofa and makes herself comfortable.

The character information that is revealed through the study of suspensions complements the information revealed by body language clusters. While Bucket appears as more engaged, communicative and hands-on in the literal sense of the word, Tulkinghorn is portrayed as distanced and passionless. Buckets gets closer

to people and the movements of his forefinger accompany his speech and actions. Tulkinghorn, in contrast, is shown to have his hands in his pockets or behind him, which in the given contexts can be interpreted as showing distance.

#### 4. Challenges for corpus stylistics

Clusters resemble lexical bundles in that they are repeated sequences of words. However, the term ‘lexical bundle’ in the sense of Biber et al. (1999) is clearly defined through criteria of frequency and distribution. Biber et al. (1999: 993) want to exclude idiosyncracies of individual speakers or writers and therefore focus on bundles that appear across at least five different texts. Thus clusters such as *not to put too fine* that are part of Mr Snagsby’s favourite phrase and only occur in one text in Dickens and nowhere in 19C would be excluded. The cluster *his hands in his pockets* occurs in eight texts in 19C so meets the minimum requirement for distribution. For five-word lexical bundles Biber et al. (1999: 993) use a cut-off of five per million, which *his hands in his pockets* meets when we look at its frequency in a corpus made up of both the Dickens corpus and 19C. When we only consider the frequency of this cluster in a corpus consisting of 19C and *Bleak House* the cluster only just misses the five per million threshold. Put differently, *his hands in his pockets* is significantly more frequent (or a ‘key cluster’) in the Dickens corpus compared to 19C (cf. also Mahlberg 2007a).

The definition of lexical bundles uses cut-offs so that characteristics of a register or “general building blocks used frequently by many different speakers/writers within a register” (Biber 2006: 174, endnote 4) can be found. By collecting clusters based on the requirement that a cluster has to occur a minimum of five times in one text, clusters are found that might be lexical bundles, but they do not have to be; the clusters might be key clusters, but do not have to be. Also finding a cluster in a text does not automatically mean that the cluster is necessarily associated with one character. We have seen that of the five occurrences of *his hands in his pockets* only two refer to Tulkinghorn, but those two still add to the picture that is created of the character Tulkinghorn. Thus, I initially want to see the relevance of clusters in the following way: clusters are building blocks of fictional worlds in the sense that they reflect aspects of the fictional world – mainly through the description of characters, comments by the narrator and references to place and time. Clusters are different from lexical bundles reflecting discourse functions of a given register, because they may be less frequent or less general. However, the areas of meanings that the clusters refer to – reflected not by individual clusters but by cluster groups – are general aspects of the fictional world. While individual clusters in the group may not be frequent, a group such as ‘body language’ clusters refers to general aspects of the novel.

To quantitatively assess the significance of the findings, however, is not straightforward. Already the notion of normalization to compare the number of occurrences of body language expressions across different texts causes some

questions. Novels will normally contain different numbers of characters, individual characters differ in the space they take up in a novel, and the descriptions of characters may vary with regard to the level of detail that can be found. Some characters may speak more, while others are presented more through description. It is easy to see how this list could be continued. However, related to the quantitative information that corpus methods can provide is the fact that already the organization and display of textual patterns can contribute to discovering information that may be less clearly visible when simply reading a text. By looking at character names in suspensions, prominent places for character information are grouped together and thus highlight the information provided in these places.

Body language in literature seems to be an area where the study of clusters (and then more flexible patterns) and suspensions can make an important contribution. While Korte's (1997) study provides a very comprehensive overview of different types of body language in literature, surprisingly little research seems to have been done to study the forms and patterns in which body language is expressed in literary texts. If we want to follow Culpeper's (2001) approach, the impressions that readers form about body language are affected by the cues in the text. At the same time, they are affected by the readers' knowledge of people in the real world. In the real world, we typically perceive body language by looking at people, while in literary texts we encounter descriptions of body language. When readers relate the character information in the text to their knowledge of people, their views on what is 'normal' body language for a person, may be different from what is 'normal' body language for a specific character simply because of the linguistic forms in which the body language is presented. Also, the description of body language has different functions in literary texts than in non-literary texts. So non-literary texts may not be very useful to find out how 'normal' body language is typically described. A search for the cluster *his hands in his pocket* in the BNC returns only one hit and the occurrence is from a classroom discourse where the discussion seems to refer to *Oliver Twist*.

Findings about body language in literature can also reveal information about the body language of the time and links to other sources of information on body language might be useful, e.g. etiquette books or guides for actors. While this option is not available for 19th century data, linking findings in literature with the study of multi-modal corpora additionally suggests ways of investigating the relationship between the body language of real people and descriptions of body language. However, methods for the analysis of multi-modal corpora might contribute to studying links between novels and their screen adaptations to find out to what extent body language in the film corresponds to what is described in the book. Such approaches will need consideration in the further development of corpus stylistics.

Another promising link to be followed is between corpus stylistics and psycholinguistics. Corpus studies can reveal patterns and highlight places in the literary text that appear to be relevant for providing cues about body language and ultimately characters. However, to what extent readers are aware of such patterns

cannot be determined by looking at corpus data alone. It is often pointed out as one of the key achievements of corpus linguistics that corpus findings can reveal information about the language that language users are not aware of. In particular, language users do not seem to be very good at making frequency judgements. Further investigations as to how readers perceive repetitions of patterns may add to our understanding of why some body language of characters is perceived as habitual while other body language may go unnoticed but still contributes to the overall impression of a character.

Particular questions that seem to call for participant testing also relate to suspensions. Lambert (1981) claims that suspensions that are five or more words long appear to be “intrusive”. He admits that he does not have firm evidence as to why five is a crucial number here, but his hunch might be related to the observation in psychology that sequences of about five words is what most people will be able to store in their short term memory. Additionally, it is not only the length of the interruption but also the kind of information that affects the readers’ perception. In example (11) above, the suspension contains information that contextualizes Bucket’s speech. In Korte’s (1997) terminology, “opening his arms” is an *illustrator*, i.e. body language that functions to support the verbal information. Example (12), in contrast, contains a comment by the narrator that does not refer to body language at all, but gives an evaluation of Bucket’s behaviour. A question to investigate further is to what extent readers notice the difference between body language that is presented as such, or comments by the narrator that suggest underlying body language.

The examples discussed here are based on observations in Dickens and the study of other authors can contribute further questions. Overall, however, crucial challenges for corpus stylistics refer to the fact that quantitative data in literary texts might have to be treated differently to quantitative data in large corpora. While large amounts of data enable the testing of statistical significance, corpus stylistics emphasizes the need for detailed qualitative interpretation. In this process the investigation of clusters and suspensions provides a starting-point for identifying textual functions and links to literary criticism. As the fictional worlds created in literary texts depend on the way in which readers process textual cues, corpus stylistics faces a major challenge in relating textual patterns to cognitive processes. Patterns identified by corpus methods and the development of descriptive concepts in corpus stylistics can inform further psycholinguistic research. The developments that corpus stylistics will be able to make with regard to linking textual cues to the description of cognitive processes as well as the wider context of criticism will determine whether the fiction we are dealing with is in the text or in the approach.

## Notes

- 1 I would like to thank Catherine Smith for the XML mark-up of the texts and her help with providing a concordance tool to search the XML texts.

## References

- Biber, D. (2006), *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999), *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Busse, B. (2010), 'Recent trends in new historical stylistics', in: D. McIntyre and B. Busse (eds.), *Language and Style*. Basingstoke: Palgrave Macmillan. 32-54.
- Carter, R. (2004), *Language and Creativity: The Art of Common Talk*. London: Routledge.
- Carter, R. (2010), 'Methodologies for stylistic analysis: Practices and pedagogies', in: D. McIntyre and B. Busse (eds.), *Language and Style*. Basingstoke: Palgrave Macmillan. 55-68.
- Chapelle, C.A. (2011), *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.
- Culpeper, J. (2001), *Language and Characterisation: People in Plays and Other Texts*. Harlow: Pearson Education.
- Jeffries, L. and D. McIntyre (2010), *Stylistics*. Cambridge: Cambridge University Press.
- John, J. (2001), *Dickens's Villains: Melodrama, Character, Popular Culture*. Oxford: Oxford University Press.
- Korte, B. (1997), *Body Language in Literature*. Toronto: University of Toronto Press.
- Kucich, J. (1994), 'Dickens', in: J. Richetti (ed.), *The Columbia History of the British Novel*. New York, NY: Columbia University Press. 381-406.
- Lambert, M. (1981), *Dickens and the Suspended Quotation*. New Haven, CT: Yale University Press.
- Leech, G. and M. Short (2007), *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. 2nd edition. London: Pearson Education.
- Lindquist, H. (2009), *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.
- Mahlberg, M. (2007a), 'Clusters, key clusters and local textual functions in Dickens', *Corpora*, 2: 1-31.
- Mahlberg, M. (2007b), 'Corpus stylistics: Bridging the gap between linguistic and literary studies', in: M. Hoey, M. Mahlberg, M. Stubbs and W. Teubert (eds.), *Text, Discourse and Corpora: Theory and Analysis*. London: Continuum. 219-246.
- Mahlberg, M. (2007c), 'A corpus stylistic perspective on Dickens' *Great Expectations*', in: M. Lambrou and P. Stockwell (eds.), *Contemporary Stylistics*. London: Continuum. 19-31.
- Mahlberg, M. (2009), 'Corpus stylistics and the Pickwickian watering-pot', in: P. Baker (ed.), *Contemporary Corpus Linguistics*. London: Continuum. 47-63.

- Mahlberg, M. and C. Smith (2010), 'Corpus approaches to prose fiction: Civility and body language in *Pride and Prejudice*', in: D. McIntyre and B. Busse (eds.), *Language and Style*. Basingstoke: Palgrave Macmillan. 449-467.
- McIntyre, D. (2010), 'Dialogue and characterization in Quentin Tarantino's *Reservoir Dogs*: A corpus stylistic analysis', in: D. McIntyre and B. Busse (eds.), *Language and Style*. Basingstoke: Palgrave Macmillan. 162-182.
- McIntyre, D. and B. Busse (eds.) (2010), *Language and Style*. Basingstoke: Palgrave Macmillan.
- O'Keeffe, A. and M. McCarthy (eds.) (2010), *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge.