

## Phrases in literary contexts: patterns and distributions of suspensions in Dickens's novels

Mahlberg, Michaela; Smith, Catherine; Preston, Simon

DOI:

[10.1075/ijcl.18.1.05mah](https://doi.org/10.1075/ijcl.18.1.05mah)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Mahlberg, M, Smith, C & Preston, S 2013, 'Phrases in literary contexts: patterns and distributions of suspensions in Dickens's novels', *International Journal of Corpus Linguistics*, vol. 18, no. 1, pp. 35-56.  
<https://doi.org/10.1075/ijcl.18.1.05mah>

[Link to publication on Research at Birmingham portal](#)

### **Publisher Rights Statement:**

Eligibility for repository : checked 08/02/2016

### **General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### **Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Phrases in literary contexts

## Patterns and distributions of suspensions in Dickens's novels\*

Michaela Mahlberg<sup>1</sup>, Catherine Smith<sup>2</sup> and Simon Preston<sup>1</sup>

<sup>1</sup>University of Nottingham / <sup>2</sup>University of Birmingham

This paper addresses relations between lexico-grammatical patterns and texts. Our focus is on a specific linguistic unit, the 'suspended quotation' (or 'suspension'), which has received particular attention in Dickens studies. The suspended quotation refers to an interruption of a fictional character's speech by the narrator with a sequence of at least five words. We show how corpus linguistic methods can help to systematically study suspensions in a corpus of Dickens's novels: we investigate relationships between patterns of body language presentation and suspensions; we consider the distribution of suspensions across novels; and we illustrate how patterns in suspensions relate to meanings of reporting verbs. Overall, we argue that suspensions are discernible units that contribute to meaningful patterns in narrative prose.

**Keywords:** suspended quotation, suspension annotation, body language presentation, reporting verbs, Dickens

### 1. Introduction

There is an increasing number of studies that show how a range of corpus linguistic tools and concepts can usefully be employed to support the analysis of literary texts (cf. e.g. Semino & Short 2004, Stubbs 2005, McIntyre & Walker 2010, for an overview see Mahlberg 2013a). Such studies are now often referred to as work in a new field called 'corpus stylistics'. However, the application of computer-assisted methods for the study of literary texts is not new (cf. e.g. overviews in Schreibman et al. 2004, or the discussion in Biber 2011). The potential innovation that 'corpus stylistic' studies can bring about lies in the way in which quantitative data is interpreted to address theoretical questions. Stubbs

(2010: 21) points out that “[c]orpus linguists like nothing better than empirical findings supported by levels of statistical significance. But outside this narrow circle, people want to know how it all hangs together”. The contribution that corpus stylistic studies can make in this regard is to further our understanding of the linguistic units in literary texts and the effects these have on the way in which readers create meanings from texts.

In this paper, our focus is on a specific textual unit that has received particular attention in the study of Dickens’s novels: the ‘suspended quotation’. The suspended quotation, also referred to as ‘suspension’, was brought to the fore by Lambert (1981) who defines it as an interruption of the speech of a fictional character by the narrator, as in Example (1), where the suspension appears in italics:

- (1) “I am very glad indeed,” *said Mrs. Jellyby in an agreeable voice*, “to have the pleasure of receiving you. I have a great respect [...] (Bleak House)<sup>1</sup>

To date the suspended quotation has mainly been of interest in literary criticism and seems to be accepted as a typical feature of Dickens’s style (Newsom 2000: 556). Using a corpus of Dickens’s novels, we investigate characteristics of the suspended quotation as a linguistic unit that contributes to the creation of meaning in narrative fiction. The suspended quotation is a place in the novel that links speech presentation and narration. This link is reflected by specific lexico-grammatical patterns found in suspensions. In addition to illustrating literary phraseologies, the suspended quotation exemplifies relations between the functions of lexico-grammatical patterns and their textual contexts. It thus highlights that the study of phraseology has at least two important dimensions: the lexical one that begins with the identification of lexico-grammatical patterns and the textual one that takes into account the contexts in which phrases function.

With the help of corpus tools a range of co-occurrence patterns of words can be investigated and various descriptive concepts have been suggested to capture such patterns. Most notably, there are ‘lexical bundles’, defined as frequently recurring sequences of words such as *I don’t know* (cf. Biber et al. 1999), and ‘lexical items’, i.e. multi-word units having a specific meaning and consisting of a lexical core surrounded by patterns of words described as collocations, colligations, semantic preference and semantic prosody (cf. for instance the example of *true feelings* discussed by Sinclair 2004). In between the straightforward and fixed sequences of lexical bundles and the more elusive units accounted for by lexical items, various co-occurrence patterns of words have been investigated. Also specific tools have been developed for the retrieval of such patterns, for instance, the ConcGram Tool that is designed as a “phraseological search engine” (Greaves 2009).

Complementing the study of phraseology that begins from a lexical point of view, phrases have been investigated in relation to the types of texts they occur in. For instance, lexical bundles, or frequent clusters, as they are also called, vary in their distribution across registers (cf. e.g. Biber et al. 1999, Carter & McCarthy 2006). Relations between phrases and texts are also highlighted by Stubbs (2010), who argues that the notion of keyness can be extended from key words both in the cultural and the computational sense to key phrases. Following Francis (1993), he discusses “culturally significant units of meanings” (Stubbs 2010:28). He argues that ‘key phrases’ functioning as evaluative speech acts can extend the empirical basis of speech act theory to include, for instance, expressions of incomprehension or uncertainty. Among his examples referring to such frequent everyday sociolinguistic acts are the CANNOT FOR THE LIFE OF ME construction as in *I can't for the life of me understand what it is you see in it*, or the construction WHAT'S X DOING Y? expressing “unexpectedness” as in *what's she doing with a young man like that?* (Stubbs 2010:29). Stubbs (2010) emphasizes the relations between phraseology and texts, but also text-types and social institutions stressing how social conventions shape linguistic patterns. While his examples of culturally significant units of meanings largely seem to reflect conventions of spoken language, Stubbs (2010:35) also briefly mentions literary texts and emphasizes that they, too, are bound by conventions and social institutions including publishers and university courses.

In addition to links between phrases and the types of texts in which they occur, corpus studies have shown that the distribution of phrases varies across sections within texts (cf. Scott & Tribble 2006, Römer 2010, O'Donnell et al. 2012). In more theoretical terms, relationships between lexical and textual patterns have been captured by what Hoey (2004) calls ‘textual colligations’. Examples of such textual colligations in Hoey’s (2004) terminology are the tendency of a word to occur as the theme of sentences, or the tendency of a word to form part of cohesive lexical chains. While Hoey’s (2004) textual colligations focus on given concepts such as theme, or lexical cohesion, another concept to capture lexical and textual relations are ‘local textual functions’ (cf. Mahlberg 2005, 2007). Local textual functions describe the patterns of a (set of) lexical item(s) in a specific (set of) text(s). Thus the textual categories that can be used to describe local textual functions are less neatly identifiable than the potential range of textual colligations, but they will be specified in relation to the text(s) at hand. Examples of local textual functions include patterns of the noun *move* in newspaper story patterns, where it can be shown to have a tendency to appear in the second paragraph of the newspaper article (Mahlberg 2009), or patterns of body part nouns that contribute to the creation of characters in texts by Charles Dickens (Mahlberg

2013b). While the concept of local textual functions makes it possible to relate lexical patterns to a range of textual properties, it also highlights the need to identify and investigate textual properties that potentially relate to the behaviour of lexical items. Textual patterns that are of interest can be formal patterns based on paragraphs, sections, or chapters of texts and they depend on the type of text under investigation. Corpus studies of literary texts have investigated, for instance, differences in lexical distribution across the speech of characters (Culpeper 2009), or across different narrators in the same text (Hori 2004), but also links between key words and text segmentation (Fischer-Starcke 2010).

In this paper, we argue that the suspended quotation is a linguistic unit that contributes to meaningful patterns in fictional prose. We begin by outlining how corpus methods make it possible to study suspensions systematically (Section 2). A stylistic analysis is interested in the potential effects that textual patterns can trigger in readers. However, quantitatively significant or repeatedly occurring patterns are not necessarily directly related to a discernible effect. Therefore, based on empirical data from Dickens's novels, we approach properties of suspensions from three different perspectives: we show that suspensions are associated with patterns of body language presentation (Section 3), we scrutinize claims by Lambert (1981) about the distribution of suspensions across Dickens's novels (Section 4), and we deal with patterns of reporting verbs in suspensions (Section 5). By drawing on these three features of suspensions we illustrate how different aspects of the creation of fictional worlds hang together and how corpus linguistic methods can aid the analysis of patterns that potentially guide the readers' perception of the narrative text.

## 2. Studying suspensions in a corpus of Dickens's novels

We work with a corpus of texts downloaded from Project Gutenberg.<sup>2</sup> From the point of view of textual scholarship, the Project Gutenberg editions might be open to criticism (cf. also Mahlberg 2013b). However, the main advantage of these texts is that they are freely available – unlike scholarly editions. The corpus of Dickens's 15 novels amounts to nearly 3.9 million words. Table 1 lists the titles in the novels corpus in chronological order of first publication. We will return to the division into “early” and “late” in Section 4.

To be able to study suspensions systematically, they have to be annotated. Figure 1 shows Example (1) from above – this time with the annotation we use in our novels corpus. The annotation contains basic information, such as the title of the novel, chapter and number of paragraph or sentence (BH.c4.p24 indicates *Bleak House*, Chapter 4, paragraph 24, or BH.c4.s52 stands for *Bleak House*, Chapter 4, sentence 52). To identify suspensions, we first annotate text within

**Table 1.** Texts in the Dickens novels corpus in chronological order

PP	<i>The Pickwick Papers</i>	
OT	<i>Oliver Twist</i>	
NN	<i>Nicholas Nickleby</i>	
OCS	<i>The Old Curiosity Shop</i>	
BR	<i>Barnaby Rudge</i>	Early Dickens
MC	<i>Martin Chuzzlewit</i>	
DS	<i>Dombey and Son</i>	
DC	<i>David Copperfield</i>	
BH	<i>Bleak House</i>	
HT	<i>Hard Times</i>	
LD	<i>Little Dorrit</i>	
TTC	<i>A Tale of Two Cities</i>	Late Dickens
GE	<i>Great Expectations</i>	
OMF	<i>Our Mutual Friend</i>	
ED	<i>The Mystery of Edwin Drood</i>	

quotation marks with <qs/> (“start of quoted text”) and <qe/> tags (“end of quoted text”). Guided by Lambert (1981), we define suspensions as any text of five or more words which occur between a <qe/> tag and a <qs/> tag within a single paragraph (indicating no change of speaker). Such suspensions are then marked up with the tags <sls/> and <sle/> (see Figure 1, the “l” indicates “long” suspensions, i.e. five or more words, as we also mark up shorter suspensions that we use to address research questions beyond the present paper). The XML annotation

```

<p pid="24" id="BH.c4.p24" type="speech">
  <s sid="52" id="BH.c4.s52">
    <qs/>
      "I am very glad indeed,"
    <qe/>
    <sls/>
      said Mrs. Jellyby in an agreeable voice,
    <sle/>
    <qs/>
      "to have the pleasure of receiving you.
    </s>
    <s sid="53" id="BH.c4.s53">
      I have a great respect for Mr. Jarndyce, and no one in
      whom he is interested can be an object of indifference
      to me."
    <qe/>
    </s>
  </p>

```

**Figure 1.** Example (1) with annotation

was created with a set of Python scripts, described in more detail in Mahlberg & Smith (2012). At this stage, the annotation does not distinguish whether the text between quotation marks is quoted speech, thought or writing. While this is an important distinction in terms of discourse presentation analysis (cf. Semino & Short 2004, Busse 2010), for the points we want to address in this paper it is not crucial. Also, with regard to the links we will make to Lambert (1981), we are in line with his work that does not distinguish between speech, thought or writing (Lambert 1981: 154).

Once the definition of suspensions is operationalized, it becomes possible to investigate associations between suspensions – as places in the novels – with lexical phrases that have been identified independently of the places in which they occur. Section 3 will deal with distributions of patterns of body language presentation across different sections of the novels, i.e. within quoted text, text outside of quotes and text in suspensions. With automatically retrievable suspensions it is also possible to test claims about the effects of suspensions that seem to be related to their distributions in particular texts. Lambert's (1981) main interest is in how suspensions reflect authorial positions that differ between early and late Dickens. He finds support for his claims in manually generated suspension counts which can now be tested against a larger set of data (Section 4). Finally, by focusing on suspensions as a subset of the novels corpus we can identify patterns within suspensions, which we will illustrate in Section 5 with examples of reporting verbs. So far the suspended quotation has been studied mainly from a literary point of view. With the help of corpus methods it becomes possible to view the suspended quotation not only as an isolated stylistic device, but in the context of wider patterns within narrative texts. While this paper concentrates on a corpus of Dickens's novels, it is clear that the approach we illustrate is also applicable to texts by other authors (see also Mahlberg & Smith 2010) and our findings raise questions that deserve to be investigated more widely in fictional prose.

### 3. Suspensions and body language presentation

Suspensions are places in the text that seem to be particularly likely to contain information on characters' body language (cf. also Lambert 1981). By interrupting a character's speech, suspensions can create an impression of simultaneity between the speech and the contextual information described by the narrator, which in turn can suggest similarities to the simultaneous occurrence of speech and body language in real life. To investigate the relationship between character speech and body language, we begin with patterns of body language presentation and look at the distribution of these patterns across sections within the text.

Literary patterns of body language presentations have not received much attention as yet. Brown (1996) and Person (1999) illustrate how observations on paralinguistic features, including pitch span, tempo, and pausing, can be related to the analysis of literature. To our knowledge, the most comprehensive approach to non-verbal communication in literary texts is Korte's (1997) study of body language. Although body language is seen to play a role in the creation of characters in fiction (Culpeper 2001, Stockwell 2009), its linguistic presentation seems to be difficult to pin down. Korte (1997) points out that body language is not necessarily represented in the form of lexicalized expressions such as *frown* or *smile*, but can also be described in a variety of forms or may not even be explicitly described but only implied.

Beginning with the retrieval of five-word clusters, Mahlberg (2013b) suggests lexically-driven categories of body language. Patterns of body language presentation can be seen as 'key phrases' in narrative fiction, in the sense that body language plays an important role in the creation of characters (cf. Mahlberg 2013b). Table 2 includes a set of five-word clusters discussed in Mahlberg (2013b). The clusters were initially derived on the basis of a 4.5 million word Dickens corpus, but Table 2 focuses on their occurrence in the novels only. Mahlberg (2013b) compares different types of clusters derived from the Dickens corpus and identifies body part clusters as a group because all of them contain a body part noun, e.g. *head* in *head on one side and* or *eyes* in *his eyes on the ground*. The clusters can be shown to fulfil local textual functions in the creation of fictional characters. For instance, *with his hand to his* is used in *Our Mutual Friend* to indicate Twemlow's habit of putting his hand to his forehead. In addition to identifying characters through habitual behaviour, body language clusters can aid the authentication of a situation by creating a lively or life-like picture, as in Example (2) which illustrates the cluster *with his eyes on the*:

- (2) While he thus walked up and down *with his eyes on the ground*, [...] (DS)

Both Mr Dombey's walking up and down the room and his eye behaviour are external signs of his thought processes. In the present paper, our aim is not a detailed discussion of such functions (which can be found in Mahlberg 2013b); what we are interested in here are relations between clusters and specific places in the narrative text. The annotated corpus makes it possible to observe the distributions of the body language clusters across text inside quotes, outside of quotes and in suspensions. Table 2 shows raw figures and figures normalized per one million words. The total word counts in each section of the texts are provided under the table. Table 2 shows that the clusters tend not to occur in the reported direct speech of fictional characters. There are altogether six clusters with occurrences in direct speech (*his hand upon his shoulder*, *his head on one side*, *his head out of*



*the, the crown of his head, with his back to, with his hand to his*). In these cases, one character describes or points out other characters and thus the cluster is used in a similar way to the instance outside quoted speech – only not by the narrator, as in most of the cases outside quotes, but by a character. Additionally, the example of the cluster *the crown of his head* in speech is an illustration of the larger pattern *from the crown of his head to the sole of his foot* that can be used to add extra emphasis to a description:

- (3) “Haven’t you yourself declared that the fellow has heaped provocations, insults, and affronts on you, or something to that effect? He has done the same by me. He is made of venomous insults and affronts, *from the crown of his head to the sole of his foot*. [...] (OMF)

Table 2. Distribution of body language clusters

	Full-text		Quote		Non-quote		Suspensions	
	Freq.	pmw	Freq.	pmw	Freq.	pmw	Freq.	pmw
<i>head on one side and</i>	18	4.69	0	0	18	7.26	3	35.79
<i>her apron over her head</i>	6	1.56	0	0	6	2.42	0	0
<i>his eyes on the ground</i>	7	1.82	0	0	7	2.82	1	11.93
<i>his hand to his forehead</i>	16	4.17	0	0	16	6.45	0	0
<i>his hand upon his shoulder</i>	10	2.61	1	0.74	9	3.63	3	35.79
<i>his hands in his pockets</i>	77	20.07	0	0	77	31.05	9	107.37
<i>his head against the wall</i>	12	3.13	0	0	12	4.84	0	0
<i>his head on one side</i>	30	7.82	2	1.47	28	11.29	6	71.58
<i>his head out of the</i>	13	3.39	1	0.74	12	4.84	0	0
<i>his pipe in his mouth</i>	8	2.09	0	0	8	3.23	2	23.86
<i>laying his hand upon his</i>	20	5.21	0	0	20	8.06	7	83.51
<i>the crown of his head</i>	17	4.43	1	0.74	16	6.45	0	0
<i>with his back to the</i>	41	10.69	1	0.74	40	16.13	2	23.86
<i>with his eyes on the</i>	16	4.17	0	0	16	6.45	1	11.93
<i>with his hand to his</i>	30	7.82	2	1.47	28	11.29	1	11.93
<i>with his hands behind him</i>	18	4.69	0	0	18	7.26	3	35.79
<i>with his hand in his</i>	51	13.3	0	0	51	20.57	6	71.58
<i>with his head against the</i>	9	2.35	0	0	9	3.63	0	0

Note. Full text: 3,835,807 words, quotes: 1,356,236 words, non-quotes: 2,479,571 words, suspensions: 83,824 words.

The figures in Table 2 suggest an association between the body language clusters and suspensions: if the clusters occur in suspensions, their normalized frequencies are also highest in suspensions, i.e. particular subsections of the non-quote text.

The quantitative information in Table 2 is best treated as indicative and not as an exact measure. Overall, the table contains only relatively small numbers and suspensions make up the smallest subcorpus, which leads to inflated normalized figures. Additionally, for any corpus study based on small numbers mistakes in the annotation – which cannot be entirely avoided – may affect the figures.<sup>3</sup> The issue of limited quantitative data is one that corpus stylistic work often has to deal with especially when the focus is on one author or one text by one author. So it is even more important to gather complementary evidence and to find explanations for the textual patterns suggested by the data.

There is further support for the relationship between presentations of body language and suspensions both in quantitative terms and through functional explanations. In the body language clusters of Table 2, the body part nouns *head*, *eyes*, *hand*, and *hands* appear in more than one cluster each, reflecting that the nouns can appear in a range of body language descriptions – and not only in the fixed body language clusters. Table 3 shows that the distribution of the nouns is similar to the distribution of the clusters in that the nouns are also more frequent in text outside quotation marks. Similar to clusters, the nouns seem to be associated with suspensions: for all nouns, suspensions are the subsection of the corpus with the highest normalized frequencies.<sup>4</sup>

Table 3. Distribution of body part nouns occurring in more than one cluster

	Full-text		Quote		Non-quote		Suspensions	
	Freq.	pmw	Freq.	pmw	Freq.	pmw	Freq.	pmw
<i>head</i>	4873	1270.4	678	499.91	4195	1691.82	371	4425.94
<i>eyes</i>	4034	1051.67	458	337.7	3576	1442.18	277	3304.54
<i>hand</i>	5128	1336.88	697	513.92	4431	1787	441	5261.02
<i>hands</i>	2615	681.73	345	254.38	2270	915.48	189	2254.72

The tendencies indicated by the quantitative information complement observations on the functions of suspensions. In the real world, speech is typically accompanied by body language. In the narrative text, suspensions seem to be a useful place to suggest that speech and body language occur simultaneously. Suspensions often interrupt the fictional spoken language with information on the context in which the speech occurs and the lexico-grammatical patterns of suspensions support the presentation of information as contextual. A pattern that repeatedly occurs is repV *-ing* (“reporting verb followed by *-ing* clause”) as in the examples below. In (4), Dick walks up and down and has his hands in his pockets while he speaks. In (5), Anthony has just interrupted Pecksniff and he accompanies his speech now with a gesture that further underlines how he keeps Pecksniff

from speaking. In both examples, the body language clusters appear in non-finite subordinate clauses to the reporting clause; in Example (4), the cluster additionally is a prepositional phrase. Thus the grammatical patterns do not contribute to the progression of the story, but detail circumstantial information.

- (4) “Now,” said Dick, walking up and down *with his hands in his pockets*,  
 “I’d give something – if I had it – to know [...] (OCS)
- (5) [...] But it seems to me, my dear Pecksniff,” added Anthony, *laying his hand upon his sleeve*, “that if you and I kept up the [...] (MC)

The view of suspensions as accompanying information also relates to the zeroes in the suspension column in Table 2. There are six body language clusters that are not found in suspensions, four of them function to highlight information that is associated with specific fictional characters. The clusters refer to typical behaviour of characters, either by occurring in contexts in which they receive particular emphasis or through the fact that the clusters only occur with reference to a single character in the novel: *her apron over her head* refers to Mistress Affery in *Little Dorrit*, *his hand to his forehead* belongs to Twemlow in *Our Mutual Friend* and *his head against the wall / with his head against the* characterizes Mr Jellyby in *Bleak House*. Therefore an explanation for the distribution of these clusters as shown in Table 2 might be that they present body language that is designed to be prominent. So if they were to be associated with contextual information of speech, their effect might be reduced.

The clusters *his head out of the* and *the crown of his head* do not occur in suspensions, but they do not uniquely identify characters either. Still, they seem to be less suitable to appear in contextual information. The phrase *from the crown of his head to the sole of his foot* – as exemplified by Example (3) above – is used for emphasis, and even when the cluster *the crown of his head* does not appear as part of the longer phrase, it seems to be more emphatic than simply referring to a character’s head. In Example (6), the detail of the *crown of his head* gives further emphasis to Quilp’s comic behaviour, as it adds to the picture created by the expressions *politely* and *little* that contrast with Quilp’s very physical approach to conducting a conversation.

- (6) As it was plain that Sampson was bent on a complimentary harangue, unless he received a timely interruption, Mr Quilp politely tapped him on *the crown of his head* with the little saucepan, and requested that he would be so obliging as to hold his peace. (OCS)

The cluster *his head out of the* occurs eight of the 13 times in *The Pickwick Papers*, mostly, but not exclusively, with Mr. Pickwick. Instead of presenting a feature of a particular character, it reflects a theme of the novel. The Pickwickians travel the

country and bump into new situations and people. The fact that characters are repeatedly shown to pop their heads out of windows, gates, or coach windows is in line with the almost accidental development of the plot and the surprises the characters encounter.

Overall, the distribution of body language clusters and body part nouns that are repeatedly found in these clusters suggests that such patterns of body language presentation are associated with suspensions – as specific places in the text. The tendencies suggested by the quantitative information complement qualitative arguments about the functions of body language in narrative prose. As the body language patterns were identified independently of their occurrence in suspensions, this gives additional strength to the argument. The following section shows how corpus methods can add additional dimensions to the assessment of qualitative analyses.

#### 4. Suspensions across novels

To assess the potential effects that textual patterns may have on the reader, qualitative analyses normally focus on selected examples or text extracts. Ultimately, however, the text as a whole forms a unit of meaning where the cumulative picture of patterns across the text shapes the reader's perception. Therefore, in this section, we complement the local view of patterns of suspensions discussed in the previous section with a wider a perspective on the distribution of suspensions across Dickens's novels.

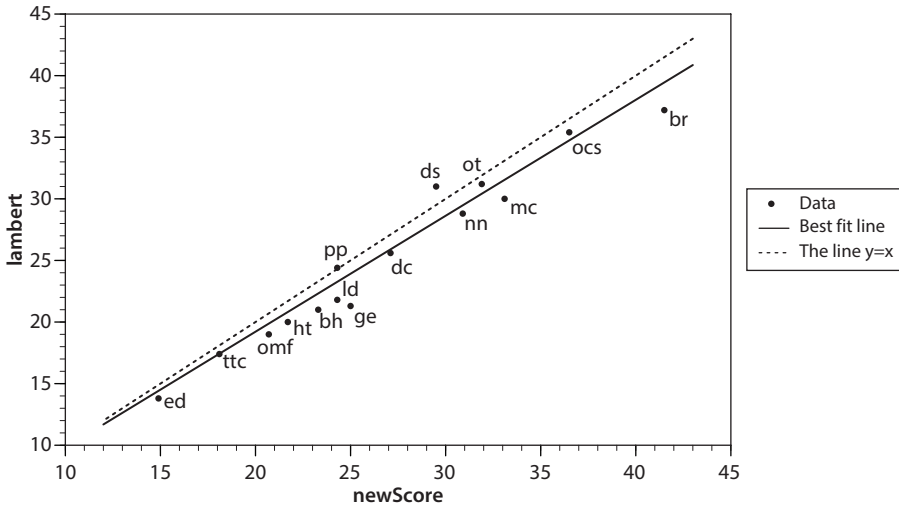
Lambert (1981) also analyses the suspended quotation across Dickens's novels and discusses its textual effects. He uses samples from each of Dickens's 15 novels to show that the number of suspensions decreases in Dickens's later work, compared to his earlier novels. Table 1 in Section 2 above shows where Lambert (1981) sees the dividing line between early and late Dickens. Lambert (1981) interprets suspensions as a reflection of the author's desire to interact with his audience. He suggests that Dickens was jealous of his characters and their speaking time. Therefore he interrupted them to make the author's presence felt more strongly. Lambert (1981) explains the decreasing number of suspensions in Dickens's later novels by the fact that Dickens found a new way of communicating with his audience through his public readings. So when the readings started, suspensions in the novels were less crucial to him. The rather provocative claims made by Lambert (1981) do not easily fall within the realm of corpus stylistics. Already the conflation of the terms 'author' and 'narrator' makes the line of argument problematic. However, in the course of his analysis, Lambert (1981) also makes observations on the usefulness of suspensions for the presentation of

body language and other paralinguistic information that support our arguments in Section 2. Therefore, in the present section, we aim to scrutinize the empirical basis on which Lambert (1981) based his claims about the distribution of suspensions. Although we might want to question Lambert's (1981) main interpretative claims, his quantitative findings still suggest that there is a perceivable effect created through the cumulative occurrence of suspensions found in a novel.

We build on initial work by Mahlberg & Smith (2012) that focuses on the rank order of Dickens's novels resulting from Lambert's (1981) findings on suspensions. Mahlberg & Smith (2012) claim that the rank order of the novels, when ordered by the number of suspensions they contain, changes when the counts are generated on the basis of full texts instead of text samples, as in Lambert's (1981) study. We now take this point further not only claiming that there is a difference in scores for suspensions in text samples versus full texts, but also examining the details of the differences. We argue that the basic trend of the distribution of suspensions across novels still holds as observed by Lambert (1981). However, by using re-sampling, we can show that there is evidence that Lambert's (1981) choice of samples is not random, but might be motivated by his literary evaluation of the texts.

Lambert (1981) manually counted the number of suspensions found in runs of 100 speech instances across the 15 novels, which is a rather small sample, considering that in 19th century narrative fiction direct speech is the most frequently occurring mode of discourse presentation (cf. also Busse 2010). As he provides detailed information on his selection of examples we can count suspensions in the same text samples. Instead of doing this manually, we make use of our suspension annotation. We refer to our automatically counted results for Lambert's (1981) samples as "newScore". Figure 2 shows a plot which illustrates Lambert's (1981) scores, referred to as "lambert", on the y-axis and our new score ("newScore") on the x-axis. Each point represents a Dickens novel. The solid line is the best-fitting straight line. If newScore were a perfect predictor of lambert we would have  $\text{lambert} = \text{newScore}$ , i.e. the points would lie on the dotted line shown on the plot. The deviation of the points from the dotted line reflects that there are several possible sources of errors, including for instance errors in the Project Gutenberg texts, mistakes in our annotation, but also human error in Lambert's (1981) manually counted suspensions or ambiguities in defining examples of suspensions.

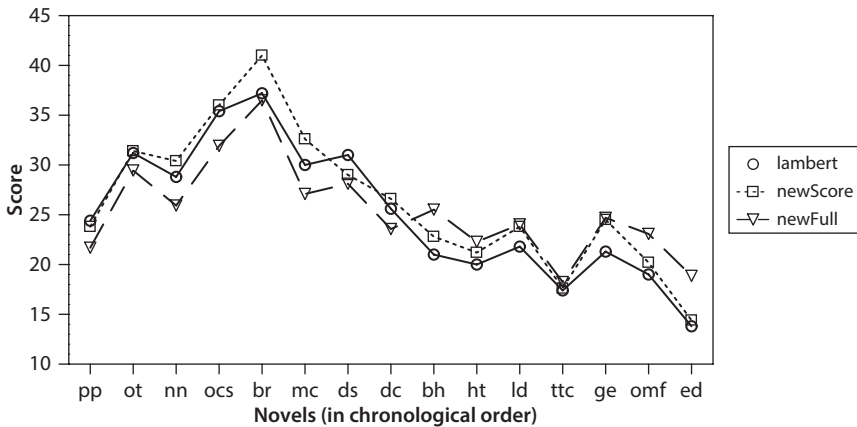
We can consider the model  $\text{lambert} = \text{newScore} + \text{err}$ , where *err* is a normally distributed error with mean zero and variance constant for each novel. An F-test reveals strong evidence ( $p$ -value = 0.02) against this model, which reflects that some of the possible errors mentioned above are probably systematic. Reasons for



**Figure 2.** Lambert's score versus "newScore", which is based on automatically counted results for Lambert's (1981) samples (dots); the best-fitting line of the form  $\text{lambert} = C1 + C2 \times \text{newScore}$  (solid); and the line  $\text{lambert} = \text{newScore}$  (dotted)

such errors might have to do with specific novels, which either contain more serious annotation errors than other books, or display particular stylistic features that might affect suspension counts. However, the gradient of the line of best fit is 0.94, which is not significantly different from 1 ( $p$ -value = 0.3) and very significantly different from 0 ( $p$ -value <  $10^{-10}$ ), so clearly newScore is a strong predictor of lambert.

If the novels are ranked by Lambert's (1981) score, then there is a division between the early novels (*Pickwick Papers* to *David Copperfield*, with scores ranging between 24 and 37 suspensions per 100 speech presentation instances) and the later novels (*Bleak House* to *Edwin Drood*, with scores between 14 and 22). The relationship between scores and book order is shown in Figure 3. This plot shows an apparent trend for Lambert's (1981) score to decrease in more recent works. The plot also shows two further lines. One for newScore (i.e. based on the same paragraphs selected by Lambert) and the other based on the full text, "newFull". The trend for newScore is similar to lambert, but for newFull, when the full text is considered, the trend that shows the number of suspensions decrease with later publication dates is less pronounced and the scores do not span as wide a range as in the other two cases.



**Figure 3.** Scores versus book order: (i) “lambert” (Spearman correlation coefficient,  $\rho = -0.793$ ,  $p$ -value  $< 0.001$ ), (ii) “newScore” ( $\rho = -0.729$ ,  $p$ -value  $\approx 0.002$ ), and (iii) “newFull” ( $\rho = -0.564$ ,  $p$ -value  $\approx 0.015$ )

Another, more quantitative, way to consider the relationship between score and novel order is to use a correlation coefficient. Since one of the variables is ordinal, it seems natural to use the Spearman correlation coefficient (Kendall 1970). All the following calculations are based on the Spearman coefficient, although we have checked that the conclusions are the same when we use either Pearson or Kendall correlation coefficients instead. The plot and the correlation coefficients show that the trend is less pronounced when comparing newFull with newScore. For each of the scores – lambert, newScore and newFull – plotted in Figure 3, we can use a permutation test to test the significance of a negative trend (i.e. the strength of evidence against the apparent trend having occurred by chance, called the ‘null hypothesis’,  $H_0$ , versus the alternative hypothesis,  $H_1$ , that there really is a negative trend). The  $p$ -values are shown in the caption. In summary, (i) using Lambert’s scores give extremely strong evidence against  $H_0$ ; (ii) using the new scores applied to Lambert’s choice of paragraphs gives less strong, but still extremely strong evidence against  $H_0$ ; (iii) using the new scores applied to the full text gives less strong but, still significant, evidence against  $H_0$ .

The less strong trend when considering the full text suggests the possibility that his literary appreciation of the novels may have led Lambert to select paragraphs in such a way that it emphasized the trend of decreasing numbers of suspensions in Dickens’s later novels. We can consider the null hypothesis  $H_0$ : Lambert chose his sections at random, versus the alternative  $H_1$ : Lambert chose specific paragraphs to emphasize a negative trend, and test for the strength of

evidence against  $H_0$ . Assuming that newScore is a good predictor of lambert (an assumption supported by Figure 2 and the tests described above) we can test  $H_0$  versus  $H_1$  by sampling under  $H_0$  – i.e. repeatedly drawing random samples of 500 paragraphs for each book and calculating the correlation coefficient. We can then see how “extreme” the correlation coefficient based on Lambert’s choice of paragraphs is, compared to what we would see by chance. Doing this gives a  $p$ -value of 0.015 under  $H_0$ , indicating strong evidence that Lambert selected his paragraphs non-randomly; in other words, that he chose his particular paragraphs to emphasize the trend.

In this section we have taken a more global view on suspensions to complement the detailed picture of Section 3. The quantitative information shows that differences in numbers of suspensions can be observed across novels. Because of the selectiveness of his data, Lambert (1981) suggests a stronger trend than the one shown with our methods. Still, there is empirical evidence that the number of suspensions decreases in Dickens’s later novels. From a qualitative point of view, it is important to investigate to what extent this fact is noticeable by the reader. Irrespective of the explanation that Lambert (1981) proposes for the existence of suspensions, his work suggests that suspensions create noticeable effects and his selection of text samples is likely to have been guided by the perception of such effects.

To more fully assess the trend of suspensions across novels it is also important to take other patterns into account that occur in the same texts. Further claims by Lambert (1981) that might be usefully investigated include the relationship between suspensions and reporting clauses that are shorter than five words. Additionally, the development of other more indirect forms of speech (thought, and writing) presentation might be relevant and the pattern for Dickens’s novels part of a more general trend across the 19th century where indirect forms become more frequent (cf. Busse 2010).

## 5. Suspensions and the reporting verbs *exclaimed* and *continued*

In this section, we focus on suspensions as interruptions of reported direct speech. From this point of view suspensions highlight another set of phrases with key functions in narrative prose: patterns of reporting verbs. While the suspended quotation as such is a concept that is mainly referred to in literary criticism, reporting verbs have also been dealt with in corpus stylistic studies (Semino & Short 2004, Busse 2010). Table 4 shows examples of reporting verbs that occur in suspensions. This set of verbs has been derived on the basis of a sample of ten suspensions per novel. The frequencies in Table 4 are totals across all suspensions in



the novels corpus.<sup>5</sup> The most frequent form, *said*, is generally frequent in fiction. Following the typology of reporting verbs by Caldas-Coulthard (1994: 305), *said* is a ‘neutral structuring’ verb that introduces “a ‘saying’ without explicitly evaluating it”. Other neutral verbs in the sample of Table 4 are *returned*, *replied*, *answered*, *asked*, *inquired* and *responded*. The verbs *added*, *pursued*, *resumed*, *continued*, *interposed* and *interrupted* are ‘discourse-signalling verbs’ (Caldas-Coulthard 1994: 306). They indicate how the quote relates to other parts of the discourse, like *added*, or they refer to the development of the discourse, e.g. *resumed*. A third group of verbs are ‘descriptive’ in the sense of marking the “manner” or “attitude” of a speaker “in relation to what is being said” (Caldas-Coulthard 1994: 306). This group includes *cried*, *exclaimed* and *whispered*. In addition to these three groups, Table 4 also contains the verb *remarked* that Caldas-Coulthard (1994: 306) subsumes under ‘metapositional verbs’ together with examples such as *urge*, *declare* and *grumble* that label “the contribution of the speaker”.

**Table 4.** Examples of reporting verbs in suspensions

Reporting verb	Frequency in suspensions
<i>said</i>	4177
<i>returned</i>	434
<i>cried</i>	364
<i>replied</i>	241
<i>added</i>	231
<i>pursued</i>	117
<i>resumed</i>	75
<i>continued</i>	66
<i>answered</i>	62
<i>exclaimed</i>	61
<i>whispered</i>	48
<i>asked</i>	42
<i>interposed</i>	33
<i>remarked</i>	30
<i>interrupted</i>	23
<i>inquired</i>	18
<i>responded</i>	4

One way to further study reporting verbs in suspensions would be to compare them with the verbs found in reporting clauses outside of suspensions and interpret the findings in view of more general distribution patterns of discourse presentations as they are accounted for in Semino & Short (2004) or Busse (2010). In the following, however, we want to concentrate on patterns of reporting verbs and their local textual functions in suspensions. We focus on an example from each

of the two main groups in Table 4 that are not 'neutral' according the typology by Caldas-Coulthard (1994): *exclaimed*, a descriptive verb that can indicate the manner/attitude of the speaker and *continued*, a discourse-signalling verb. Since we annotated suspensions in our corpus, it is now possible to retrieve concordances for reporting verbs in suspensions to support the identification of patterns.

Figure 4 shows a concordance for a selection of the examples of *exclaimed*. The data has been retrieved with the tool CLiC that was designed to enable searches across text in quotes, outside of quotes, and in suspensions.<sup>6</sup> The examples illustrate that suspensions can support the meaning of the verb by highlighting an exclamation separated from the remainder of the speech, e.g. *Lard!* (line 1), *My goodness me!* (line 4), *My good fellow!* (line 11). Exclamations can also contain a name, *Why, George* (line 21), *Why, my dear Paul* (line 24) or the vocative on its own may be separated off, *Mas'r Davy* (line 10), *Tilda* (line 16). Within the suspension the meaning of the verb *exclaimed* is enforced through details on the tone or the body language of a character. This information can show heightened emotion (e.g. *in a tone of great enjoyment*, line 1, *laughing and clapping her hands*, line 5, *exasperated beyond all bounds*, line 9, *with another contemptuous look*, line 14), but also indicate that the exclamation reflects surprise or sudden reactions (*being thrown by his surprise*, line 8, *opening her eyes wide*, line 18, *suddenly firing another volley*, line 20, *who had been evidently roused*, line 22, *halting very abruptly*, line 23).

#	Left	Node	Right
1	resigning it. 'It's a treat to me to stand.' 'Lard!'	<b>exclaimed</b>	Mr Boffin, in a tone of great enjoyment, as he
2	'Just another!' said Mr Toots. 'Go along with you!'	<b>exclaimed</b>	Susan, giving him a push 'Innocents like you, too!
3	nothing,' replied a man's voice. 'Dear me!'	<b>exclaimed</b>	the matron, in a much sweeter tone, 'is that Mr.
4	or anything of that kind? My goodness me!'	<b>exclaimed</b>	Mrs Nickleby, with a half-simper, 'suppose he was
5	the still pensive Mr Boffin. 'Lor-a-mussy!'	<b>exclaimed</b>	Mrs Boffin, laughing and clapping her hands, and
6	with a view, if possible, to its removal?' 'Ha!'	<b>exclaimed</b>	Mr. Jingle, with another start--'removal! remove
7	bonnet, as if preparing to be gone. 'Why, Nance!'	<b>exclaimed</b>	the Jew, starting back as he put down the candle,
8	her husband then?' 'Why, Lord bless my soul!'	<b>exclaimed</b>	Mr. Omer, after being thrown by his surprise into a
9	responded the girl, with a laugh. 'Change it!'	<b>exclaimed</b>	the Jew, exasperated beyond all bounds by his
10	tell me his name's Steerforth!' 'Mas'r Davy,'	<b>exclaimed</b>	Ham, in a broken voice, 'it ain't no fault of
11	cried Mark. 'What! Don't you?' 'My good fellow!'	<b>exclaimed</b>	Martin, clutching him by both arms, 'I have never
12	done it without her assistance.' 'Upon my word,'	<b>exclaimed</b>	Mr. Pickwick, who by this time had resumed his
13	prostration from drink. 'Ugh, you disgraceful boy!'	<b>exclaimed</b>	Miss Wren, attracted by the sound of his
14	Miss Sally, 'and don't talk so much.' 'Right again!'	<b>exclaimed</b>	Quilp, with another contemptuous look at
15	Joe. 'Dolly! 'Ay, call me that; call me that always,'	<b>exclaimed</b>	the locksmith's little daughter; 'never speak coldly
16	and she had thought so a long time. 'Tilda,'	<b>exclaimed</b>	Miss Squeers with dignity, 'I hate you.' 'Ah!
17	rarity and excellence. 'Sitch a rabbit pie, Bill,'	<b>exclaimed</b>	that young gentleman, disclosing to view a huge
18	very well ourselves.' 'Oh, what an accusation,'	<b>exclaimed</b>	Dora, opening her eyes wide; 'to say that you
19	most dimly. 'Would anybody believe, ma'am,'	<b>exclaimed</b>	Mrs. Raddle, turning fiercely to the first-floor
20	he keeps us," suggested Richard. "By my soul,"	<b>exclaimed</b>	Mr. Boythorn, suddenly firing another volley, "that
21	or not. It meets my approval." "Why, George,"	<b>exclaimed</b>	Mrs. Bagnet, who had been unpacking her basket,
22	then a hesitating knock at the door. 'Dear, dear,'	<b>exclaimed</b>	Mr. Pickwick, who had been evidently roused by
23	he be innocent? He looked like--Bye the bye,'	<b>exclaimed</b>	the old gentleman, halting very abruptly, and
24	I should designate him.' 'Why, my dear Paul!'	<b>exclaimed</b>	his sister, as he returned, 'you look quite pale!
25	and said I was a good fellow. 'Now, mind! she	<b>exclaimed,</b>	turning back on her way to the door, and looking

Figure 4. The reporting verb *exclaimed* in suspensions (25 of 61 examples)

The patterns of the meanings of *exclaimed* become more perceivable if we compare them to patterns of the verb *continued*. Figure 5 shows that the meaning of *continued* can entail a comparison with preceding speech: *in a lower voice* (line 7), *confident again by this time* (line 17), *even more gravely and impressively than he had spoken yet* (line 20). The basis for the comparison, however, does not have to be made explicit through comments by the narrator. Line 7, presented as extended example in (7), for instance, refers to *a lower voice* without providing details on the tone or voice in the preceding text. Instead, the fact that Pell lowers his voice relates to the content of his speech. The suspension follows a stretch of speech that introduces a topic shift: *Excuse my asking the question, Mr. Samuel*.

- (7) “Very good, very good,” said Pell, “you’re quite right, Mr. Weller. Mrs. Pell was a very elegant and accomplished woman; her manners were the theme of universal admiration in our neighbourhood. I was proud to see that woman dance; there was something so firm and dignified, and yet natural, in her motion. Her cutting, gentlemen, was simplicity itself. Ah! well, well! *Excuse my asking the question, Mr. Samuel*,” continued the attorney *in a lower voice*, “was your mother-in-law tall?” (PP)

Another pattern shown by the concordance of *continued* refers to temporal aspects of the speech when either the suspension qualifies the speed of speech, as in line 22 *continued Mr. Guppy rapidly*, or indicates the existence of pauses or silences as in lines 1, 2, 3 and 21. Example (8), which provides extended context for line 2, shows how a pause is only narrated retrospectively. The text suggests that *If we had come up with them this morning* follows a pause, but the reader only experiences a pause at the time of the suspension.

- (8) “You are there, are you, my friend?” he repeated, greedily biting his nails. “I am suspected and thrown aside, and Kit’s the confidential agent, is he? I shall have to dispose of him, I fear. *If we had come up with them this morning*,” *he continued, after a thoughtful pause*, “I was ready to prove a pretty good [...] (OCS)

The examples of *exclaimed* and *continued* show that patterns in the suspensions add detail to the meanings of the reporting verbs. Whether *exclaimed* describes the manner or the attitude of the speaker is indicated by information such as *opening her eyes wide* or *in a tone of great enjoyment*. The patterns also underline that a typology of reporting verbs can overemphasize the meaning of the verb in isolation. The verb *continued* is discourse-signalling in that it refers to the progress of the discourse, but the patterns around it can also emphasize the manner or attitude of the speaker in reaction to the preceding discourse. Patterns of the reporting verbs that are associated in particular with their occurrence in

suspensions, such as separating off exclamations or indicating pauses, show how the break that a suspension creates in the reported direct speech contributes to the presentation of the discourse and complements the meaning of the verb.

#	Left	Node	Right
1	yours to such an extent. I hope, Mrs Dombey,' he	<b>continued,</b>	after a few moments, during which, in his
2	If we had come up with them this morning,' he	<b>continued,</b>	after a thoughtful pause, 'I was ready to prove a
3	had he been upon the rack. 'There,' he	<b>continued,</b>	after a short silence, 'you have your message and
4	Carker. 'But why did you tell her? You see,' he	<b>continued</b>	with a smile, and softly laying his velvet hand, as
5	I have often wished to say this to you,' Tom	<b>continued</b>	with an embarrassment that became him very
6	and held it before her eyes. 'Clara,' he	<b>continued,</b>	looking at my mother, 'you surprise me! You
7	well! Excuse my asking the question, Mr. Samuel,'	<b>continued</b>	the attorney in a lower voice, 'was your mother-
8	Perhaps you may be interested, Mr. Jarndyce,"	<b>continued</b>	Mr. Bayham Badger, leading the way into the next
9	people right in their logic? 'Johnny, my pretty,'	<b>continued</b>	old Betty, caressing the child, and rather
10	if they had heard it. 'A porochial life, ma'am,	<b>continued</b>	Mr. Bumble, striking the table with his cane, 'is a
11	of the defendant. 'The plaintiff, gentlemen,'	<b>continued</b>	Serjeant Buzfuz, in a soft and melancholy voice,
12	have heard from my learned friend, gentlemen,'	<b>continued</b>	Serjeant Buzfuz, well knowing that, from the
13	his truth to me!" "He is not so sanguine, Ada,"	<b>continued</b>	Richard, casting his dejected look over the
14	a perspiration and do no good whatever. I say,	<b>continued</b>	Perker, checking off each position on a different
15	the expense of a wife, no more than nothing,'	<b>continued</b>	Mrs. Cluppins, with great volubility; 'why there
16	in him so to do,' said Sam. 'One of which,'	<b>continued</b>	the cobbler, 'he left to me, 'cause I married his
17	so constantly upon the scene of action. So what,"	<b>continued</b>	Richard, confident again by this time, "do I
18	neighbour, what's the matter? I swear to you,'	<b>continued</b>	the dwarf dismounting from the chair and sitting
19	g them, 'to restore a parent his child. Ay, sir,' he	<b>continued,</b>	bending eagerly forward, and addressing
20	Heaven forbid! Besides; how, Mr Tigg,'	<b>continued</b>	Pecksniff even more gravely and impressively than
21	you desire to speak to me before I go, sir,'	<b>continued</b>	that gentleman, after another pause, 'you may
22	time touching the matters now in question,"	<b>continued</b>	Mr. Guppy rapidly, as if he were repeating a
23	himself again—he did not pulverise him. 'Here,'	<b>continued</b>	the hardened traitor, tossing the licence at Mr.
24	room, and looking round. 'Shall I tell you,' she	<b>continued,</b>	with her eyes fixed on her mother, 'who already
25	dislike. 'The great advantage of this, Mr. Weller,'	<b>continued</b>	Job, his eyes filling with tears as he spoke, 'will

Figure 5. The reporting verb *continued* in suspensions (25 of 66 examples)

## Conclusions

In this paper we have approached suspensions as linguistic units from three different perspectives. Patterns of body language presentation and patterns of reporting verbs both indicate associations between key phrases in narrative fiction and suspensions as particular places in the narrative text. There are also relationships between patterns of reporting verbs and presentations of body language, especially if we consider descriptions of the tone and manner of speech along with body language. The examples of reporting verbs further show that typologies of such verbs provide a limited picture. The consideration of lexico-grammatical patterns within suspensions and across speech and suspensions is equally important.

The distribution of the suspended quotation across different texts provides a complementary perspective to the patterns associated with it locally. It seems that the occurrence of suspensions across a text creates a cumulative effect so that differences between texts become perceivable. Our findings raise further questions

in this regard: are effects created by suspensions mainly due to their presence as such, or are the types of suspensions (i.e. types of body language patterns, types of reporting patterns) even more important than their frequency? Additionally, the lengths of suspensions deserve further attention – so far we have only worked with the criterion that there have to be at least five words in a suspension. Similarly, a closer investigation of the lengths of the text that is interrupted is needed in future work.

Although this paper is based on data from Dickens's novels, it is clear that the concept of the suspension and the methods to analyse suspensions within and across texts are also applicable to texts by other authors.

## Notes

\* This paper is dedicated to Michael Stubbs who is inspirational, healthily critical and a good friend.

1. Throughout the paper all quotes are from e-texts. Therefore, we do not provide page numbers. For details on the Dickens novels corpus see Table 1, Section 2.
2. <http://www.gutenberg.org/> accessed September 2012.
3. Our precision and recall figures for the annotation of quote and non-quote text are: precision of 0.97 and a recall of 0.98 (Mahlberg & Smith 2012).
4. The significance of this association can also be shown by comparing the frequencies of the nouns in suspensions with their frequency in non-quote text that does not contain suspensions. Using a likelihood ratio test, for each noun the test statistic ('keyness value') is so large that it corresponds to a vanishingly small  $p$ -value.
5. For reporting verbs, we do not compare normalized frequencies across the subsections of the texts as in Tables 1 and 2, because a meaningful comparison for the verbs in Table 4 would be with their occurrences specifically in reporting clauses and not simply in quoted or non-quoted text.
6. CLiC (Corpus Linguistics in Cheshire) is currently a prototype that still needs further development and is not available online as yet. For more details on CLiC see Mahlberg & Smith (2012).

## References

- Biber, D. 2011. "Corpus linguistics and the scientific study of literature: Back to the future?". *Scientific Study of Literature*, 1 (1), 15–23.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Brown, G. 1996. *Listening to Spoken English*. 2nd ed. London: Longman.

- Busse, B. 2010. *Speech, Writing and Thought Presentation in a Corpus of Nineteenth-Century English Narrative Fiction*. Bern: University of Bern.
- Caldas-Coulthard, C. 1994. "On reporting reporting: The representation of speech in factual and factional narratives". In M. Coulthard (Ed.), *Advances in Written Text Analysis*. London: Routledge, 295–308.
- Carter, R. & McCarthy, M. 2006. *Cambridge Grammar of English: A Comprehensive Guide: Spoken and Written English Grammar and Usage*. Cambridge: Cambridge University Press.
- Culpeper, J. 2001. *Language and Characterisation. People in Plays and other Texts*. Harlow: Pearson Education.
- Culpeper, J. 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.
- Fischer-Starcke, B. 2010. *Corpus Linguistics in Literary Analysis: Jane Austen and her Contemporaries*. London: Continuum.
- Francis, G. 1993. "A corpus-driven approach to grammar: Principles, methods and examples". In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins, 137–156.
- Greaves, C. 2009. *ConcGram 1.0. A Phraseological Search Engine*. Amsterdam: John Benjamins.
- Hoey, M. 2004. "Lexical priming and the properties of text". In A. Partington, J. Morley & L. Haarman (Eds.), *Corpora and Discourse*. Bern: Peter Lang, 386–412.
- Hori, M. 2004. *Investigating Dickens' Style. A Collocational Analysis*. Basingstoke: Palgrave Macmillan.
- Kendall, M. G. 1970. *Rank Correlation Methods*. 4th ed. London: Griffin.
- Korte, B. 1997. *Body Language in Literature*. Toronto: University of Toronto Press.
- Lambert, M. 1981. *Dickens and the Suspended Quotation*. New Haven/ London: Yale University Press.
- Mahlberg, M. 2005. *English General Nouns: A Corpus Theoretical Approach*. Amsterdam: John Benjamins.
- Mahlberg, M. 2007. "Lexical items in discourse: Identifying local textual functions of sustainable development". In M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert, *Text, Discourse and Corpora. Theory and Analysis*. London: Continuum, 191–218.
- Mahlberg, M. 2009. "Local textual functions of move in newspaper story patterns". In U. Römer & R. Schulze (Eds.), *Exploring the Lexis-Grammar Interface*. Amsterdam: John Benjamins, 265–287.
- Mahlberg, M. 2013a. "Corpus analysis of literary texts". In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.
- Mahlberg, M. 2013b. *Corpus Stylistics and Dickens's Fiction*. New York/ London: Routledge.
- Mahlberg, M. & Smith, C. 2010. "Corpus approaches to prose fiction: Civility and body language in *Pride and Prejudice*". In D. McIntyre & B. Busse (Eds.), *Language and Style*. Basingstoke: Palgrave Macmillan, 449–467.
- Mahlberg, M. & Smith, C. 2012. "Dickens, the suspended quotation and the corpus". *Language and Literature*, 21 (1), 51–65.
- McIntyre, D. & Walker, B. 2010. "How can corpora be used to explore the language of poetry and drama?". In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, 516–530.
- Newsom, R. 2000. "Style of Dickens". In P. Schlicke (Ed.), *The Oxford Reader's Companion to Dickens*. Oxford: Oxford University Press, 553–557.

- O'Donnell, M. B., Scott, M., Mahlberg, M. & Hoey, M. 2012. "Exploring text-initial words, clusters and congrams in a newspaper corpus". *Corpus Linguistics and Linguistic Theory*, 8 (1), 73–101.
- Person, R. F. 1999. *Structure and Meaning in Conversation and Literature*. Lanham: University Press of America.
- Römer, U. 2010. "Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews". *English Text Construction*, 3 (1), 95–119.
- Schreibman, S., Siemens, R. & Unsworth, J. (Eds.) 2004. *A Companion to Digital Humanities*. Oxford: Blackwell.
- Scott, M. & Tribble, C. 2006. *Textual Patterns. Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.
- Semino, E. & Short, M. 2004. *Corpus Stylistics. Speech, Writing and Thought Presentation in a Corpus of English Writing*. London: Routledge.
- Sinclair, J. 2004. *Trust the Text. Language, Corpus and Discourse*. London: Routledge.
- Stockwell, P. 2009. *Texture. A Cognitive Aesthetics of Reading*. Edinburgh: Edinburgh University Press.
- Stubbs, M. 2005. "Conrad in the computer: Examples of quantitative stylistics methods". *Language and Literature*, 14 (1), 5–24.
- Stubbs, M. 2010. "Three concepts of keywords". In M. Scott & M. Bondi (Eds.), *Keyness in Texts*. Amsterdam: John Benjamins, 21–42.

#### *Authors' addresses*

Michaela Mahlberg  
School of English  
University of Nottingham  
University Park  
Nottingham  
NG7 2RD  
UK

michaela.mahlberg@nottingham.ac.uk

Simon Preston  
School of Mathematical Sciences  
The Mathematical Sciences Building  
University Park  
Nottingham  
NG7 2RD  
UK

simon.preston@nottingham.ac.uk

Catherine Smith  
Institute for Textual Scholarship and Electronic Editing  
Department of Theology and Religion  
ERI Building  
Pritchatts Road  
University of Birmingham  
Edgbaston  
Birmingham  
B15 2TT  
UK

c.j.smith@bham.ac.uk