



University of HUDDERSFIELD

University of Huddersfield Repository

McIntyre, Dan

Towards an integrated corpus stylistics

Original Citation

McIntyre, Dan (2015) Towards an integrated corpus stylistics. *Topics in Linguistics*, 16 (1). pp. 59-68. ISSN 1337-7590

This version is available at <http://eprints.hud.ac.uk/id/eprint/29616/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>



DOI: 10.2478/topling-2015-0011

Towards an integrated corpus stylistics

Dan McIntyre

University of Huddersfield, United Kingdom

Abstract

Over recent years, the use of corpora in stylistic analysis has grown in popularity. However, questions still remain over the remit of corpus stylistics, its distinction from corpus linguistics generally and its capacity to explain complex stylistic effects. This article argues in favour of an integrated corpus stylistics; that is, an approach to corpus stylistics that integrates it with other stylistic methods and analytical frameworks. I suggest that this approach is needed for two main reasons: (i) it is analytically necessary in order to fully explain stylistic effects in texts, and (ii) integrating corpus methods with other stylistic tools is what will distinguish corpus stylistics from corpus linguistics. My argument is supported by reference to examples from Mark Haddon's novel *The Curious Incident of the Dog in the Night-time* and the HBO TV series *Deadwood*. Both these examples rely for their explanation on a combination of corpus stylistic analytical techniques and other stylistic methods of analysis.

Keywords

Cognitive stylistics, corpora, corpus stylistics, *The Curious Incident of the Dog in the Night-time*, *Deadwood*, dialogue, direct speech, methodologies.

1. Introduction

The early 1990s saw the development and sudden growth of what have become two major areas of stylistics. These have come to be known commonly as corpus stylistics and cognitive stylistics. Research in these two areas has led to considerable insights into textual elements of style (see, for example, Ho, 2011; Bednarek, 2012; Malhberg, 2013a) as well as the nature of the reading process (e.g. Sandford and Emmott, 2012; Stockwell, 2013 and Harrison, *et al.*, 2014), and stylistics has undoubtedly been re-energized by their development. However, the popularity of the two areas brings with it potential problems for the discipline of stylistics as a whole. A glance at recent programmes for the annual conference of the Poetics and Linguistics Association (see, for example, PALA, 2013) shows that most papers are either cognitive or corpus-based in focus. That is, corpus stylistics and cognitive stylistics are no longer niche areas impacting on the mainstream; they *are* the mainstream. And because of this it is incumbent on us to ensure that the two

areas do not develop in isolation from each other. There is, certainly, a difference in focus for corpus and cognitive stylisticians: the former tend to be interested in deriving generalizations from the observation of patterns in large bodies of texts, while the latter are interested in the mechanics of individual readings. Nonetheless I want to argue that both areas have much to offer each other and that a stylistic analysis should not be restricted by the boundaries of a particular sub-discipline. Indeed, I would argue that seeing corpus and cognitive stylistics as sub-disciplines of stylistics is counterproductive. Rather, they are better seen as convenient labels for stylistic analysis which is focused on one particular area.

This article is primarily concerned with corpus stylistics and its future development. I argue that what is needed is an integrated corpus stylistics; that is, an approach to corpus-based stylistic analysis that takes account of all appropriate analytical frameworks (including but not limited to those from cognitive stylistics) and is not

restricted by only utilizing the tools and techniques of corpus linguistics. Indeed, for corpus stylistics to distinguish itself from corpus linguistics generally, it needs to incorporate theories, models and methods from qualitative stylistic analysis to augment computational techniques. I begin by considering existing definitions of corpus stylistics before going on to consider its relation to current work in cognitive stylistics. I then discuss two short texts, both of which rely on corpus analytical techniques for their explanation but which also necessitate insights from other areas of stylistics for a full elucidation of their stylistic effects.

2. Defining corpus stylistics

Despite for a long time enjoying only peripheral status within linguistics and literary studies, the practice of corpus stylistic analysis has grown in recent years, to the extent that corpus stylistics is fast becoming a recognizable field within stylistics generally. Evidence of this can be seen in the increasing number of encyclopaedia articles discussing its practice (see, for example, Mahlberg, 2013b and McIntyre, 2013) and in the small but growing number of monographs demonstrating the approach (e.g. Fischer-Starcke, 2007; Toolan, 2007; Ho, 2011; Mahlberg, 2013a; Hoover, et al., 2014 and Demjen, 2015). In addition, there are a number of books describing approaches to text analysis that would be recognized by most stylisticians as clearly fitting the general remit of this emerging area (see, for instance, Adolphs, 2006). However, there are a number of problems with many current definitions of corpus stylistics, not least of which is the tendency to define corpus stylistics rather narrowly as the analysis of literary texts using corpus linguistic techniques. This is the definition used by both Fischer-Starcke (2010) and Ho (2011), for example. However, if this is all that corpus stylistics is, then it is difficult to see a justification for the use of a distinct term to describe it. Under this definition, corpus stylistics is simply corpus linguistics with a different object of study (literature as opposed to non-literary language). This makes little sense, since corpus linguistic work that focuses on phonetic analysis (e.g. Honga, et al., 2014) is not commonly known as corpus phonetics (or corpus speech science), for instance. Nor is the kind of syntactic analysis that relies on corpus methods (e.g. Ai and

Lu, 2013) generally referred to as corpus syntax. Corpus linguistics as understood from the methodologist position (see Hardie and McEnery, 2010) is nothing more than a methodology for linguistic analysis and, as such, can be applied to any area of language study. That is, the same methods and analytical tools are used regardless of the sub-discipline of linguistics they are being applied to. Distinguishing corpus stylistics by its own specialist moniker therefore runs the risk of implicitly suggesting that stylistics is not a constituent sub-discipline of linguistics. The additional issue is that stylistics (of the corpus and non-corpus varieties) is not solely concerned with the analysis of literature. So distinguishing corpus stylistics from corpus linguistics on the grounds that its object of study is different makes little sense either. What this suggests is that for corpus stylistics to be a useful label, there must be something more to it than the simple application of corpus linguistic tools in the analysis of literary texts.

A second problem with many current definitions of corpus stylistics is an implicit assumption that traditional (i.e. non corpus-based) stylistics lacks rigour. Fischer-Starcke (2010), for example, discussing the application of corpus methods in the analysis of Jane Austen's prose, notes that:

New insights into the data can be gained since (1) the data is studied in a systematic and detailed way and (2) a larger number of units of meaning in language is analysed than in literary studies.

(Fischer-Starcke, 2010, p.11)

While Fischer-Starcke's second point is true (corpus tools do indeed afford the opportunity to identify units of meaning in language that would be difficult to unearth through manual qualitative analysis), her first point is an implicit misrepresentation of traditional stylistic methods. The aim of stylistics has always been to produce analyses of texts that are falsifiable, systematic and rigorous. The analytical checklists presented in Short (1996) and Stockwell (2010) are aimed squarely at ensuring systematicity, while Leech and Short's (1981) now classic *Style in Fiction* is a watchword for rigour and detail. The point is that corpus linguistic techniques are not necessary to make stylistic analysis systematic and detailed; good stylistics

should already be so. What corpus techniques offer are new ways of examining texts that can supplement insights gained via traditional methods (what Carter, 2010 affectionately calls 'steam stylistics'). Corpus stylistics is not in and of itself any more or less systematic than traditional stylistics, so systematicity cannot be seen as a distinguishing feature of the practice. To put it another way, systematicity is a necessary condition of corpus stylistics but not sufficient to distinguish it from any other kind of stylistics.

A third issue with many current definitions of corpus stylistics is one that extends to stylistics generally, and this is the definition of the discipline as the study of the language of literature. Stylistics emerged out of the work of the Russian Formalists and that of Charles Bally (1909) at the turn of the 20th century (Busse and McIntyre, 2010), and one of the earliest concerns of this movement was to isolate the linguistic properties of literary language. It was quickly discovered, however, that *literary language* is something of a misnomer, since there is nothing in the language of literary texts that is not found in myriad other text-types. There is, then, no reason why stylistics should not also be practised on non-literary texts; and indeed it is (see, for example, the early work of Crystal and Davy 1969, the critical stylistic work of Jeffries, 2007 and 2010, and the sociolinguistically inclined work of Coupland, 2007). Stylistics is best understood as the linguistic study of style (Leech and Short, 2007, p. 11) and how this can be affected by such non-linguistic variables as genre, author, historical period etc. (Jeffries and McIntyre, 2010, p. 1). While literature remains the primary object of study for most stylisticians, there is no intrinsic reason why this should be the case (indeed, it may be argued that this is a rather Anglocentric view; the Slavonic tradition of stylistics, for instance, is one in which the analysis of non-literary texts takes prominence). Consequently, the definition of corpus stylistics as the corpus linguistic study of literary language is unacceptably reductive. A suitable definition of corpus stylistics needs to take this into account while acknowledging that literature remains the object of study for many stylisticians.

All of this is to say that if corpus stylistics is to be useful as a descriptor, and if we are to make the case for it as having the kind of status that distinguishes it from mainstream corpus linguistics, we must be clear about

what it is that sets it apart from other corpus-related work. To this end, I define corpus stylistics as the application of theories, models and frameworks from stylistics in corpus analysis.

3. Corpus stylistics in relation to cognitive stylistics

The other dominant area of stylistics at the moment is cognitive stylistics. This has been (and continues to be) one of the most significant movements in stylistics over the last two decades, though its roots go back even further (West, 2010). Its popularity stems in part from the democratizing principles on which it is founded and which arise from its concern with how real readers respond to texts, including both literary and popular fiction. While cognitive stylistics aims to develop theories and models of the reading process that can be extrapolated to all readers, it is careful that such theories and models are able to explain individual readers' responses to texts. In this respect cognitive stylistics is concerned with the individual. Corpus stylistics, on the other hand, is concerned with discerning patterns in language use through the study of large quantities of language data, and while this may sometimes provide evidence for individual interpretations of specific texts (see, for example, Jeffries and McIntyre's 2010 corpus-informed stylistic analysis of the Roger McGough poem, 'italic'), it is not in and of itself concerned with accounting for the practices of individual readers. Rather, it is aimed at generalizing about linguistic behaviour beyond the sample studied. Consequently, corpus stylistics and cognitive stylistics are concerned with different (though often related) issues. This is perhaps one reason why corpus stylistics has to date made few inroads into cognitive stylistics; for reasons of focus, cognitive stylisticians have sometimes assumed that corpus methods have little to offer to the cognitive enterprise (this situation is changing; see, for example, Stockwell and Mahlberg, 2015). Added to this is the fact that doing corpus stylistics involves familiarizing oneself with corpus linguistic software (assuming, that is, that the analyst is not sufficiently skilled in programming to develop their own; see Gries, 2010 for a criticism of the over-reliance on commercially available software), principles of statistics, techniques of data collection and storage and data manipulation. It is, then, easy to see why many cognitive

stylisticians have taken the view that corpus stylistics is not for them. However, it should also be noted that corpus stylistics has, from some quarters at least, been reluctant to take on board insights from cognitive work. In some cases this reluctance has extended to outright hostility. This is a problem because it is often the case that without insights from other areas of stylistics, corpus stylistics can become simply an exercise in counting linguistic patterns, with no means of accounting for the interpretative significance of these. Indeed, one of the factors that has arguably made corpus stylistics unattractive to some stylisticians is the fact that some corpus stylistic 'analyses' fail to engage sufficiently at a functional level for meaningful interpretations of the data to be made. It is interesting to note that this is the kind of criticism that was made of some stylistics in its early days. Sinclair's (1966) stylistic analysis of Larkin's poem, "First Sight", for example, was heavily criticized for being mechanistic rather than interpretatively revealing (see Vendler, 1966 and Melia, 1974), a criticism that is sometimes now often levelled at contemporary stylistic work that makes use of corpus methods. Nonetheless, some corpus stylisticians have been similarly critical of cognitive stylistics for a number of reasons. One of the most vocal critics has been Louw, who has been vehement in his criticism of the cognitivists. Here he is being interviewed on the issue:

[...] cognitivists (Stockwell, 2002; Gavins and Steen, 2003) will tell you that you have a schema for *pubs* and that as a result you know that you can purchase food and drink in them. We accept this form of claptrap far too readily and the system behind it works because most of us will never see the *proof* of what only a corpus can show: that the most frequent collocates of *pub* are the terms *groups*, *chains* and *organisations*. Pubs organise our drinking habits and their own profitability, far more than anything else they do, and only a corpus will show you that.

(Louw, 2011, p. 179)

There are a number of problems with Louw's argument, however. First of all, evidence from the corpus that *pub* collocates with *groups*, *chains* and *organisations* does not, in and of itself, falsify the concept of a schema. A schema is a structured collection

of world knowledge stored in our long-term memory (Eysenck and Keane, 2010, p. 403-6) and while some specific concepts from schema theory (such as the notions of scripts and frames; see Jeffries and McIntyre 2010 for a summary) may be difficult to confirm empirically, experiments in psycholinguistics and neuroscience have found evidence for the existence of some kind of schematic structuring of world knowledge (see, for example, Bransford and Johnson, 1972 and Pratt, *et al.*, 2010). Evidence for the existence of a schema of some kind can also be found in the fact that when talking to someone from the same culture as us, there are a number of elements of world knowledge that we can take for granted, a pub being one of these. Put simply, we do not need to explain what pubs are when we use the word; we can instead simply assume that our addressee will possess the requisite background knowledge to be able to understand what we are talking about. Collocation, then, does not refute schema theory.

Louw's criticism of cognitive work extends beyond that concerned specifically with stylistics. Louw is seemingly dismissive of any endeavour which does not embrace empiricism, asking "What price are we to set upon 'theories' that specialise in telling you what you already know in the hope of keeping you away from the empiricism in which truth manifestly resides and is easily revealed?" (Louw, 2011). Louw is right in counselling against the unquestioning acceptance of claims that lack evidence. However, while empiricism can validate or invalidate such claims, in some cases so too can logic; that is, empiricism is not the be-all and end-all when it comes to demolishing mentalistic theories and models. Furthermore, the assumption that cognitive stylistics eschews empiricism is simply not true. While many cognitive theories are indeed difficult to test, significant work has been done on precisely this topic (see, for example, the work of Sandford and Emmott [2012] on testing the psychological reality of the processing of narratives). The sweeping aside of all non-empirically derived theories also seems short-sighted when we consider their value in other disciplines. Physics, for instance, has made substantial steps forward as a result of the postulation of non-empirically derived theories such as String Theory (see Polchinski, 1998).

There is, then, a danger in setting corpus stylistics in opposition to cognitive stylistics.

The two endeavours are not mutually exclusive and if we follow this path, we are likely to overlook significant insights that might be gained in one area and have relevance for the other. To this end, I would argue that neither corpus stylistics nor cognitive stylistics should be aiming for disciplinary isolation. Rather, they should be practised with the aim of developing a new mainstream in stylistics, so that it becomes inconceivable that a corpus stylistic analysis that would benefit from a cognitive dimension (or any other perspective) ignores this, and vice versa.

4. Direct speech in *The Curious Incident of the Dog in the Night-time*

To support my general argument, in this section I discuss an example of an extract from a novel that necessitates the use of corpus stylistic analysis for the explanation of its stylistic effects. But additionally, I want to make the point that a corpus stylistic analysis can only take us so far in understanding the effects associated with this particular text. In the case of this particular example, insights from a corpus stylistic analysis need to be integrated with a further stylistic concept, that Fowler (1977, 1996) terms *mind style*.

The Curious Incident of the Dog in the Night-time is a 2005 novel by Mark Haddon. The novel is narrated in the first person by Christopher, a teenager whom we assume from various contextual cues to have Asperger's Syndrome. This condition makes forming social relationships difficult for Christopher. The novel begins with Christopher's discovery of his next-door neighbour's dog, which has been killed with a garden fork. The police are called and their suspicion of Christopher panics him. Because he does not like being touched, he lashes out when one of the police officers attempts to grab his arm:

The policeman looked at me for a while without speaking. Then he said, "I am arresting you for assaulting a police officer."

This made me feel a lot calmer because it is what policemen say on television and in films.

Then he said, "I strongly advise you to get into the back of the police car, because if you try any of that monkey business again, you little shit, I will seriously lose my rag. Is that understood?"

I walked over to the police car, which was parked just outside the gate. He opened the back door and I got inside. He climbed into the driver's seat and made a call on his radio to the policewoman, who was still inside the house. He said, "The little bugger just had a pop at me, Kate. Can you hang on with Mrs. S. while I drop him off at the station? I'll get Tony to swing by and pick you up."

And she said, "Sure. I'll catch you later."

The policeman said, "Okeydoke," and we drove off.

(Haddon, 2005)

One of the intuitively striking elements of the above extract is the amount of direct speech in the passage. This quantity seems unusual but this is an unverifiable claim unless we use corpus techniques to check it. To this end, I annotated the novel for categories of speech, writing and thought presentation using a stylistic model based on that originally presented in Leech and Short (1981) and an annotation scheme for this model developed by Semino and Short (2004). The example below is of an annotated extract from the novel:

```
<dptag cat="N"> The policeman looked at me for a while without speaking.
</dptag> <dptag cat="NRS"> Then he said, </dptag> <dptag cat="xDS"> "I am arresting you for assaulting a police officer." </dptag> <dptag cat="N"> This made me feel a lot calmer </dptag> <dptag cat="N"> because it is what policemen say on television and in films. </dptag> <dptag cat="NRS"> Then he said, </dptag> <dptag cat="xDS"> "I strongly advise you to get into the back of the police car, because if you try any of that monkey business again, you little shit, I will seriously lose my rag. Is that understood?" </dptag>
```

The tags (i.e. everything within angle brackets) all follow the same format in order that they be retrievable using corpus software. Tags consist of an element *dptag* (i.e. "discourse presentation tag"), followed by an attribute *cat* (i.e. "category"). The attribute value is then placed within inverted commas; this is the relevant category of speech, writing and/or thought presentation. For example, the *N* attribute value indicates that everything following the tag is narration, while the attribute value "xDS"

indicates a stretch of direct speech (x is a placeholder which indicates that a particular slot is not necessary for recording the category in question). The full range of categories in the model are as follows:

Category	Descriptor	Speech presentation example
FD[S/W/T]	Free direct speech/writing/thought	I'm exhausted!
D[S/W/T]	Direct speech/writing/thought	He said, "I'm exhausted!"
FI[S/W/T]	Free indirect speech/writing/thought	He was exhausted!
I[S/W/T]	Indirect speech/writing/thought	He said that he was exhausted.
NP[S/W/T]A	Narrator's presentation of a speech/writing/thought act	He complained of tiredness.
NV/W/T	Narrator's presentation of voice/writing/thought	He droned on and on.
NR[S/W/T]	Narrator's report of voice/writing/thought [i.e. a reporting clause]	He said [...]

Having annotated the novel, I then used Multilingual Corpus Toolkit (Piao, *et al.*, 2002; available at <https://sites.google.com/site/scottpiaosite/software/mlct>) to extract all the direct speech. This amounts to 9 933 words from a total of 63 087 words in the novel as a whole. The next step was to compare these figures to those of a reference corpus of contemporary fiction. I used the similarly annotated fiction section of the Lancaster Speech, Writing and Thought Presentation Corpus (see Semino and Short, 2004). The corpus as a whole is a 260 000-word representative sample of news texts, fictional prose and (auto)biographical writing. The fiction section constitutes a representative sample of serious and popular fiction in English. In this section of the corpus there are 5 165 words of direct speech out of 87 570 words in total. Finally, I carried out a log-likelihood test to determine whether the difference in direct speech between *The Curious Incident of the Dog in the Night-time* and the reference corpus was statistically significant. The log-likelihood score was 3 499.46, above the cut-off value of 15.13 ($p < 0.0001$), indicating that there is indeed more direct speech in *The Curious Incident of the Dog in*

the Night-time than we find in novels generally.

This finding validates my initial hypothesis concerning the amount of direct speech in the novel but the corpus annotation techniques and statistical analysis will only take us so far in explaining the stylistic effects associated with this relative overuse of direct speech. In effect, we have described the source of one element of foregrounding in the novel but not evaluated its consequences. To do this, it is necessary to draw on a further concept from stylistics. This is Fowler's (1996) notion of mind style. Fowler defines mind style as "The world-view of an author, or a narrator, or a character, constituted by the ideational structure of the text" (Fowler, 1996, p. 21) and Semino (2007) has demonstrated how the narrator in *The Curious Incident* might be characterized as having an abnormal mind style as a result of his apparent inability to process metaphor. I would add to this that the abundance of direct speech in the novel can also be connected to Christopher's abnormal mind style, in that it suggests an inability to report speech using normal conventions (i.e. a lack of awareness of other possible discourse presentation categories, such as the Narrator's Presentation of Voice for propositional content that is not of primary importance to the narrative). Simply knowing that direct speech is comparatively overused in the novel, then, is not enough. The corpus stylistic element of the analysis needs to be elucidated by reference to additional stylistic concepts in order to explain its function in the text.

5. Dialogue in *Deadwood*

My example in this section comes from *Deadwood*, a TV series produced by HBO which ran from 2004 to 2006. Set in the American West in the 1870s, it was critically acclaimed and won plaudits particularly for the quality of the show's writing. The dialogue in particular is stylistically inventive. As Feeney (2004) puts it:

Milch's attempt to capture a sense of historical distance with the speech patterns of *Deadwood* succeeds marvelously, but not because the dialogue achieves true realism or gritty accuracy. *Deadwood*'s characters don't talk quite like us, but neither do they talk like Dakota scalawags in 1876 probably talked. (Feeney, 2004)

While it is easy to demonstrate that the dialogue in *Deadwood* is not like present day English, Feeney's claim that it is anachronistic is one that can only be validated using corpus techniques. However, as with *The Curious Incident of the Dog in the Night-time*, corpus analysis alone is not enough. One of the interesting issues surrounding *Deadwood* is the fact that, even though the dialogue is anachronistic, it has been highly praised by critics. That is, the anachronisms do not appear to get in the way of an immersive viewing experience. Elsewhere (McIntyre, 2015), I have demonstrated how credibility in fictional speech does not necessarily stem from authenticity (i.e. non-anachronistic dialogue). Here I want to draw on a cognitive approach to deixis known as deictic shift theory to further explain why the lack of authenticity in *Deadwood's* dialogue does not appear to cause problems for viewers' feelings of involvement in the fictional world. First, though, I will consider the issue of anachronisms. To illustrate how these are used in the series, consider the following extract from a scene which takes place in The Gem saloon:

[Context: Al Swearingen is the owner of The Gem and has a controlling interest in many of the town's businesses. Dan is his hired hand who has just collected rent from two men intending to set up a hardware store on one of Al's vacant plots. Ellsworth is a prospector who has just struck gold.]

1. **Al** 8 ounces of gold at \$20 an ounce is a 160, plus \$10 for a half-ounce is a 170 total.
2. **Ellsworth** Inform your dealers and whores of my credit, and pour me a goddamned drink.
3. **Al** Honor and a pleasure my good man. 170 credit, Dan, for Ellsworth.
4. **Dan** Yes, sir, 170 for Ellsworth. I'll let everybody know. Lot four, some hardware guys.
5. **Ellsworth** First one today with this hand. And pour me another, my good man.
6. **Al** Here comes another. Lot four a stayer?
7. **Dan** Wagon loaded with goods.
8. **Ellsworth** Now, with that Limey damn accent of yours, are these rumors true that you're descended from the British nobility?

9. **Al** I'm descended from all them cocksuckers.

10. **Ellsworth** Well here's to you, your majesty. I'll tell you what. I may've fucked my life up flatter than hammered shit, but I stand here before you today beholden to no human cocksucker. And workin' a payin' fuckin' gold claim. And not the U.S. government sayin' I'm trespassin' or the savage fuckin' red man himself or any of these limber dick cocksuckers passin' themselves off as prospectors had better try and stop me.

11. **Al** They better not try it in here.

12. **Ellsworth** Goddamn it, Swearingen, I don't trust you as far as I can throw ya, but I enjoy the way you lie.

13. **Al** Thank you, my good man.

14. **Ellsworth** You're welcome! You conniving, heavy thumbled motherfucker.

(*Deadwood*, Series 1, episode 1, my transcription)

The anachronisms in the extract above can only be identified with recourse to corpus data. Using the Corpus of Historical American English (COHA; <http://corpus.byu.edu/coha/>), a 385 000 000-word corpus covering the years 1810 to 2009, we can identify a range of both anachronistic and non-anachronistic words and phrases, including the following:

<i>Honor and a pleasure</i>	first attested 1870s
<i>Goddamn</i>	first attested 1910s
<i>Limey</i>	first attested 1910s
<i>Motherfucker</i>	first attested 1950s
<i>Cocksucker</i>	first attested 1960s
<i>Fucked * up</i>	first attested 1960s
<i>Trust * as far as I can</i>	first attested 1950s

The seeming authenticity of the dialogue is likely to stem from the fact that it does include words and phrases that were in use in the 1870s, such as *honour and a pleasure*. Even some of the apparent anachronisms (e.g. *goddamn* and *limey*) can be explained by the fact that COHA is composed of written texts, and written language lags behind the spoken language. Hence, something first attested in the 1910s is likely to have been in use in speech before then. Some of the words and phrases in the extract though are clearly anachronistic. The question then arises of why the anachronisms do not appear to affect viewers' perceptions of the fictional world of *Deadwood* as realistic. To understand this,

we need to consider the means by which we become immersed in a fictional world. As Sanford and Emmott (2012) point out:

To understand fiction one has to suspend disbelief, and suppose that the events being depicted actually occur. Only through automatic and possibly effortless suspension of disbelief is narrative immersion possible.

(Sanford and Emmott, 2012, p. 46)

Deictic shift theory (Duchan, et al., 1995; see also McIntyre, 2006; 2007) offers one explanation for this process of immersion. As Segal puts it, “[T]he metaphor of the reader getting inside the story is cognitively valid” (Segal, 1995, p. 14-15). According to deictic shift theory, we become immersed in a fictional world by taking up a position in a particular deictic field within the text world. And we move around the various deictic fields of the fictional world as and when we are directed by textual and contextual triggers. In an idealized reading situation, the reader pushes into a deictic field in the fictional world and the real world deictic field in which they exist decays as a result of not being regularly reinstated. In a non-ideal reading situation, on the other hand, readers are frequently reminded of their real-world deictic field, which prevents full immersion. I would suggest that the same cognitive procedures are undertaken by viewers as well as readers and that immersion in the fictional world is supported by realistic elements of *mise-en-scene* such as costumes and sets. Anachronistic dialogue, though, ought to act as a reminder of the real world, thereby reinstating our real world deictic field, causing the fictional world to decay and reducing our sense of immersion in the fiction. But this does not appear to happen for viewers of *Deadwood*. The reason, I suggest, is connected to the principled decisions that the writers of *Deadwood* made concerning their use of anachronisms.

Let us consider authentic taboo language of the 1870s. Attested taboo words of the period (from COHA) include *goldarn*, *darned*, *tarnation* and *gosh*. These are all minced oaths; that is, euphemisms derived from more pragmatically forceful taboo words, all of which are in this case religious in meaning. *Goldarn* derives from *God damn*, *darned* from *damned*, *gosh* from *God* and *tarnation* from *eternal damnation*. But as Hughes (2006) explains, “The force of the

traditional taboos against using religious oaths has generally diminished in modern times with the secularization of Western society” (Hughes, 2006, p. 389). According to Hughes, religious swearing has given way to sexual swearing with a consequent semantic weakening of previously taboo words. Since the original pragmatic force of the authentic taboo words of the period has been lost, using these terms in *Deadwood* would have been likely to create more of a comic effect, thereby detracting from the gritty realism of the drama. The writers’ decision to replace these with anachronistic sexual swearing preserves the pragmatic force of the archaic terms. The words themselves may be anachronistic but the pragmatic force is not. In this case, then, perhaps counterintuitively, use of authentic taboo language would have been more likely to reconstitute the viewer’s real world deictic field, since the weakened pragmatic force of the authentic lexemes would have been incongruous with the realistic *mise-en-scene* of the fictional world.

6. Conclusion

The examples from *The Curious Incident of the Dog in the Night-time* and *Deadwood* exemplify the need for corpus techniques in stylistics while also demonstrating that these need to be supplemented by other models and analytical frameworks. Without these, we are left with partial analyses. Indeed, I would argue that it is the integration of corpus techniques with non-corpus derived stylistic theories, models and frameworks that distinguishes corpus stylistics from general corpus linguistics. Such a combination of methods, I would suggest, ought to be a core principle of mainstream stylistics. Stylistics has always been eclectic in the sense of borrowing methods and frameworks from other disciplines (though not indiscriminately; see Jeffries and McIntyre, 2010: Chapter 1). Consequently, I would argue strongly in favour of a greater amalgamation of methods and frameworks developed *within* the discipline of stylistics. A state-of-the-art stylistic analysis ought not to be limited by the restrictions of a particular sub-area. Rather, an inclusive approach to stylistic analysis ought to expand the range of research questions it is possible to ask answer. These concerns are what motivate an integrated corpus stylistics.

References

- ADOLPHS, S., 2006. *Introducing electronic text analysis*. London: Routledge.
- AI, H. and LU, X., 2013. A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In: A. Díaz-Negrillo,, N. Ballier and P. Thompson, eds. *Automatic treatment and analysis of learner corpus data*, Amsterdam: John Benjamins, pp. 249-264.
- BALLY, C., 1909. *Traité de stylistique française*. Heidelberg, C. Winter.
- BEDNAREK, M., 2012. "Get us the hell out of here": key words and trigrams in fictional television series. *International Journal of Corpus Linguistics*, vol. 17, no.1, pp. 35-63.
- BRANSFORD, J. D. and JOHNSON, M. K., 1972. Contextual prerequisites for understanding: some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, vol. 11, pp.717-26.
- BUSSE, B. and MCINTYRE, D., 2010. Language, literature and stylistics. In: D. McIntyre and B. Busse, eds. *Language and Style*. Basingstoke: Palgrave, pp. 3-14.
- CARTER, R., 2010. Methodologies for stylistic analysis: practices and pedagogies. In: D. McIntyre and B. Busse, eds. *Language and Style: In Honour of Mick Short*. Basingstoke: Palgrave, pp. 34-46.
- COUPLAND, N., 2007. *Style: Language, variation and identity*. Cambridge: Cambridge University Press.
- CRYSTAL, D. and DAVY, D., 1969. *Investigating English style*. Bloomington: Indiana University Press.
- Deadwood*. (US) (2004-06) HBO. Writer: David Milch. Directors: Various.
- DEMJEN, Z., 2015. *Sylvia Plath and the language of mental states*. London: Bloomsbury.
- DUCHAN, J. F., BRUDER, G. A. and HEWITT, L. E., eds. 1995. *Deixis in narrative: A cognitive science perspective*. Hillsdale: Lawrence Erlbaum Associates.
- EYSENCK, M. W. and KEANE, M. T., 2010. *Cognitive psychology: A student's handbook*. 6th edn. New York: Psychology Press.
- FEENEY, M., 2004. Talk pretty: the linguistic brilliance of HBO's *Deadwood*. *Slate*: http://www.slate.com/articles/arts/television/2004/05/talk_pretty.html. Accessed 28 November 2015.
- FISCHER-STARCKE, B., 2007. *Corpus linguistics in literary analysis: Jane Austen and her contemporaries*. London: Continuum.
- FOWLER, R., 1977. *Linguistics and the novel*. London: Methuen.
- FOWLER, R., 1996. *Linguistics criticism*. Oxford: Oxford University Press.
- GRIES, S. TH., 2010. Corpus linguistics and theoretical linguistics: a love-hate relationship? Not necessarily.... *International Journal of Corpus Linguistics*, vol. 15, no. 3, pp.327-43.
- HADDON, N., 2005. *The curious incident of the dog in the night-time*. London: Jonathan Cape.
- HARDIE, A. and MCENERY, T., 2010. On two traditions in corpus linguistics, and what they have in common. *International Journal of Corpus Linguistics*, vol. 15, no.3, pp.384-94.
- HARRISON, C., NUTTALL, L., STOCKWELL, P. and YUAN, W., eds. 2014. *Cognitive grammar in literature*. Amsterdam: John Benjamins.
- HO, Y., 2011. *Corpus stylistics in principles and practice*. London: Continuum.
- HONGA, H., KIMB, S. and CHUNGA, M., 2014. A corpus-based analysis of English segments produced by Korean learners. *Journal of Phonetics*, vol. 46, pp. 52-67.
- HOOVER, D., CULPEPER, J. and O'HALLORAN, K., 2014. *Digital literary studies: Corpus approaches to poetry, prose and drama*. Abingdon: Routledge.
- HUGHES, G., 2006. *An encyclopedia of swearing: The social history of oaths, profanity, foul language, and ethnic slurs in the English-speaking world*. Armonk, NY: M. E. Sharpe.
- JEFFRIES, L., 2007. *Textual construction of the female body*. Basingstoke: Palgrave.
- JEFFRIES, L., 2010. *Critical stylistics*. Basingstoke: Palgrave.
- JEFFRIES, L. and MCINTYRE, D., 2010. *Stylistics*. Cambridge: Cambridge University Press.
- LEECH, G. and SHORT, M., 1981. *Style in fiction*. London: Longman.
- LEECH, G. and SHORT, M., 2007. *Style in fiction*. 2nd edition. London: Pearson.
- LOUW, B., 2011. Philosophical and literary concerns in corpus linguistics. In: V. Viana, S. Zyngier and G. Barnbrook, eds. *Perspectives on corpus linguistics*. Amsterdam: John Benjamins, pp. 171-96.
- MAHLBERG, M., 2013a. *Corpus stylistics and Dickens's fiction*. Abingdon: Routledge.

- MAHLBERG, M., 2013b. Corpus analysis of literary texts. In: C. A. Chapelle, ed. *Wiley Blackwell Encyclopedia of Applied Linguistics*. Oxford: Wiley Blackwell.
- MAHLBERG, M. and MCINTYRE, D., 2011. A case for corpus stylistics: Ian Fleming's *Casino Royale*. *English Text Construction*, vol. 4, no. 2, pp. 204-27.
- MCINTYRE, D., 2006. *Point of view in plays*. Amsterdam: John Benjamins.
- MCINTYRE, D., 2007. Deixis, cognition and the construction of viewpoint. In: M. Lambrou and P. Stockwell, eds. *Contemporary stylistics*, London: Continuum, pp. 118-30.
- MCINTYRE, D., 2012. Corpora and literature. In: C.A. Chapelle, ed. *Wiley Blackwell Encyclopedia of Applied Linguistics*. Oxford: Wiley Blackwell.
- MCINTYRE, D., 2015. Dialogue: credibility versus realism in fictional speech. In: V. Sotirova, ed. *The Bloomsbury companion to stylistics*. London: Bloomsbury.
- MELIA, D. F., 1974. Review of essays on style and language edited by Roger Fowler. *Foundations of Language*, vol. 11, pp. 591-94.
- PALA, 2013. *Annual conference of the Poetics and Linguistics Association*. University of Heidelberg, Germany. <http://tinyurl.com/njctuxk>
- PIAO, S., WILSON, A. and MCENERY, T., 2002. A multilingual corpus toolkit. In: AAACL 2002 Conference. Indianapolis, Indiana, USA.
- POLCHINSKI, J., 1998. *String Theory*. Cambridge: Cambridge University Press.
- SANDFORD, A. and EMMOTT, C., 2012. *Mind, brain and narrative*. Cambridge: Cambridge University Press.
- SEGAL, E. M., 1995. Narrative comprehension and the role of deictic shift theory. In J.F. Duchan, G.A. Bruder and L.E. Hewitt, L. E. eds. *Deixis in narrative: A cognitive science perspective*. Hillsdale: Lawrence Erlbaum Associates, pp. 61-78.
- SEMINO, E., 2007. Mind style 25 years on. *Style*, vol. 41, no.2, p. 153-73.
- SEMINO, E. and SHORT, M., 2004. *Corpus stylistics: Speech, writing and thought presentation in a corpus of English writing*. London: Routledge.
- SHORT, M., 1996. *Exploring the language of poems, plays and prose*. London: Longman.
- SINCLAIR, J., 1966. Taking a poem to pieces. In: R. Fowler, ed. *Essays on style and language*. London: Routledge, pp. 68-81.
- STOCKWELL, P., 2010. The eleventh checksheet of the apocalypse. In: D. McIntyre and B. Busse, eds. *Language and Style: In Honour of Mick Short*. Basingstoke: Palgrave.
- STOCKWELL, P., 2013. Creative reading, world and style in Ben Jonson's "To Celia". In: M. Borkenent, B. Dancygier and J. Hinnell, eds. *Language and the creative mind*. Chicago: University of Chicago Press.
- STOCKWELL, P. and MAHLBERG, M., 2015. Mind-modelling with corpus stylistics in David Copperfield. *Language and Literature*, vol. 24, no.2, p. 129-47.
- TOOLAN, M., 2007. *Narrative progression in the short story*. Amsterdam: John Benjamins.
- VENDLER, H., 1966. Review of essays on style and language edited by Roger Fowler. *Essays in Criticism*, vol. 16, no. 4, pp. 457-63.

Author's address and contact details

Dan McIntyre
University of Huddersfield
Queensgate
Huddersfield, HD13DH
United Kingdom
Phone: + 01484 478444
E-mail: d.mcintyre@hud.ac.uk