

Literary dialogues as models of conversation in English Language Teaching

Christian Jones and David Oahey, University of Liverpool

Abstract

This study explores dialogues from a corpus of 19th century fiction as a possible model of conversation for use in English as a second or foreign language classrooms. We employ corpus linguistics methods to investigate the extent to which such dialogues contain lexical chunks and conversation strategies. This is then compared to usage in unscripted 21st century conversations, as found in a large spoken corpus. Findings show that there are a number of similarities to modern day spoken language in the chunks used in the 19th century data but also some key differences in how some common chunks function. It is also clear that many conversation strategies are significantly underused in the fictional dialogues. Overall, the study shows that dialogues from fiction have some potential as a model of conversation and could be used to supplement materials offering very contrived dialogues.

Keywords: *spoken language; conversation; fictional dialogues; models of conversation; corpus linguistics*

Introduction

It has been argued that conversation should occupy a central place in English Language Teaching (ELT) as it is the most common form of interaction between speakers and hence one which most learners want to develop (Thornbury & Slade, 2006). Most teachers, however, will have met learners who have studied the language for a number of years and still feel unable to have a successful conversation. One reason for this may be that, as McCarthy and McCarten (2018) argue, communicative language teaching has often valued speaking practice above practice which develops conversational ability. Learners are commonly asked to undertake activities which require students to speak, but often this is in order to practise specific language points or perhaps for the slightly vague notion of 'developing fluency'. Learners seem therefore to be given speaking activities in the hope that the ability to develop conversations will simply develop naturally. McCarthy and McCarten (2018) argue instead that the ability to converse in English requires a specific skillset, and that strategies which learners need, such as developing their own turn or showing they are interested in what someone is saying, can and should be highlighted, taught and practised with learners. At the same time, many models of conversations provided in English language textbooks are designed with the primary intention of presenting language points. As a result, such dialogues may serve to show particular language at work but do not always help

learners develop common conversation strategies such as those outlined above, and as such are not always useful models.

This article applies corpus linguistics methodology to the second issue highlighted here - the need for useful models of conversations in ELT. It explores a possible alternative conversational model to that found in many coursebooks - dialogues found in literature. Such dialogues have the potential to be both interesting and motivating when viewed from a pedagogical perspective (McRae, 1991; Carter & McRae, 1996). Many learners read literature in their first and second language (either in the original or a simplified form) and we thus assume that many will have an interest in the dialogues contained within novels, short stories or plays. The aim of this paper is, therefore, primarily to investigate the extent to which conversations in literature a) contain examples of frequent features of spoken English, in this case common lexical chunks, and b) contain examples of the language used to realise typical conversation strategies as outlined by McCarthy and McCarten (2018, pp.13-14). In doing so, we hope to discuss whether such dialogues offer a useful model of conversation for learners of English which could be used at least as a supplement to standard materials. Once this data has been examined, we discuss how the use of such dialogues could develop learners' awareness of language and strategies used to maintain and develop conversations. This is achieved by examining data from the Corpus Linguistic in Context (CLiC) corpus (Mahlberg, Stockwell, de Joode, Smith., & O'Donnell, 2016) in particular, a large corpus of 19th century fiction which allows us to look at quoted data - speech as used in dialogues between characters. The data for CLiC is examined in comparison to a 21st century spoken corpus of conversations contained in the Spoken British National Corpus 2014 (2018) (hereafter Spoken BNC2014).

Literature review

Research in corpus linguistics has produced detailed descriptions of common features of conversational language. There is now a general understanding that conversational language has fundamental features which make it different to many forms of writing (e.g. Biber et al. 1999, Carter & McCarthy, 2006). Conversations have a primarily social goal, and consist of two (or more) people taking turns and interacting rather than two or more individuals making sentences. Conversations also contain forms, such as particular lexical chunks and spoken discourse markers, which are generally not used in written texts (see Carter and McCarthy, 2006 for a more detailed explanation). This section reviews previous work on written literary dialogue and unscripted conversations before stating the specific questions this research attempts to answer.

Written literary dialogue and unscripted conversations

It is perhaps obvious that conversations which we find in literature are not the same as those we find in real life. One obvious difference is the fact that conversations in literature serve a different purpose. In real life, the goals of conversation are, at least in part, interpersonal ones, while in literature they will often serve the purpose of telling us about plot, theme or characters. There are, however, a number of spoken language and discourse features we can find in literary dialogues. Some of these have similarities to real conversations and some occur but with a fundamentally different function.

Leech and Short (2007, pp.128-134) provide a useful comparison of unscripted conversation and dialogue in fiction. They suggest that features common in conversation (Carter & McCarthy, 2006) such as discourse marking or the tendency to use fewer complex sentences may occur with less frequency in fictional dialogues. This, again, is likely to be because dialogues are in fiction to tell us about characters or to signal a plot or thematic development and not always to create a completely realistic picture of conversation. The examples below show this, with two contrasting dialogues employing the discourse marker 'well', the first taken from unscripted, naturally occurring conversation and the other from a work of fiction:

Example 1. *Use of 'well' in Spoken BNC2014*

S0094: mm (.) yeah cos I like their they had erm organic Fairtrade just normal tea

S0021: mm

S0094: which was really quite nice and

S0021: mm

S0094: quite cheap and good and stuff but they stopped stocking that now here anyway

S0021: d' you reckon like people thought in the recession people aren't gonna want organic stuff so we just won't 't ?

S0095: mm

S0021: **well** that was my theory

File S23A

Example 2. *Use of 'well' in literary dialogue*

'I don't mind summer rain. In fact I like it. It's my favourite sort.'

'Your favourite sort of rain?' said Thea.

I remember that she was frowning, and pondering these words, and then she announced: 'Well, I like the rain before it falls.'

Coe, (2008), pp. 161-2

Here, we can clearly see the discourse marker 'well' being used in a similar way in each conversation. Similarly, there are features of conversation common in both texts: speakers co-operate to construct the dialogue together, although this is clearly more strongly apparent in the Spoken BNC2014 sample in Example 1. However, we can also see that there is more complexity and author explanation in the literary dialogue in Example 2 than in the British National Corpus sample. We see, for example, a 'suspension' (Mahlberg & Smith, 2012) which is the interruption of a character's speech by at least five words from the narrator to tell us what the character was doing as she spoke. Such suspensions can serve to help us to visualise the scene more easily. The fictional dialogue therefore has an 'air' of realism but is also clearly designed to tell us something about a person who would describe rain in this unusual way.

Aside from the tendency for greater complexity in literary dialogues, Short (2012, p. 21) also shows that in the fictional world an author has many more options when representing what speakers say, which are unlikely to be employed by speakers in conversations. When writers report speech they have several choices, including direct speech as in Example 3 and indirect speech as in

Example 4, both taken from Conan Doyle's *Hound of the Baskervilles* in the CLiC corpus (2018), where the reporting has been underlined and the suspensions are in italics.

Example 3. 'Really Watson, you excel yourself,' said Holmes, *pushing back his chair and lighting a cigarette*.

Example 4. The man looked surprised and a little embarrassed. 'Why, there's no good my telling you things, for you seem to know as much as I do already,' said he. 'The truth is that the gentleman told me that he was a detective and that I was to say nothing about him to anyone.'

In these examples, we can see the use of suspensions to paint a clearer picture for the reader and these would not, of course, be present in naturally occurring conversations. When people in real conversations report what someone said, the reported speech is rarely

an exact quote of the previous speech (Carter & McCarthy, 2006; Biber & Conrad, 2009). Instead, speakers are far more likely to simply report what others have said directly, indirectly or in summary as in the examples from the Spoken BNC2014 below, with the reporting underlined. Example 5 shows a direct report, Example 6 an indirect report and Example 7 a summary. Speakers may also add comments or views on what they are reporting (in italics in the examples below) and in addition to 'say' often use structures such as *BE + like* and *NP + GO*, or *NP + BE all*. (Biber & Conrad, 2009, p109). Examples 8 and 9 show this.

Example 5. However for like months cos –ANON said that I 'm making a secret thing you know that kind of thing (.) and *I like guessed almost straight off that he was making like bootleg alcohol for me for Christmas.*

File S23A

Example 6. Yeah it's er he says he (.) he came home from work and just said that his back hurt and that he's had like some like muscle spasm at work or whatever and he said he was going to lie down erm and *then I went up like half an hour later and just got him like a hot water bottle (.) just to give him some like heat it so erm basically his back ...*

File SC29

Example 7. ... which I think is good and all sorts of things but they come round and inspect it from time to time and and she was saying they're going to get into trouble cos their hedge was too high and she couldn't use the hedge trimmer so she was hoping –ANON would get back in time to do it.

File S2UT

Example 8. I'm like are you from Idaho City and she's like no do I look like it? (Biber & Conrad, 2009, p109)

Example 9. He was all I love you sweetie (Biber & Conrad , 2009, p109)

Another common technique which writers use (and which is largely absent from naturally occurring speech) is to report what someone said and also how s/he said it, as in the invented examples 10 and 11:

Example 10. 'Go!' he said angrily.

Example 11. 'Go!' he screamed.

Many creative writers tend to frown on the use of adverbs, as in Example 10, to modify what Stephen King calls 'verbs of dialogue attribution' since they are an indicator of poor verb choice. He advises the writer to 'use the adverb in dialogue attribution only in the rarest and most special of occasions ... and not even then, if you can avoid it' (King 2000, p.125) but rather choose a stronger verb of attribution as in Example 11.

More specific features of speech such as hesitation have also been discussed. Short (1996) notes that hesitation marked by filled pauses such as 'er' or unfilled pauses are, not surprisingly, more common in unscripted conversation, whether this be from native speakers or learners (Jones, Byrne, & Halenko, 2017). Hesitation in fictional dialogues may also serve a different purpose. While pauses are used in conversation to buy speakers time, to mark a pause and/or to hold the floor, Short (1996) argues that in literature they often function differently. A pause by a character may, for example, be used to build up tension between speakers or to tell us that a character is hesitant and nervous.

Related to these differences in speech, work in corpus stylistics (see Hoover, Culpepper & O'Halloran, 2014 for a helpful exploration) has enabled researchers to examine large bodies of fiction and explore the characteristics of the language used by writers in dialogues. Mahlberg and Smith (2012), for example, examine the use of suspended quotations, as already mentioned, the interruption of a character's speech by at least five words from the narrator. Examining a corpus of Dickens' work, Mahlberg and Smith show how suspensions can also be used by an author to contribute to characterisation. Focusing on the character Mrs Sparsit (from *Hard Times*), their analysis shows how Dickens uses suspensions to paint a picture of a character who is lofty. Stockwell and Mahlberg (2015) also use a corpus to explore what the non-quoted text in *David Copperfield* by Dickens can tell us about the characterisation of Mr Dick if examined through the lens of what they term 'mind modelling'. This type of research shows, as mentioned previously, a marked dissimilarity between

literary dialogues and unscripted conversation, where speakers display their character through choice of language, gesture, intonation and so on.

Overall, the studies reviewed illustrated some key differences between unscripted, naturally occurring conversations and dialogues we can find in literature. Table 1 (adapted from Byrne & Jones, forthcoming) below summarises some of the key general differences discussed so far.

Table 1: *Key differences between dialogues in fiction and unscripted conversations*

Dialogues in fiction	Unscripted conversations
Characters	Participants
Used to develop the plot/theme(s)/character(s) in some way	Interpersonal/transactional goals
Constructed by the author	Co-constructed by participants
Topics chosen by the author	Topics chosen and developed by participants
Features of spoken English such as hesitation and false starts may be used to inform us about a character or to contribute to the theme. They may also contribute to the overall style of the literature	Features of spoken English such as hesitation and false starts are a standard 'performance feature' of conversations
More complexity	Less complexity

Despite these clear differences, there are of course also similarities between dialogues in fiction and unscripted conversations. Example 2 discussed previously shows that literary dialogues will often contain features such as spoken discourse markers. In addition, Short (1996) shows that ellipsis is one feature of unscripted conversation which can be found in dramatic dialogues, whereas certain other features such as hesitation devices may be absent. Carter (1998) makes a similar argument, suggesting that dialogues from plays often contain many features of spoken English found in spoken corpora and gives an example of a dialogue from Pinter to illustrate this point. Fictional conversations also have other similarities to naturally occurring data. Leech and Short (2007, pp.128-134) give examples such as the use of contracted forms and informal lexical items, which can and do feature in fictional dialogue. We can see this if we return to the dialogue from Example 2, with these characteristics (contractions, informal lexical items and discourse markers) underlined:

Example 12. *Sample literary dialogue with common spoken features underlined*

'I don't mind summer rain. In fact I like it. It's my favourite sort.'

'Your favourite sort of rain?' said Thea.

I remember that she was frowning, and pondering these words, and then she announced:

'Well, I like the rain before it falls.'

Coe, (2008) pp. 161-162.

One other key area of similarity, that of lexical chunks, is discussed in the next section.

Lexical Chunks

Research in corpus linguistics over the last three decades has established that a significant amount of speech consists of pre-formed strings or clusters of words, which we here refer to as 'lexical chunks'. Adolphs and Carter suggest that chunks 'are extremely frequent, are necessary in discourse and are fundamental to successful interaction' (Adolphs & Carter 2013, p. 36). More than half of all language choices are formulaic in some way, a feature more prevalent in spoken rather than written language (Biber et al., 1999; Erman & Warren, 2000) and many other studies have highlighted the importance of chunks in spoken interaction (e.g. Pawley & Syder, 1983; Wood, 2010). Chunks are typically expected to be between two and four words in length (McCarthy 2010) and while longer chunks are possible, they tend to be much lower in frequency (O'Keeffe, McCarthy & Carter, 2007). With regards to function, although they are relevant for conveying meaning they are also considered to be one key aspect of fluency. It is for this reason they are examined in this study, something we expand upon at the end of this literature review. It is also undoubtedly the case that fictional dialogues will contain such chunks. Mahlberg and Wiegand (2018), for example, explore speech in *Great Expectations* and show how chunks (which they term common clusters) function in quotes and non-quotes, i.e. in dialogue and non-dialogue. Their analysis also demonstrates many similarities to those chunks used in the BNC spoken sub-corpus (1993) such as the frequency of the cluster 'what do you mean'. Example 13, taken from the *Hound of the Baskervilles* (CLiC 2018) shows that even a somewhat archaic text can contain lexical chunks (underlined) which we would expect in modern-day usage.

Example 13. *Chunks in Hound of the Baskervilles*

'What do you think of that, Watson?' cried Holmes in high glee, rubbing his hands together with satisfaction. "Don't you think that is an admirable sentiment?"

Dr. Mortimer looked at Holmes with an air of professional interest, and Sir Henry Baskerville turned a pair of puzzled dark eyes upon me. 'I don't know much about the tariff and things of that kind'" said he, 'but it seems to me we've got a bit off the trail so far as that note is concerned.'

Dialogues in EFL/ESL Textbooks

Finally, we briefly illustrate some features of dialogues in EFL textbooks which we believe mean they are not satisfactory models of conversation. The artificial nature of some EFL coursebook dialogues was pointed out by McCarthy and Carter (1994) and Gilmore (2004). These dialogues can sound more like interrogations, during which the first speaker asks questions and the second speaker gives complete answers, so that every turn is completely efficient in getting its message across. This, as Carter (1998) suggested, reflects a 'can-do' society, 'in which interaction is generally smooth and problem free, the speakers cooperate with each other politely, the conversation is neat, tidy, and predictable' (Carter 1998, p47). Example 14 below lacks standard 'performance features' of conversations in spoken English mentioned in table 1 above, such as hesitations, repetitions and false starts, the absence of which make the exchange seem contrived.

Example 14. *Sample textbook dialogue.*

Dave and Neil are discussing what type of food they like.

DAVE: Do you like Indian food?

NEIL: Yes, I do! It's my favourite.

DAVE: Why don't we go out for a curry, then?

NEIL: No – we can't do that tonight.

DAVE: Why not?

NEIL: Paul's coming out with us tonight, and he doesn't like curry.

DAVE: Doesn't he? What kind of food does he like, then?

NEIL: I think he likes Chinese food. Shall we go to the *Peking*?

DAVE: No, I'd rather not – I don't like Chinese food very much.

NEIL: All right, then – let's all stay in and phone for a pizza, shall we?

DAVE: Does Paul like pizza?

NEIL: It's his favourite food!

DAVE: OK, that's what we'll do!

King (2014: 74)

The focus of this article is therefore on whether common features of conversations are common in a corpus of fiction and the extent to which they function in similar ways. The intention is to explore their potential as models of spoken language. One reason for this is because, as mentioned, textbook dialogues often present a very unrealistic model of conversation. In addition to this, we feel that unedited conversations taken directly from a corpus are unlikely to be very motivating for learners, though corpus data can of course be adapted and used to inform classroom materials (see McCarthy & McCarten, 2018 for a recent discussion and useful examples). Dialogues from literature could potentially act as a useful halfway point between heavily contrived course materials such as the dialogue in Example 14 and unedited corpus data. In this article we therefore hope to answer the following research questions:

RQ1. What are the frequencies and functions of the most common 4-word lexical chunks used in dialogues from a 19th century literature corpus and how do these compare to usage in 21st century unscripted conversations?

RQ2. What are the frequencies and functions of common forms used to realise conversational strategies in dialogues from a 19th century literature corpus and how do these compare to usage in 21st century unscripted conversations?

Methodology

In answering the research questions set, we used the 19th century reference corpus of approximately 4.5 million tokens taken from the open access CLiC corpus (Mahlberg et al., 2016). The data is divided into five sub-corpora: all text, short suspensions, long suspensions, direct quotes and non-quotes and contains a number of popular texts including Bram Stoker's *Dracula* and Conan Doyle's *Hound of the Baskervilles* (see Mahlberg et al. 2016 for more detail). Although these texts are not of course from the present day, it was felt that the corpus contains many texts (such as the Conan Doyle novel) which learners may be familiar with and /or motivated to read at some stage in their learning. Besides, the differences between archaic and modern dialogue are not so clear-cut. C.P Snow, for example, writing a novel *The Masters* in the 1950s, set in a 1930s Cambridge College, acutely observed the differing usages of his characters, applying a geological metaphor to describe how the idiolects of speakers acquired in different eras overlapped in conversations:

It was interesting to hear so many strata of speech round one table. Old Gay, for example, used 'absolutely', not only in places where the younger of us might quite naturally still, but

also in the sense of 'actually' or even 'naturally' - exactly as though he were speaking in the 1870's. Pilbrow, always up to the times, used an idiom entirely modern, but Despard-Smith still brought out slang that was fresh at the end of the century - 'crab', and 'josses', and 'By Jove'. Crawford said 'man of science', keeping to the Edwardian usage which we had abandoned. So, with more patience it would have been possible to construct a whole geological record of idioms, simply by listening word by word to a series of college meetings.

(Snow 1951: 160)

The point here is that speakers in many interactions bring with them individual speech repertoires which are not necessarily co-synchronous. While there may be differences in usages like those described by Snow, there may well also be features, such as chunks, which have become fossilised across different generations of speakers and which are still to be found in present day conversations. It was therefore considered that 19th century fictional dialogues had the potential to act as motivating and interesting models of conversation to modern learners.

Due to this study's focus on conversation, we analysed data from the 1,611,083 token quote sub-corpus of the 19th century reference corpus. This was because the quote sub-corpus shows characters in dialogue with each other (being quoted as speaking by the author) as opposed to characters in monologue or description. Example 15 illustrates this difference with two short samples from the *Hound of the Baskervilles*, based on a search for the word 'no' taken from the CLiC Corpus (2018).

Example 15. *Sample quote and non-quote from Hound of the Baskervilles*

Quote

'But that was all.'

'No, no, my dear Watson, not all--by no means all. I would suggest, for example, that a presentation to a doctor is more likely to come from a hospital than from a hunt...'

Non-quote

The most of them would by no means advance, but three of them, the boldest, or it may be the most drunken, rode forward down the goya.

The approach adopted was to first analyse the data by searching for the most common four-word lexical chunks in the CLiC data. These frequency counts were then compared to 21st century spoken data. To undertake this we compared the number of occurrences of each chunk found with the Spoken BNC2014. This corpus consists of 11,422,617 tokens of

conversations collected in the UK. It is the most up to date record of British conversation currently available in an open access format. The corpus was chosen so a comparison of frequency could be made between the CLiC data and naturally occurring, unscripted data. When we compared the number of occurrences of each item across both corpora, log-likelihood scores were also computed using Rayson’s (2018) UCREL log-likelihood online calculator. Log-likelihood scores are useful as they do not assume a normal distribution of the item in question within a text and allow us to observe the significance of different levels of frequency in corpora of different sizes. Four-word chunks were chosen as a focus because items of this size tend to be reasonably meaningful units and, occurrences of larger chunks tend to reduce considerably so that five- and six-word chunks are considerably less frequent (O’Keeffe, McCarthy and Carter, 2007).

Following this, the data was explored to find examples of conversational strategies described by McCarthy and McCarten (2018, p.13-14) and discussed earlier in this paper. The broad strategies they outline are as follows: managing the conversation, constructing your own turn, listenership and taking account of others. These strategies can also be subdivided and exemplified by typical language used to realise them, and it was by doing so that the data were analysed in this study. The 19th century sub-corpus was initially searched for typical language which McCarthy and McCarten give as examples of each strategy based on findings from spoken corpora. These are shown in table 2.

Table 2. *Conversation strategies and typical language used to realise them (from McCarthy and McCarten, 2018, pp.13-14).*

Strategy	Sub strategy	Typical language used to realise the strategy
Managing the conversation	Referring back to an earlier comment	As I was saying
	Ending a conversation	Better go
Constructing your own turn	Taking time to think of an answer	Let me think
	Elaboration	In other words

Listenership	Responding to news or information	That's wonderful
	Showing understanding	Uh huh
Taking account of others	Projecting shared understanding	And things like that
	Telling new information	What happened was

Once the initial search for language was conducted, concordance lines were examined to check each item was actually being used for the strategy above. For some items such as 'As I was saying' this was straightforward but for others, such as 'Better go,' items had to be closely checked to ensure they were fulfilling the function given in table one. Only the occurrences where the item fits the function were explored. As with the lexical chunks, the Spoken BNC2014 was used as a point of comparison with the 21st century and log-likelihood scores were also computed using Rayson's (2018) UCREL log-likelihood online calculator. In order to answer both research questions, concordance lines were checked and then examples viewed in context to obtain a picture of how the items functioned in each corpus. This type of qualitative analysis, based on an initial frequency analysis, can help to give a more fine-grained picture of the data.

Results and Discussion

For ease of reference, the data is discussed in relation to each research question in turn.

RQ1. What are the frequencies and functions of the most common 4-word lexical chunks used in dialogues from a 19th century literature corpus and how do these compare to usage in 21st century unscripted conversations?

Table 3 shows the most frequent four-word chunks in CLiC in comparison to the Spoken BNC2014. Log-likelihood comparison are also given, with + indicating overuse in CLiC and – underuse in comparison to the Spoken BNC2014 when occurrences in each corpus are compared. Significance levels are given under the table.

Table 3. *The twenty most frequent 4-word chunks in CLiC 19th century quotes in comparison with frequencies in the Spoken BNC2014*

Rank	CLiC	Spoken BNC2014	Log-likelihood comparison
1	I should like to (221)	I should like to (1)	+ 911.52 ****
2	I do not know (140)	I do not know (14)	+495..24 ****
3	I am going to (126)	I am going to (55)	+319.05 ****
4	I beg your pardon (103)	I beg your pardon (14)	+ 348.66 ****
5	What do you think (101)	What do you think (440)	+17.55 ****
6	I don't know what (101)	I don't know what (1746)	-99.63 ****
7	What do you mean (95)	What do you mean (571)	+ 2.14
8	I don't want to (91)	I don't want to (682)	-0.25
9	I have no doubt (87)	I have no doubt (2)	+345.16 ****
10	I am sure you (83)	I am sure you (1)	+336.46 ****
11	I am sure I (82)	I am sure I (1)	+332.40 ****
12	At the same time (81)	At the same time (546)	+ .018
13	For the sake of (78)	For the sake of (69)	+141.11 ****
14	I want you to (74)	I want you to (100)	+98.49 ****
15	It would have been (70)	It would have been (194)	+ 38.50 ****
16	I will tell you (67)	I will tell you (12)	+216.01 ****
17	To speak to you (62)	To speak to you (13)	+193.50 ****
18	I'll tell you what (62)	I'll tell you what (128)	+ 53.03 ****
19	I do not think (62)	I do not think (1)	+249.23 ****
20	I wish I could (61)	I wish I could (57)	+106.65 ****

**** = $p < .0001$, *** = $p < .001$, ** = $p < .01$, * = $p < .05$

Differences

Some clear differences can be observed when we examine the log-likelihood comparisons. A number of the significantly more frequent examples do, to a certain degree, reflect features of the more archaic style of dialogues in the CLiC corpus. 'I should like to', for

example frequently functions in the CLiC data as a demand, as in 'I want to/ would like to', a use that has only one example in the Spoken BNC2014. The edited concordance lines in figure 1 show some examples of this.

Figure 1. *Concordance lines for 'I should like to' from CLiC*

1	. 'Indeed! pray tell us what it is	.I should like to	be a governess.'
2	you'll be ready to worship me--you really must stay	I should like to	see you very much; but I shall have many opportunities'
3	I could be always young, I would be always single	.I should like to	enjoy myself thoroughly, and coquet with all the world, till
4	Well, Miss Grey, if it's all the same to you,	I should like to	hear that chapter in the First Epistle of St. John
5	'How could I ride in the rain,	I should like to	know. That damned pelting shower was vexatious enough
6	so well, and is so active, and clever, and kind	I should like to	know your mother. Will you introduce me to her some
7	smile on him. Or his right either-- and that's what	I should like to	see; so now you've heard it.'You know you are
8	You know you are talking nonsense. So	I should like to	see it. And so would mamma too, I'm sure

Other chunks such as 'I beg your pardon' in the 19th century corpus also seem to reflect the language of the era, in a similar way to the expressions noticed by C.P. Snow above. This expression is clearly much less frequent in the Spoken BNC2014. However, in the Spoken BNC2014 'pardon' as a stand-alone item is more frequent (479 occurrences) as against 245 occurrences in CLiC. The functions also differ to a certain degree. While in both sets of data, the predominant use is to ask somebody to repeat themselves, in the CLiC data it is also often used by speakers to excuse themselves as part of the chunk 'pardon me'.

Other chunks may be significantly more frequent in the CLiC data because they are of use as devices to help develop plots or tell us something about a character. Two examples of this are 'I am sure I' and 'I am sure you'. In the CLiC data 'I am sure I' tends to function to tell us what a character thinks they will do or what their intentions are. As a result, it tells us something about the kind of character they are. In the Spoken BNC2014 (as 'I'm sure I'), it functions as either a prediction or to suggest what a speaker thinks they remember doing in the past.

Examples of this can be seen below.

Example 16. *Samples of 'I am sure /I'm sure I' in CLiC and Spoken BNC2014*

CLiC

James Vane looked into his sister's face with tenderness. 'I want you to come out with me for a walk, Sibyl. I don't suppose I shall ever see this horrid London again .**I am sure** I don't want to.'

(from *The picture of Dorian Gray*, Oscar Wilde)

Spoken BNC2014

S0315: I like that cos I like green anyway

S0255: >> --UNCLEARWORD that'll be the next colour so next week when I come I **'m sure** I'll have blue all over

File S28F

It is also notable that 'what do you think' is significantly more frequent in the CLiC corpus data. This may also be because, as well as contributing to the realism of the dialogue, it acts as a useful device for writers to bring other characters into conversations; just as in a real conversation, speakers include others by using this chunk when seeking others' views.

Example 17. *'What do you think' in CLiC and Spoken BNC2014*

CLiC

'Well, Miss Grey, **what do you think** of the new curate?' asked Miss Murray, on our return from church the Sunday after the recommencement of our duties.

'I can scarcely tell,' was my reply: 'I have not even heard him preach.'

(from *Agnes Grey*, Anne Brontë)

Spoken BNC2014

S0680: think so I 'm going to go and knit

S0679: >> erm **what do you think** we should do tomorrow ? er can we do anything useful in --ANONnameM 's kitchen ? Can we clear I think you were quite interested in investigating was it on top of the table ?

S0680: erm well if I move those boxes erm that box on the on the table is I believe is quite heavy

File S2CY

Other differences we can observe may simply reflect a writer's need to create a certain sense of realism. One example of this is the use of 'I do not know'. In the CLiC data, this chunk is significantly more common than in the Spoken BNC2014, while 'I don't know' is significantly more common in the Spoken BNC2014 (504 occurrences in the CLiC data and 14,621 in the BNC). The high frequency of this item in naturally occurring conversations is, as O'Keeffe, McCarthy and Carter (2007) observe, because it serves several interpersonal functions including the need to gain time and to be indirect. The greater frequency of 'I do not know' may be because the writers did not see the need to display such functions in order to further their narratives or develop characterisation. As we will discuss later, in literary dialogues, there is clearly less need for interpersonal uses of language than in face-to-face unscripted conversations.

Similarities

Despite these differences, there are also many similarities in the chunks used in both corpora. One obvious example is the predominance of 'I' and 'you' in each set of chunks, something we would expect in conversational exchanges. There are also other notable similarities when we look in particular at table 3. As mentioned before, 'what do you think' is highly frequent in both corpora and in fact significantly more frequent in the CLiC data when we observe the log-likelihood scores. Other chunks in table three show similarities including 'I don't know what' and other uses of 'I don't want to' and 'I don't know if'. This suggests that even in texts from the 19th century there are similarities in frequent forms used in spoken language when we compare to 21st century unscripted conversations.

RQ2. What are the frequencies and functions of common forms used to realise conversational strategies in dialogues from a 19th century literature corpus and how do these compare to usage in 21st century unscripted conversations?

Table four shows the frequency of items used to realise common conversation strategies. Please note that as mentioned on the methodology section, the table does not show the absolute frequency of each item but only when it fulfils the conversation strategy.

Table 4. Occurrences of common items to fulfil conversational strategies in CLiC 19th century quotes in comparison to Spoken BNC2014

Main strategy	Item	Occurrences in CLiC	Occurrences in Spoken BNC2014	Log-likelihood score
Managing the conversation	As I was saying	11	21	+ 10.35 ***
	Better go	5	35	+ 0.00
Constructing your own turn	Let me think	5	29	+ 0.16
	In other words	5	52	-0.74
Listenership	That's wonderful	1	17	-0.94
	Uh huh	0	116	-30.61 ****
Taking account of others	And things like that	2	420	-93.80 ****
	What happened was	2	62	-6.92 **

**** = p <.0001, *** = p < .001, ** = p <. 01, * = p<.05

What is notable from this data is that many of the items used to realise these strategies are clearly more frequent in the Spoken BNC2014. We would argue that this is because conversations in literature, as noted in the literature review above, are constructed by authors to tell us about characters, themes or plots. They therefore will often seek to create what Leech and Short (2007) call an illusion of reality, whereas participants in real conversations have a much greater need to use such strategies to interact.

Having noted this, the differences in frequency of strategy are only significant in terms of listenership and taking account of others. We can see, for example, that the use of 'uh-huh' is significantly more frequent in the Spoken BNC2014. This is of course because characters do not need to show they are listening. There is no interpersonal reason to choose particular language and so it will depend on the author's wish for naturalness. Although this may seem only a minor omission, it is worth noting because listenership is an important aspect of conversation skills. We will return to this in the conclusion.

The frequency of vagueness and the accompanying assumptions of shared knowledge show how such items serve an important function in unscripted conversations. Some examples of this are shown in figure 2.

Figure 2. *Concordance lines of and things like that' from the Spoken BNC2014.*

1	S24A 59	clearly you can see that erm by doing these kind of forums	and things like that	he's probably thinking longer term that he can't be out
2	S24E 336	routine to suit me in other words typically we 'd work late	and things like that	S0520: mm S0521: but the Norwegians they were out of the door at four
3	S26N 537	S0152: ah S0013: but lots of people coming from the Eden project you know	and things like that	but S0152: oh so you were that side okay yeah cos it 's
4	S28F 253	is really helping S0315: good S0255: so er S0315: >> so does it stop the anxieties	and things like that	as well ? S0255: >> yeah I think er it stops the huh all
5	S28F 1888	getting so anxious I got to the point where I was hyperventilating	and things like that	S0255: yeah S0315: she gave me these techniques to do S0255: yeah S0315: and to visualise
6	S28F 2020	here is where it 's groin S0255: yeah	S0315: and	so pelvis S0255: mm yeah S0315: ah S0255: yeah S0315: so that's why then I

	S0315: and ba- lower	things	
	back S0255: yeah	like that	
	know when you 've got	and	
7	S28F 2527	things	S0255: oh yeah S0315: I mean it
	your own business	like that	has been a good summer and erm
	you have work		
	stresses		

The significantly higher frequency of chunks such as ‘and things like that’ show that in unscripted conversations there is a clear need for speakers not to over-elaborate and presume, at least to some degree, shared knowledge. Characters do not have the need to do so, unless, again, the author wishes to create a better illusion of reality or to tell us something about a particular character - perhaps he or she is deliberately vague, for example. The greater frequency of ‘what happened was’ also shows this interpersonal function. In the Spoken BNC2014 it is used to clarify a narrative recount and involve the other person so the function is more interpersonal. In fictional conversations, there is little need to do so, as we, the readers, have already read what happened previously. When it does occur, it is because one character wishes to recount and clarify something to another and thus the function is remind readers of what happened previously or to inform us of something which happened to a character of which we were unaware.

Example 18. ‘*What happened was*’ in CLiC

What happened was simply this. After I left you yesterday evening, Harry, I dressed, had some dinner at that little Italian restaurant in Rupert Street you introduced me to, and went down at eight o'clock to the theatre.

(from *The picture of Dorian Gray*, Oscar Wilde)

Despite these noted differences, it is interesting to note that ‘as I was saying’ is used significantly more in the CLiC as a means of organising a conversation. This seems to be because it has a clear use as a narrative device: it reminds readers of something said by a character previously and thus can help to reinforce the plot, theme or characters, as well as to add a sense of realism to the dialogue. The function is largely the same as in the BNC data and serves to help organise what a character is saying, perhaps because they have lost the thread and wish to link what they are currently saying to something mentioned previously. The examples from CLiC and the BNC show this.



Example 19. 'As I was saying' in CLiC and Spoken BNC2014

CLiC

That will be a matter of course. But, **as I was saying**--Let me see. Yes--all that waiting will be intolerable to me .It is such a bore for a man when he has made up his mind on such a matter as marriage, not to make the change at once, especially when he is going to take to himself such a little angel as you are.

(from *The Small House at Allington*, Anthony Trollope)

Spoken BNC2014

S0255: yeah so we used to sit and do that and (.) and again when you 're there we 'd maybe have one beer one glass of wine with it but not lots

S0315: yeah yeah

S0255: erm (.) so

S0315: >> keep it to a minimum

S0255: yeah so **as I was saying** cos I er when he went out he obviously came back with a van and loaded up with all the beer and the wine erm including like the boxes of wine which are real cheap

File 28F

Other items did not differ significantly in terms of the log-likelihood comparison but did differ in terms of the way they functioned. The most common function of 'Better go' in CLiC, for example, was for characters to give each other advice or for characters advising themselves. This function also existed in the BNC data but was not as common. In the BNC, in contrast, the most common use of this was by an individual to signal they wished to end a conversation soon. This usage existed in the CLiC data but was mainly as a device to signal the movement of characters from one scene to the next. The examples below show these uses.

Example 20. 'Better go' in Spoken BNC2014 and CLiC

Spoken BNC2014

S0083: but so er just when you get used to the idea that jazz night was such a night at this pub then it changes (.) but there 's usually a lot of it around

S0086: mm (.) well anyway I 'd **better go** now cos I want it to be light while I 'm out

S0038: mm

File S35U

CLiC

'Thank you: that contents me for to-night. Now you had **better go**; for if you stay longer, you will perhaps irritate me afresh by some mistrustful scruple.'

'And the school, Miss Eyre? It must now be shut up, I suppose?'

'No. I will retain my post of mistress till you get a substitute.'

He smiled approbation: we shook hands, and he took leave.

(from *Jane Eyre*, Charlotte Brontë).

The final item we wish to focus upon was 'Let me think'. In the CLiC data, this had a very different function. It was not really used by characters to signal that they wanted to take time to think of an answer but rather as a plea to or an imperative to themselves to think of another person. The example below shows this:

Example 21. 'Let me think' in CLiC

No I will never write, I will never think of Armadale again.

'Yes! Let me write once more--**let me think** once more of him, because it quiets me to know that he is going away, and that the sea will have parted us before I am married.'

(from *Armadale*, Wilkie Collins)

Conclusions

Overall, the analysis shows that there are some clear similarities between the conversations in the quotes from CLiC. We have shown, for example, that a number of the four-word chunks in the literature dialogues are similar in form and function to those found in the Spoken BNC2014. Examples we have illustrated include 'I don't know what'. This suggests that dialogues in literature have the potential to act as useful models of spoken language, giving learners exposure to common chunks used to realise conversational goals. Some

chunks such as 'I should like to' are significantly more frequent in the literary dialogues and highlight the more archaic world of 19th century literature we have examined. We have argued that such archaic uses must be noted and understood but they could also act as a useful point of comparison to help learners notice 21st century usage, should dialogues from 19th century texts be employed in class.

When we examined the language used to realise typical conversation strategies (as identified in McCarthy & McCarten, 2018) we have shown that these functions are noticeably less frequent in the literary dialogues, aside from the use of 'As I was saying' as a conversation/narrative management tool. What was particularly noticeable here were interpersonal uses of language such as showing listenership, which is significantly less frequent in the CLiC literature examined. As we have argued, this is largely because the functions of conversation in literature are to tell us something about character, theme or plot and, at times, to paint a picture of a setting. Therefore, the interpersonal goals common in unscripted conversations do not need to be present, a finding which Jones (2017) also made in relation to scripted soap opera dialogues. Such differences also need to be noted, and again, it is a useful exercise for learners to be made aware of 'missing' conversational strategies.

Overall, we would suggest that such dialogues could act as a helpful model of conversations for learners. In motivational terms, as we have argued, they are likely to be more interesting, and accessible to learners than unedited conversational exchanges from a corpus and could be a useful supplement to conversational models informed by corpora (see McCarthy & McCarten, 2018 for useful examples of such an approach).

In order to use such dialogues we would advocate that teachers take a text-based approach (Timmis, 2018) whereby dialogues are chosen for use in class firstly on the basis that they are likely to engage a particular group of learners, rather than because they contain certain forms. Following this, teachers can then examine the forms and functions a literary conversation contains and assess the extent to which learners may find it useful in linguistic terms. A text may be interesting and potentially engaging but if it does not contain any language which students can use to in some way develop their awareness then they are likely to find it ultimately demotivating. Dialogues used in this article serve as examples of such texts: we considered them to be potentially engaging but also to contain useful aspects of conversational language such as lexical chunks. We would argue that such dialogues offer a more useful, engaging and realistic model of conversation than that offered in some published materials.

Once a suitable dialogue has been found, we would suggest the use of an Access, Activity and Awareness framework (Jones & Carter, 2012) with such dialogues in class. Access simply means to choose an interesting text which has an 'access' point for learners and exploit this access point in lead-in activities to help learners engage with a text. Activity involves students actively participating in working with the texts. This can include typical communicative activities such as prediction, re-assembling texts, discussing questions related to meaning, or acting out dialogues. Awareness activities involve students discussing and highlighting language features in the text and discussing the connection between form(s) and meaning, language choice and differences to written language. A short sample of a simple class activity is given in appendix one to demonstrate this, using a dialogue from the 19th century CLiC corpus.

In summary, the results from this article suggest that dialogues have some potential for use as models of spoken language for use with learners of English as a second or foreign language. In examining 19th century literature, we are not of course advocating that this is the only or even the best form of literature to use in class but as we are unable to examine a corpus of modern-day literature, we have instead explored this potential in older texts.

Biodata

Christian Jones is a Senior Lecturer in Applied Linguistics and TESOL at the University of Liverpool. His main research interests are connected to spoken language and he has published a number of research papers related to spoken corpora, lexis, lexico grammar and instructed second language acquisition. He is the co-author (with Daniel Waller) of *Corpus Linguistics for Grammar: A guide for research* (Routledge, 2015), *Successful Spoken English: Findings from Learner Corpora* (with Shelley Byrne and Nicola Halenko) (Routledge, 2017) and editor of *Practice in Second Language Learning* (Cambridge University Press, 2018).

David Oakey is a Lecturer in Applied Linguistics and TESOL at the University of Liverpool. He has taught undergraduate and graduate classes in EAP writing, grammar, lexis and corpus linguistics at universities in China, Turkey and the USA and has published research on the phraseology of disciplinary and interdisciplinary discourse.

References

- Adolphs, S., & Carter, R. (2013.) *Spoken corpus linguistics: From monomodal to multimodal*. New York: Routledge.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Biber, D., & Conrad, S. (2009). *Real grammar: A corpus-based approach to English*. White Plains, NY: Pearson Longman.
- Byrne, S., & Jones, C. (forthcoming). The realism of conversation in literature. In C. Jones (Ed.), *Literature, spoken language and speaking skills in second language learning*. Cambridge: Cambridge University Press.
- Carter, R. (1998). Orders of reality: CANCODE communication, and culture. *English Language Teaching Journal*, 52(1), pp.43–56.
- Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English*. Cambridge: Cambridge University Press.
- Carter, R., & McRae, J. (Eds.) (1996). *Language, literature and the Learner: Creative classroom practice*. Oxon: Routledge.
- CLiC. (2018). Retrieved from www.CLiC.bham.ac.uk
- Coe, J. (2008). *The rain before it falls*. London: Penguin.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), pp. 29–62.
- Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal*, 58(4), 363-374. doi:10.1093/elt/58.4.363
- Hoover, D.L., Culpepper, J., & O'Halloran, K. (2014). *Digital literacy studies: Corpus approaches to poetry, prose and drama: The corpus, the computer and the study of literature*. London: Routledge.
- Jones, C., & Carter, R. (2012). Literature and language awareness: Using literature to achieve CEFR outcomes. *Journal of Second Language Teaching and Research*, 1(1), 69 –82.
- Jones, C. (2017). Soap operas as models of authentic conversations: Implications for materials design. In B. Tomlinson & A. Maley (Eds.), *Authenticity in materials development for language learning* (pp. 158-175). Newcastle: Cambridge Scholars Publishing
- Jones, C., Byrne, S., & Halenko, N. (2017). *Successful Spoken English: Findings from learner corpora*. London: Routledge
- King, G. (2014) *Colloquial English - the complete course for beginners* (2nd Edition) London: Routledge.

- Leech, G., & Short, M. (2007.) *Style in fiction: A linguistic introduction to English fictional prose* (2nd Edition). Harlow: Longman.
- Mahlberg, M., Stockwell, P., de Joode, J., Smith, C., & O'Donnell, B. (2016.) CLiC Dickens: Novel uses of concordances for the integration of corpus stylistics and cognitive poetics. *Corpora*, 11(3), 433–463.
- Mahlberg, M., & Smith, C. (2012.). Dickens, the suspended quotation and the corpus. *Language and Literature*, 21(1), 51–65.
- Mahlberg, M., & Wiegand V. (2018). Corpus stylistics, norms and comparisons: studying speech in *Great Expectations*. In R. Page, B. Busse., & N. Norgaard (Eds.) *Rethinking language, ext and Ccontext. Interdisciplinary Research in Stylistics in Honour of Michael Toolan* (pp. 123-143). New York & London: Routledge.
- McCarthy, M. (2010). Spoken fluency revisited. *English Profile Journal*, 1(1), pp. 1–15.
- McCarthy, M., & McCarten, J. (2018). Now you're talking! Practising conversation in second language learning. In C. Jones (Ed.) *Practice in second language learning* (pp 7 – 29). Cambridge: Cambridge University Press.
- McCarthy, M., & Carter, R. (1994). *Language as discourse: Perspectives for language teaching*. Harlow: Longman.
- McRae, J. (1991). *Literature with a small 'l'*. London: Macmillan.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J.C. Richards and R.W. Schmidt (Eds.) (pp. 191 – 227). *Language and communication*. New York: Longman.
- Rayson, P. (2018) *UCREL's Log-Likelihood and Effect Size Calculator*. Retrieved from: <http://ucrel.lancs.ac.uk/llwizard.html>
- Short, M. (1996). *Exploring the language of poems, plays and prose*. London: Longman.
- Short, M. (2012). Discourse presentation and speech (and writing, but not thought) summary. *Language and Literature*, 21(1), pp.18–32.
- Spoken BNC2014* (2018). Retrieved from <https://cqpweb.lancs.ac.uk/bnc2014spoken/>
- Snow, C. P. (1951). *The masters*. Harmondsworth: Penguin.
- Stockwell, P., & Mahlberg, M. (2015). Mind modelling with corpus stylistics in David Copperfield. *Language and Literature*, 24(2), 129 –147.
- Thornbury, S., & Slade, D. (2006). *Conversation: From description to pedagogy*. Cambridge: Cambridge University Press.
- Timmis. I. (2018). A text-based approach to grammar practice. In C.Jones (Ed.) *Practice in second language learning* (pp 79 –108). Cambridge: Cambridge University Press.



Journal of Second Language Teaching and Research. Volume 7

Wood, D. (Ed.) (2010). *Perspectives on formulaic language: Acquisition and communication*.
London: Continuum.

Appendix one

Sample activity using a literary dialogue

From Conan Doyle, A (1902.) *The Hound of the Baskervilles*. Extract from CLiC (2018)

'I have in my pocket a manuscript,' said Dr. James Mortimer.

'I observed it as you entered the room,' said Holmes.

'It is an old manuscript.'

'Early eighteenth century, unless it is a forgery.'

'How can you say that, sir?'

'You have presented an inch or two of it to my examination all the time that you have been talking. It would be a poor expert who could not give the date of a document within a decade or so. You may possibly have read my little monograph upon the subject. I put that at 1730.'

'The exact date is 1742.' Dr. Mortimer drew it from his breast-pocket. 'This family paper was committed to my care by Sir Charles Baskerville, whose sudden and tragic death some three months ago created so much excitement in Devonshire. I may say that I was his personal friend as well as his medical attendant. He was a strong-minded man, sir, shrewd, practical, and as unimaginative as I am myself. Yet he took this document very seriously, and his mind was prepared for just such an end as did eventually overtake him.'

Access:

1. Ask students how good they are remembering details when they see things.
2. Play 'Kim's game' in groups. Present students with a tray of objects for a few seconds and then cover it. Groups compete to remember the most objects and where they were placed.
3. Ask students to recall what they know about Sherlock Holmes' character e.g. he is clever/a good observer/he remembers things. Explain that you will be looking at a short dialogue which shows this.

Activity:

4. Give students the dialogue above to read. As they read, ask them to 'picture' the scene i.e. the room, the people in etc. They then describe that to each other and note differences.
5. Ask students some simple comprehension questions: what does Sherlock notice here? How? Why is Mortimer surprised?
What do you think is written on the document? Why do you think this might be important for the story? What do you think will happen next?
These are obviously open questions with no set answers



Awareness:

6. Underline all the examples of 'it' and 'that' in the conversation. When are they used to refer back to things already mentioned? What do they refer to? Do you use these items in the same way when you speak?
7. Underline the phrase which means 'I do not understand how you know that' (How can you say that?). When we use this phrase, how do we normally feel? (surprised or annoyed). Think of a situation where you might say this to someone. Do you have a similar expression in your first language? What is another way to say this? (How do you know that?)