

Dickens, the suspended quotation and the corpus

Mahlberg, Michaela; Smith, Catherine

DOI:

[10.1177/0963947011432058](https://doi.org/10.1177/0963947011432058)

License:

Creative Commons: Attribution-NonCommercial (CC BY-NC)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Mahlberg, M & Smith, C 2012, 'Dickens, the suspended quotation and the corpus', *Language and Literature*, vol. 21, no. 1, pp. 51-65. <https://doi.org/10.1177/0963947011432058>

[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 providing details and we will remove access to the work immediately and investigate.



Dickens, the suspended quotation and the corpus

Language and Literature

21(1) 51–65

© The Author(s) 2012

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0963947011432058

lal.sagepub.com



Michaela Mahlberg

University of Nottingham, UK

Catherine Smith

University of Birmingham, UK

Abstract

This article presents a computer-assisted approach to the study of character discourse in Dickens. It focuses on the concept of the ‘suspended quotation’ – the interruption of a character’s speech by at least five words of narrator text. After an outline of the concept of the suspended quotation as introduced by Lambert (1981), the article compares manually derived counts for suspensions in Dickens with automatically generated figures. This comparison shows how corpus methods can help to increase the scale at which the phenomenon is studied. It highlights that quantitative information for selected sections of a novel does not necessarily represent the patterns that are found across the whole text. The article also includes a qualitative analysis of suspensions. With the help of the new tool CLiC, it investigates interruptions of the speech of Mrs Sparsit in *Hard Times* and illustrates how suspensions can be useful places for the presentation of character information. CLiC is further used to find patterns of the word *pause* that provide insights into how suspensions contribute to the representation of pauses in character speech.

Keywords

Character discourse, corpus stylistics, Dickens, suspended quotation

1 Introduction

In this article we look at a linguistic device that has received attention as one of the techniques characteristic of Dickens’s style (see Newsom, 2000: 556). The ‘suspended quotation’, that is, an interruption of a character’s words by the narrator, has been extensively

Corresponding author:

Michaela Mahlberg, School of English Studies, The University of Nottingham, University Park Nottingham, NG7 2RD, Nottingham, NG7 2RD, UK.

Email: Michaela.Mahlberg@nottingham.ac.uk

discussed by Lambert (1981), who draws both on quantitative and qualitative methods. Although he uses his data to make some rather provocative claims that we do not want to follow up in our study, his observations also suggest interesting questions from a corpus stylistic point of view. The computer-assisted study of literary texts can make an important contribution to the study of linguistic devices by increasing the scale at which they are analyzed.¹ In this way, the study of the suspended quotation provides insights into the presentation of character information as well as the organization of character discourse. Although we raise some questions about the way in which Lambert approaches suspensions we also want to show that the concept is a useful one to include in the stylistician's checklist. The article begins with an outline of Lambert's (1981) argument in section 2. Section 3 introduces our approach to the quantification of suspensions in a corpus of Dickens's novels and compares our results with those by Lambert. Sections 4 and 5 then focus on detailed accounts of functions of suspensions, while at the same time illustrating how the new corpus tool CLiC can support stylistic analyses by searching specified sections of texts. Section 4 deals with the presentation of character information in suspensions illustrated with the example of Mrs Sparsit in *Hard Times*. Section 5 focuses on the presentation of pauses in character speech, before we conclude the article with more general observations.

2 The suspended quotation

Lambert (1981: 6) defines the 'suspended quotation' as a 'protracted interruption by the narrator of a character's speech. And here, "protracted" means containing at least five words.' Lambert (1981: 6ff.) explains that the criterion of five words is based on an intuitive judgement: an interruption of five words seems to be intrusive while a shorter one mostly does not. One of his examples is the following with the suspended quotation in italics.

- (1) 'I am proud to see,' *said Mr. Carker, with a servile stooping of his neck, which the revelations making by his eyes and teeth proclaim to be a lie, 'I am proud to see that my humble offering is graced by Mrs. Dombey's hand ...'*

(*Dombey and Son*)²

Lambert (1981: 41) points out that the suspension is 'a handy place to put information, gestures, facial contortions' and other details that he subsumes under the heading of 'suprasegmentals'. He points out that suprasegmentals play an important role for the representation of life-like dialogue, but are difficult to capture because of the linearity of the text (Lambert, 1981: 42). Suspensions thus are useful places to create an impression that comes close to synchronicity of presentation. Korte (1997: 97) makes a similar observation when she discusses ways in which body language accompanies fictional speech. Korte (1997: 97) points out that the two devices that are most frequently used to suggest that body language and speech occur simultaneously are '(a) the interruption of the character's speech by a description of the body language, and (b) the syntactical subordination of the body language to the character's speech'. While Lambert (1981) acknowledges that the description of suprasegmentals is one of the functions of suspensions, he sees the main reason for the frequency of the phenomenon elsewhere. In his

discussion of example (1), he points out that the information in the suspension is 'distinctly unflattering', which he takes as evidence for his provocative explanation of suspensions in terms of Dickens's aggressiveness towards his characters (Lambert, 1981: 59). Lambert (1981: 35) claims that suspended quotations in Dickens's early novels occur more frequently than in his late novels, because 'the heavy Dickensian use of suspended quotation seems to be fundamentally a sort of aggression' – even if the content of the suspensions is not overtly aggressive. He views the suspended quotation as a device employed by a jealous author who expresses hostility towards the characters he created. As this hostility stems from Dickens's resentment of the characters' attractiveness to the audience, Lambert claims, the use of suspended quotations decreases in Dickens's later novels. For him the division into early and late novels is between *David Copperfield* and *Bleak House*. He argues that with the start of the public readings in 1853, the year *Bleak House* was completed, Dickens found another way of seeking contact with his audience. Thus the indirect audience contact that was enabled through the suspended quotation became less relevant (Lambert, 1981: 140).

The claims that Lambert makes about Dickens's intentions when using suspended quotations cannot be tested by a linguistic analysis of Dickens's novels. However, claims about frequencies and functions of suspensions can be scrutinized in more detail. A methodological point that is crucial to Lambert's (1981) approach is the deliberate focus on striking cases when aiming to explain the prevalence of suspensions. He suggests: 'when trying to make sense of an uncommon pattern of usage, we may begin with those examples of the pattern which are the strangest of the strange and hope those oddest cases and most bizarre subvarieties will suggest something about the less startling examples' (Lambert, 1981: 44). What Lambert spells out is a way of approaching linguistic examples that is not uncommon but usually not made explicit. However, developments in corpus linguistics raise questions about what is to be regarded as 'uncommon patterns of usage', because it is the common patterns that are far less noticeable than the uncommon ones. This point also comes into play in Carter's (2004) approach to the concept of literariness. Instead of making a distinction between literary and non-literary language, Carter suggests a 'cline of literariness', which makes it possible amongst other things to account for literary language in everyday conversations. The notion of the cline plays an important role in the description of linguistic patterns, especially when those patterns are observed on the basis of corpus data. Sinclair's (1991: 100) observation that '[t]he language looks rather different when you look at a lot of it at once' still holds true for the analysis of literary texts. In sections 4 and 5, we will show that functions of suspensions can be analyzed from a different angle than the one taken by Lambert (1981) when a number of them are viewed together in the form of a concordance. Although Lambert did not have this option, he grants that not all suspensions can be clearly seen as cases of aggression when viewed in isolation. To strengthen his argument, Lambert draws on the overall frequency of suspensions in Dickens's novels to provide a context for his interpretative claims. In section 3, however, we aim to show that this context also needs to be treated with caution now that corpus methods can assist us to quantify the occurrence of suspensions.

Lambert's approach to quantifying suspensions in Dickens was to manually count the suspensions found in a number of runs of 100 speech instances. Five or six speech

Table 1. Average counts per novel from Lambert's (1981) study.

Novel	Lambert's average
Barnaby Rudge	37.2
Old Curiosity Shop	35.4
Oliver Twist	31.2
Dombey and Son	31
Martin Chuzzlewit	30
Nicholas Nickleby	28.8
David Copperfield	25.6
Pickwick Papers	24.4
Little Dorrit	21.8
Great Expectations	21.3
Bleak House	21
Hard Times	20
Our Mutual Friend	19
Tale of Two Cities	17.4
Edwin Drood	13.8

sections were selected from each book so that an even number of sections was studied for the novels that he divided into early and late Dickens. The counts for each section were recorded and the mean of the counts for each book was taken as the score for that book. The study of suspensions was only the first of two studies contributing to Lambert's concept of an 'officiousness index'. His second study counted the number of untagged quotations in the same sections used for the first study. The officiousness index was calculated by subtracting the average number of untagged quotations per book from the average number of suspensions in each book. This resulted in all of the early novels having a positive officiousness index and the late novels having a negative officiousness index. If we focus just on the results of Lambert's (1981) first study and rank the novels purely on those figures then the early books still come at the top (shaded in grey) and the later books at the bottom (see Table 1).

3 Automatic retrieval of suspensions

While Lambert's (1981) approach is based on manual counting, corpus linguistic methods enable us to build on his study by scaling it up to look at the full texts of the novels rather than just a few sections from each book. In order to be able to count all of the suspensions in a text they need to be annotated with XML. The text files for the corpus are taken from project Gutenberg³ and the annotation process for our corpus is fully automated using a set of Python scripts. Key to the automated annotation is the use of regular expressions. These are essentially complex patterns of characters that specify certain features of the text and can be used to find matching strings in the text, in this case they are used to find sentences and text in quotes. In a first pass through the text files, chapters and paragraphs are identified using the graphological conventions present in the base text. Sentences are

```

<p type="speech" id="BH.c6.p114">
  <s id="BH.c6.s340">
    <q</q>
      "My dear Miss Summerson,"
    <qe/>
    said Richard in a whisper,
    <q</q>
      "I have ten pounds that I received from Mr. Kenge.
  </s>
  <s id="BH.c6.s341">
    I must try what that will do."
  <qe/>
</s>
</p>

```

Figure 1. Example of annotation.

then annotated using a sentence tokenizer,⁴ which is based on regular expressions. In a third pass regular expressions are used again to identify passages of text between quotation marks. These sections are marked as quotes using milestone elements, which are empty XML elements containing no text, as they overlap with the sentence hierarchy. An example of the resulting XML is given in Figure 1: <q</q> is the milestone element for the start of a quotation and <qe/> the milestone element for the end of a quotation.

The suspension annotation is based on the XML structure shown in the figure. We take as our definition of suspensions any text of five or more words which occurs between a <qe/> tag and a <q</q> tag within a single paragraph. Based on this definition another pass through the corpus annotates the suspensions with more milestone elements. It is important to point out that at this stage, the annotation does not distinguish whether the text between quotation marks is speech or thought, but this is not crucial to points we aim to make in this article and in line with Lambert's (1981: 154) approach.⁵

As a way of determining how good our annotation is we can use 'precision and recall' figures. Our suspension tagging is based entirely on our quotation tagging. Therefore the quotation tagging is key in determining the accuracy of our suspension tagging and so is used to calculate precision and recall. In this case, precision is a measure of how many text sections we tagged as being quotes were actually quotes. Recall is a measure of how many quotes we tagged compared to the number of quotes that actually occur in the text. For our calculation, we took random samples of our text and manually corrected the annotation to give us a set of accurately annotated samples. Our automated annotation was then compared to the accurate samples and the precision and recall calculated as shown below. The calculations produce a number between 0 and 1. The closer to 1 the better the precision and/or recall. The figures for our quotation tagging are very good with precision of 0.97 and a recall of 0.98 so we can be confident that our annotation is accurate enough to form the basis for further study.

Precision = Correctly Tagged Quotes/All Quotes Tagged

Recall = Correctly Tagged Quotes/All Quotes

Table 2. Our ranks and averages for suspensions compared to Lambert's (1981).

Lambert's figures		Our figures	
Average	Novel	Novel	Average
37.2	Barnaby Rudge	Barnaby Rudge	41
35.4	Old Curiosity Shop	Old Curiosity Shop	36
31.2	Oliver Twist	Martin Chuzzlewit	32.6
31	Dombey and Son	Oliver Twist	31.4
30	Martin Chuzzlewit	Nicholas Nickleby	30.4
28.8	Nicholas Nickleby	Dombey and Son	29
25.6	David Copperfield	David Copperfield	26.6
24.4	Pickwick Papers	Great Expectations	24.5
21.8	Little Dorrit	Pickwick Papers	23.8
21.3	Great Expectations	Little Dorrit	23.8
21	Bleak House	Bleak House	22.8
20	Hard Times	Hard Times	21.2
19	Our Mutual Friend	Our Mutual Friend	20.2
17.4	Tale of Two Cities	Tale of Two Cities	17.6
13.8	Edwin Drood	Edwin Drood	14.4

The first step in extending Lambert's study to the full text of the novels is to try to recreate the counts for his sections and see how closely our automated annotation and counting matches Lambert's manual counts for his selected sections. Lambert (1981) explains in detail where each of his sections begin, so this aspect is easy to implement. However he does not define precisely what he considers as a 'speech instance'. Therefore, we take a speech instance to be any paragraph that contains speech. To recreate Lambert's study we take the first 100 paragraphs that contain speech from each of the starting points selected by Lambert. As with Lambert's study, we count the number of suspensions in each set of 100 paragraphs and take the mean for the sections in each book as the average. Our figures compared to Lambert's can be seen in Table 2. Although the figures are not exactly the same there is very little difference in the overall ranking of the books⁶ and the early-late distinction is largely maintained.

As Table 2 shows, our figures are not identical to those of Lambert (1981). On the whole our figures are slightly higher but the difference between them is not consistent. This aspect of the research needs further investigation to see if there are any underlying inaccuracies in the data that causes this difference in average counts. Unfortunately, Lambert does not publish the results by section but only provides the average count figures, which makes this kind of investigation more complicated to carry out. The fact that the two groups of early and late novels are on the whole maintained with our counts gives us enough confidence in the process to extend the analysis to the full texts and see what, if any, difference this makes to the averages and the ranks.

Having established that our recreation of Lambert (1981) is reasonably accurate we can expand the same technique to look at the full texts to see if the sections selected by

Table 3. A comparison between our counts for Lambert's (1981) selected sections compared to a full text count.

Our version of Lambert sections		Including full text	
Average	Novel	Novel	Average
41	Barnaby Rudge	Barnaby Rudge	36.5
36	Old Curiosity Shop	Old Curiosity Shop	31.9
32.6	Martin Chuzzlewit	Oliver Twist	29.4
31.4	Oliver Twist	Dombey and Son	28.1
30.4	Nicholas Nickleby	Martin Chuzzlewit	27.1
29	Dombey and Son	Nicholas Nickleby	25.9
26.6	David Copperfield	Bleak House	25.5
24.5	Great Expectations	Great Expectations	24.7
23.8	Pickwick Papers	Little Dorrit	24
23.8	Little Dorrit	David Copperfield	23.6
22.8	Bleak House	Our Mutual Friend	23.1
21.2	Hard Times	Hard Times	22.3
20.2	Our Mutual Friend	Pickwick Papers	21.7
17.6	Tale of Two Cities	Edwin Drood	18.9
14.4	Edwin Drood	Tale of Two Cities	18.2

Lambert are representative of the books as a whole. In order to do this, the calculation needs to be changed slightly. Rather than count the suspensions in 100 speech instances, we count all of them and work out the percentage of speech instances containing a suspension using the following formula. Thus our figures are directly comparable with Lambert's.

$$\frac{\text{Total number of suspensions}}{\text{total number speech paragraphs}} \times 100$$

Table 3 contains a comparison between our counts for Lambert's selected sections compared to the counts from the full text of each novel. As the table shows, the distinction between the early and late novels is eroded to a greater extent than we have seen before. However, the core of each section seems to remain intact despite the shifts in ranks present here.⁷ The two books which seem to fall furthest outside of their respective section are *David Copperfield* and *Pickwick Papers*.

We will need to extend the study to include the untagged quotations before further conclusions can be drawn as to the issues that Lambert raises with his concept of the 'officialness index'. However, the present study already shows that the novels are perhaps less homogeneous with respect to their use of suspensions than Lambert's (1981) results suggest. There are several differences in figures and ranks when our counts focusing entirely on his selected sections are compared to our counts across the full texts. It is also important to investigate the reasons behind the differences in order to ensure that our automated process is not introducing a greater error margin in particular books compared to others because of special features in the book or the quality of the underlying text taken from project Gutenberg.

Further reasons for differences in ranks between sample and full text counts could be due to the sections selected by Lambert. A possible explanation for this is provided by Lambert himself when he suggests that suspensions are used by Dickens to interrupt the speech of a particular type of character (see also section 4). In addition to being able to carry out more expansive quantitative studies, our annotated corpus also allows us the scope to conduct in-depth qualitative studies, which can be used to investigate the link between suspensions and characterization as well as other features of character discourse.

4 Studying suspensions with CLiC – the case of Mrs Sparsit

In order to investigate our XML annotated corpus, the web interface CLiC (Corpus Linguistics in Cheshire) has been developed. CLiC has been created on top of Cheshire 3, an open source search retrieval engine for XML data (cheshire3.org, n.d.). It allows searches to be carried out in different sections of the corpus. In particular, CLiC makes it possible to search for words and phrases in suspensions – as illustrated in Figure 1 with a search for the name *Sparsit*. The example of Mrs Sparsit in *Hard Times* serves to illustrate some of the characteristics of suspensions and the opportunities that CLiC offers.⁸

The screenshot shows the CLiC interface for the Dickens Corpus. The search term 'Sparsit' is entered in the search bar. The search scope is set to 'suspensions (5 word +)'. The results table is as follows:

#	Left	Node	Right
1	'A singular world, I would say, sir,' pursued Mrs.	Sparsit;	after acknowledging the compliment with a
2	I believe.' 'Don't say that, sir,' returned Mrs.	Sparsit,	almost with severity, 'because that is very unkind
3	'Your foot on the last step, my lady,' said Mrs.	Sparsit,	apostrophizing the descending figure, with the aid
4	and claims to reverence. 'The clerks,' said Mrs.	Sparsit,	carefully brushing an imperceptible crumb of
5	for past favours. And I hope, sir,' said Mrs.	Sparsit,	concluding in an impressively compassionate
6	pursued the stranger, fluttering over Mrs.	Sparsit's	eyebrows, with a propitiatory air, 'but you know
7	way!' 'I wish with all my heart, sir,' said Mrs.	Sparsit,	in a highly superior manner; somehow she
8	I don't like his ways at all.' 'Bitzer,' said Mrs.	Sparsit,	in a very impressive manner, 'do you recollect my
9	to my poor judgment - ' 'Oh! Pray, sir,' Mrs.	Sparsit	interposed, with sprightly cheerfulness, 'don't
10	another.' 'It is much to be regretted,' said Mrs.	Sparsit,	making her nose more Roman and her eyebrows
11	'All is shut up, ma'am.' 'And what,' said Mrs.	Sparsit,	pouring out her tea, 'is the news of the day?
12	not hurried. 'My dear child,' said Harthouse; Mrs.	Sparsit	saw with delight that his arm embraced her; 'will

Figure 2. Screen shot of CLiC for search of *Sparsit* in suspensions.

1 world, I would say, sir,' pursued Mrs. Sparsit; after acknowledging the compliment wit
 2 . 'Don't say that, sir,' returned Mrs. Sparsit, almost with severity, 'because that is
 3 on the last step, my lady,' said Mrs. Sparsit, apostrophizing the descending figure,
 4 to reverence. 'The clerks,' said Mrs. Sparsit, carefully brushing an imperceptible
 5 st favours. And I hope, sir,' said Mrs. Sparsit, concluding in an impressively compassi
 6 sued the stranger, fluttering over Mrs. Sparsit's eyebrows, with a propitiatory air,
 7 wish with all my heart, sir,' said Mrs. Sparsit, in a highly superior manner; somehow s
 8 e his ways at all.' 'Bitzer,' said Mrs. Sparsit, in a very impressive manner, 'do you r
 9 poor judgment - ' 'Oh! Pray, sir,' Mrs. Sparsit interposed, with sprightly cheerfulness,
 10 It is much to be regretted,' said Mrs. Sparsit, making her nose more Roman and her
 11 shut up, ma'am.' 'And what,' said Mrs. Sparsit, pouring out her tea, 'is the news of
 12 'My dear child,' said Harthouse; Mrs. Sparsit saw with delight that his arm embraced
 13 Gradgrind's daughter?' 'Yes,' said Mrs. Sparsit, suddenly compressing her mouth, 'he ha
 14 ss all the morning. 'Bitzer,' said Mrs. Sparsit that afternoon, when her patron was
 15 dence! Come out, ma'am!' then said Mrs. Sparsit, to some one inside, 'come out, or
 16 , 'but it is my duty to remember,' Mrs. Sparsit was fond of observing with a lofty
 17 you.' 'I certainly, sir,' returned Mrs. Sparsit, with a dignity serenely mournful, 'was
 18 luxury.' 'I do not, sir,' returned Mrs. Sparsit with a shake of her head, 'deny it.' Mr.
 19 urables.' 'I trust, sir,' rejoined Mrs. Sparsit, with decent resignation, 'it is not
 20 consider, I cannot consider,' said Mrs. Sparsit, with a most extensive stock on hand of
 21 to understand these things,' said Mrs. Sparsit, with dignity, 'my lot having been
 22 heard in Dutch clocks. Not,' said Mrs. Sparsit, with a lofty sense of giving strict
 23 ur feelings?' 'Yes, sir,' returned Mrs. Sparsit, with a meek shake of her head, 'he

Figure 3. The 23 occurrences of *Sparsit* in suspensions in *Hard Times*.

There are 23 suspensions in the Dickens Corpus that contain the name *Sparsit*. All of them come from the novel *Hard Times*. As the screen shot in Figure 2 only shows part of the examples, Figure 3 focuses just on the concordance with all 23 suspensions. Of those 23 examples, 21 interrupt the speech of Mrs Sparsit and the suspensions contribute to revealing character information. Among the repeated patterns visible in the concordance are non-finite clauses and prepositional phrases that contain circumstantial information, as in examples (2) and (3) which correspond to concordance lines 11 and 17 respectively. Both *-ing* clauses and prepositional phrases support the effect of synchronicity between the speech and the body language in those examples. In example (2), Mrs Sparsit accompanies her words with practical action body language: she pours out her tea. In example (3), there is no explicit description of body language, but the narrator provides an interpretation of Mrs Sparsit's body language: she speaks 'with a dignity serenely mournful'. In other examples where the narrator interprets body language, reference is also made to Mrs Sparsit's manner of saying something (the noun *manner* appears in concordance lines 7 and 8, as well as 5, which is not visible in the amount of context included in Figure 2 though).

- (2) 'And what,' said Mrs. Sparsit, pouring out her tea, 'is the news of the day? Anything?'
- (3) 'I certainly, sir,' returned Mrs. Sparsit, with a dignity serenely mournful, 'was familiar with the Italian Opera at a very early age.'

The picture that is presented through the suspensions in Figure 3 depicts Mrs Sparsit as having a lofty way of behaving. The adjective *lofty* occurs twice (lines 16 and 22), but

there are also examples referring to her impressive or superior manner (lines 5, 7, 8) and the dignity and severity with which she behaves (lines 2, 17, 19, 21). All these examples reflect that Mrs Sparsit acts as if she were better than other people. She now is Bounderby's housekeeper, but she 'had not only seen different days, but was highly connected' (*Hard Times*, Chapter 7). When she is introduced in Chapter 7, her nose and her eyebrows are two features that are highlighted:

- (4) And here she was now, in her elderly days, with the Coriolanian style of nose and the dense black eyebrows ...

The references to her 'Coriolanian' or 'Roman' nose then recur as external features of her lofty manner. The suspension in concordance line 10, corresponding to example (5) here, also refers to Mrs Sparsit's nose. It shows how her nose and her eyebrows are signs of her severity – this severity being part of her loftiness, as highlighted by the patterns in the concordance.

- (5) 'It is much to be regretted,' said Mrs. Sparsit, making her nose more Roman and her eyebrows more Coriolanian in the strength of her severity, 'that the united masters allow of any such class-combinations.'

These examples show how the information in the suspensions relates to similar information in other descriptions of Mrs Sparsit. The suspensions do not introduce a new view of her, but by repetition enforce features typical of this character.

The function of suspensions as places to describe habitual behaviour or typical character features is not limited to the example of Mrs Sparsit but also found with other characters (see, for example, Mahlberg forthcoming b). Suspensions seem to be a useful place for such information. When presented as circumstantial information, character information that has already been mentioned appears less strikingly repetitive. If a character has a typical manner of saying something it appears almost natural that the same body language occurs again. In this sense, suspensions also help to show what behaviour is to be regarded as character information in contrast to behaviour that may be untypical and a reaction triggered by a specific situation. Behaviour that is due to a particular situation would have to be discounted in the assessment of what constitutes character information, as Culpeper (2001) also explains in his discussion of characterization.

An important point to note is that the discussion so far has focused on suspensions that contain Mrs Sparsit's name and also interrupt her own speech. There are at least two other cases that deserve attention because they show the limitations of the method at its present stage. One case is illustrated by concordance line 12, reproduced as example (6) below. Here it is not Mrs Sparsit's speech that is interrupted, but the speaker is Harthouse. The suspension highlights Mrs Sparsit's point of view when she is spying on Harthouse and Louisa. Although she is hidden from the two characters who are having a conversation, her presence intrudes into the conversation through the suspension. Thus the fact that she pokes her 'Roman nose' into other people's business is also reflected in the presentation of speech.

- (6) ‘My dear child,’ said Harthouse; Mrs. Sparsit saw with delight that his arm embraced her; ‘will you not bear with my society for a little while?’

The other case to consider is illustrated by example (7). This example is not picked up by a concordance search for *Sparsit* in suspensions, because the suspension uses a pronoun and not a name to refer to the speaker (the reporting clause *rejoined Mrs. Sparsit* is not a suspension in the sense of the definition above, because it only contains three words). In the current version of CLiC, suspensions that refer to a particular character can only be retrieved by searching for the character name in suspensions. If pronouns or other reference expressions are used for a character the suspensions cannot be found automatically. Such examples raise questions that have to be tackled in future work, for instance, with the introduction of anaphor resolution to the automated annotation. However, even without this technical solution the current limitations can to some extent be addressed by looking for character names in the surrounding text. In example (7) for instance, the character name appears in the reporting clause that forms only a short speech interruption towards the beginning of the speech instance. The name predicts the pronoun reference in the suspension that occurs later in the passage.

- (7) ‘Sir,’ rejoined Mrs. Sparsit, ‘say no more. In yielding up my trust here, I shall not be freed from the necessity of eating the bread of dependence:’ *she might have said the sweetbread, for that delicate article in a savoury brown sauce was her favourite supper:* ‘and I would rather ...

Example (7) also illustrates another aspect of suspensions that is dealt with in more detail in the next section – the interplay between character and narrator. Mrs Sparsit uses the figurative expression ‘the bread of dependence’ which the narrator comments on by referring to the concrete meaning of ‘sweetbread’. Thus Mrs Sparsit’s pretentious way of speaking is contrasted with very practical details of her life, and, at the same time, the narrator exposes the fact that Mrs Sparsit’s situation is less arduous than she wants to suggest if she can enjoy luxury food. The narrator’s comment, which is given particular emphasis by separating it from the reporting clause *rejoined Mrs. Sparsit* that already interrupted the character’s speech once, illustrates the narrator’s strong presence and his ironic treatment of Mrs Sparsit. Lambert (1981: 61) points out that suspensions can often be found to make fun of characters as well as to attack characters that deserve punishment.

5 Pauses in character speech

Lambert finds further evidence for his claim that Dickens is jealous of his characters in the use of suspensions to handle pauses in character speech. Even without interpreting them as a sign of authorial jealousy, pauses in character speech can be seen as techniques of an intrusive narrator. Lambert (1981: 67ff.) describes the use of an ‘artificial sequence’ as an effective device to make the reader feel to some extent similar discomfort to that experienced by the characters throughout the pause. One of Lambert’s 1981 examples is example (8) below – from *Hard Times*. Here the narrator only refers to a silence once the character started speaking again. This way the character’s speech is interrupted to refer to a pause at a point where – in the story – no pause actually occurs in the character’s

speech. With regard to example (8), Lambert (1981: 68) describes the effect of presenting two narrated pauses in the form of artificial sequences as recreating for the reader ‘the painfulness’ of the silences.

- (8) “Mrs. Bounderby,” said Harthouse, *after a short silence*, “may there be a better confidence between yourself and me? Tom has borrowed a considerable sum of you?”

“You will understand, Mr. Harthouse,” she returned, *after some indecision*: she had been more or less uncertain, and troubled throughout the conversation, and yet had in the main preserved her self-contained manner; “you will understand [...]” [our italics]

Claims about the way in which pauses are actually perceived by readers can ultimately only be verified by experiments involving participants. However, corpus methods can contribute to formulating hypotheses about potential effects. Corpus evidence seems to suggest that pauses described in suspensions may not necessarily be perceived as such striking interruptions of character speech as Lambert claims. The effect that the narrating of a pause in the form of a suspension will have depends not only on the presence of a suspension but also on textual patterns around it, as well as on the patterns that are found repeatedly in suspensions. To see links between pauses and other patterns in suspensions, example (9), which corresponds to concordance line 1, in Figure 3, might be helpful. The example also illustrates how the order of presentation departs from the order of the narrated activities. We have pointed out earlier that suspensions can be seen as devices to suggest synchronicity between speech and body language. Example (9), in effect, explicitly sequences the speech and body language, but employs an order of presentation that does not match the order of the speech and body language. However, as the body language is presented as circumstantial information, the reversal of the actual order of speech and body language does not seem to be highlighted. Similarly, the prepositional phrase *after a short silence*, in example (8), might make the narration of the pause less striking than Lambert (1981) wants to argue.

- (9) ‘A singular world, I would say, sir,’ pursued Mrs. Sparsit; *after acknowledging the compliment with a drooping of her dark eyebrows ...*

While individual examples raise some interesting points, we can get a better picture of options for presenting pauses in character speech, if we look at a larger number of instances. Patterns of the word *pause* can serve as a useful starting point. The word occurs altogether 418 times in the Dickens corpus⁹, of which 83 occurrences appear in suspensions, which means 335 occurrences appear in text outside of suspensions. Of the occurrences of *pause* that do not appear in suspensions, not all refer to a silence, some are forms of verbs or refer to the meaning of ‘break’, as in ‘We made a pause at the toy shop in Fleet Street ...’ (*David Copperfield*). In suspensions, *pause* always refers to a silence. Of the 83 occurrences 78 appear in a prepositional phrase with *after*. Table 4 shows the most frequent patterns of these prepositional phrases and compares them to the numbers of the respective patterns outside suspensions. The log-likelihood value¹⁰ for all three patterns, as well as for the frequencies of the word *pause*, shows significant overuse in the suspensions.

Table 4. Patterns of *pause* in suspensions compared to text outside suspensions.

	Text without suspensions	Suspensions	LL
<i>pause</i>	335	83	242.53
<i>after a pause</i>	47	20	75.76
<i>after a short pause</i>	16	15	82.08
<i>after a moment's pause</i>	7	13	75.47

For all three patterns, the numbers refer only to cases where the pattern constitutes the whole prepositional phrase, in other words there is no postmodification as in example (10). For each of the patterns most of the suspensions in which they occur have the form 'reporting verb + subject (or subject + reporting verb, in the case of a pronominal subject) + Adverbial,¹¹ as in examples (11) and (12).

- (10) 'And when,' asked my guardian, rising *after a pause, during which Mr. Kenge had rattled his money and Mr. Vholes had picked his pimples*, 'when is next term?'
(*Bleak House*)

- (11) 'Still, it is a triumph to me to know that he is so true to himself, and to his name of Dombey; although, of course, I always knew he would be. I only hope,' said Mrs Chick, *after a pause*, 'that she may be worthy of the name too.'
(*Dombey and Son*)

- (12) 'That's all I say. And I suppose,' *added the lofty young man, after a moment's pause*, 'that visitor will ...'
(*Little Dorrit*)

The point that Table 4 makes is that the word *pause* tends to occur in suspensions and typically in the three patterns shown. Corpus linguistic findings emphasize that the patterns that are frequent are not necessarily the same patterns that are easily noticeable. The *after ... pause* patterns may also fall into this category. In examples (11) and (12), the emphasis is less on the pause as such, but the patterns support the organization of the character discourse. In both examples, the sentences that are interrupted add another thought to what is said before and the interruption follows the prefaces *I only hope* and *And I suppose*. The description of pauses in examples such as (11) and (12) is different from examples where the description of the pause receives more emphasis as in (13) and (14). Here the discourse between characters is interrupted, but the interruption does not appear within a turn of a character, but between turns. Also the pause is not presented in the form of an adverbial, but in more prominent grammatical structures.

- (13) There was a *pause*. The countenance of Nicholas fell, and he gazed ruefully at the fire.
(*Nicholas Nickleby*)

- (14) A *pause* ensued, in which the schoolmaster looked very awkward.
(*Our Mutual Friend*)

6 Conclusions

The comparison of Lambert's (1981) data with automatically generated counts for suspensions illustrates that quantitative information can potentially be misleading when claims made on the basis of specified samples are extended to generalizations about whole novels. We have also shown how the annotation of a corpus of novels of Dickens and the tool CLiC that is designed to handle such annotation can open new ways of studying linguistic phenomena in literary texts. Suspensions analyzed in the form of concordances can be useful places to study character information. The example of *pause* further points to similarities between the presentation of body language that accompanies speech and pauses in character speech. While they can be highlighted in the text, common patterns in suspensions describe pauses in the form of circumstantial information so that the pause as such might receive less attention. However, assumptions about what readers do and do not notice will ultimately need to be verified by psycholinguistic experiments. The definition of suspensions as consisting of at least five words thus also requires further study. While our examples are specific to Dickens and only provide some initial insights into the range of functions of suspensions, it is clear that the more general approach illustrated in this article has far-ranging potential. The mark-up of 'quotes' and 'non-quotes' enables a range of key comparisons that can contribute to the analysis of character discourse on various levels. Additionally, the approach is not limited to the study of Dickens's fiction, but can be extended to other authors,¹² as well as comparisons between fictional speech and real spoken language.

Notes

1. For a wider discussion of the application of corpus methods to the study of literary texts see, for instance, Biber (2011); Mahlberg (forthcoming a).
2. For quotations from Dickens we do not provide page references, as we refer to the texts in the corpus that we use for our study, see section 3.
3. Project Gutenberg www.gutenberg.org/
4. The sentence tokenizer uses a regular expression which defines the graphological pattern equivalent to a sentence. Every time a section of text is found that matches the pattern that section of text is taken to be a sentence and is annotated as such.
5. Unlike Lambert (1981), our approach also includes writing in our quoted material.
6. The Spearman's Rho score for these two ranks is 0.97 indicating a very strong positive correlation.
7. Despite more movement in the ranks compared to the results reported in Table 2 the Spearman's Rho score still remains high at 0.9.
8. At present, CLiC is still a prototype, so not freely accessible as yet.
9. The corpus used in this section contains the novels as well as eight further texts by Charles Dickens and amounts to about 4.5 million words.
10. For $p < 0.0001$ the critical value is 15.13. The total number of words in suspensions is 91,953 and in text outside suspensions 4,394,155. The LL has been calculated with Paul Rayson's log-likelihood calculator (see Rayson, n.d.).
11. As these suspension patterns have no equivalent to be compared to outside suspensions, no counts are given for them in Table 4.
12. For a study of suspensions in Jane Austen's *Pride and Prejudice* see Mahlberg and Smith (2010).

References

- Biber D (2011) Corpus linguistics and the scientific study of literature: Back to the future? *Scientific Study of Literature* 1(1): 15–23.
- Carter RA (2004) *Language and Creativity: The Art of Common Talk*. London: Routledge.
- Cheshire3.org (n.d.) Cheshire 3 Information Framework. Available at: <http://cheshire3.org> (accessed 31 December 2011).
- Culpeper J (2001) *Language and Characterisation: People in Plays and Other Texts*. Harlow: Pearson Education.
- Korte B (1997) *Body Language in Literature*. Toronto: University of Toronto Press.
- Lambert M (1981) *Dickens and the Suspended Quotation*. New Haven, CT and London: Yale University Press.
- Mahlberg M (forthcoming a) Corpus analysis of literary texts. In: Chapelle CA (ed.) *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.
- Mahlberg M (forthcoming b) The corpus stylistic analysis of fiction or the fiction of corpus stylistics? In: Mukherjee J and Huber M (eds) *Corpus Linguistics and Variation in English: Theory and Description*. Amsterdam: Rodopi.
- Mahlberg M and Smith C (2010) Corpus approaches to prose fiction: Civility and body language in *Pride and Prejudice*. In: McIntyre D and Busse B (eds) *Language and Style*. Basingstoke: Palgrave Macmillan, pp. 449–467.
- Newsom R (2000) Style of Dickens. In: Schlicke P (ed.) *The Oxford Reader's Companion to Charles Dickens*. Oxford: Oxford University Press, pp. 553–557.
- Project Gutenberg (n.d.) Available at: <http://www.gutenberg.org/> (accessed 31 December 2011).
- Rayson P (n.d.) Log-Likelihood Calculator. Available at: <http://ucrel.lancs.ac.uk/llwizard.html> (accessed August 2011).
- Sinclair J (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Author Biographies

Michaela Mahlberg is Associate Professor in English Language and Applied Linguistics at the University of Nottingham, where she is also the Director of the Centre for Research in Applied Linguistics (CRAL). She is the Editor of the *International Journal of Corpus Linguistics* (John Benjamins), and Co-editor of the series *Corpus and Discourse* (Continuum). Her recent work includes the forthcoming monograph *Corpus Stylistics and Dickens's Fiction* (Routledge).

Catherine Smith is a technical officer in the Institute for Textual Scholarship and Electronic Editing (ITSEE) in the College of Arts and Law at the University of Birmingham. She has research experience in biblical studies, corpus linguistics and computer science. Her main research interests are the application of corpus linguistic methods to literary stylistics and the development of software tools to support research in the humanities.