A corpus-based approach to mind style

DAN MCINTYRE and DAWN ARCHER

Abstract

Fowler's (1977) original definition of mind style emphasised consistency as a defining feature of the phenomenon, something that is (i) difficult to measure, and (ii) often missed in qualitative analyses. In this paper we investigate how a computational semantic analysis might be used to address this difficulty, with particular reference to McIntyre's (2005) analysis of the deviant mind style of the character of Miss Shepherd in Alan Bennett's play The Lady in the Van*. To do this we analyse the speech of all the characters in* The Lady in the Van *using Wmatrix (Rayson 2003, 2008), to see whether it provides quantitative support for the interpretative conclusions reached by McIntyre. Wmatrix utilises the UCREL Semantic Annotation System (USAS) which has been designed to undertake the automatic semantic analysis of English. The initial tag-set of the USAS system was loosely based on McArthur's* Longman Lexicon of Contemporary English *(McArthur 1981), but has since been considerably revised in the light of practical tagging problems met in the course of previous research, and now contains 232 category labels (such as* MEDICINE AND MEDICAL TREATMENT, MOVEMENT, OBLIGATION AND NECESSITY, *etc.). We use Wmatrix's facility for identifying key semantic domains in pursuit of our two main aims: (i) to determine whether Miss Shepherd's odd mind style is consistent, as Fowler's definition suggests it should be; and (ii) to determine the usefulness of computational semantic analysis for investigating mind style.*

1. The phenomenon of mind style

Mind style, according to Roger Fowler, with whom the term originated, is "the world-view of an author, or a narrator, or a character, constituted by the ideational structure of the text" (1996: 21). It differs from point of view as a result of mind style being unique to the individual. Research into mind style has tended to concentrate on uncovering the linguistic and paralinguistic means by

JLS 39 (2010), 167–182
DOI 10.1515/jlse.2010.009
0341-7638/10/039–167

which such unique world views are conveyed, and among the indicators of mind style that have been identified are transitivity patterns (see Halliday 1972; Bockting 1994a and Hoover 1999), conceptual metaphors (Semino and Swindlehurst 1996 and Semino 2002) and the idiosyncratic use of logical reasoning (McIntyre 2005). Some examples will illustrate the phenomenon. The first is from William Faulkner's novel *The Sound and the Fury*, an oft-quoted exemplar of mind style:

> Through the fence between the curling flower spaces, I could see them hitting. They were coming toward where the flag was and I went along the fence. Luster was hunting in the grass by the flower tree. They took the flag out, and they were hitting. Then they put the flag back and they went to the table, and he hit and the other hit. (Faulkner 1989: 3)

The first-person narration in the above example comes from Benjy, whom it transpires has some form of cognitive impairment. His handicap and the resultant conceptual difficulties that he experiences are conveyed through the unusual linguistic choices he makes. For example, his use of the transitive verb *hit* as if it were intransitive suggests a lack of awareness of cause and effect (Leech and Short 1981: 204–207).

Mind style conveyed through conceptual metaphor is apparent in the extract below from J. M. Barrie's *Peter Pan*:

> For a week or two after Wendy came it was doubtful whether they would be able to keep her, as she was another mouth to feed. Mr Darling was frightfully proud of her, but he was very honourable, and he sat on the edge of Mrs Darling's bed, holding her and calculating expenses, while she looked at him imploringly. She wanted to risk it, come what might, but that was not his way; his way was with a pencil and a piece of paper, and if she confused him with suggestions he had to begin at the beginning again. (Barrie 1911: 8)

Mr Darling's concern over whether or not it would be financially viable to keep his daughter is clearly an odd mind style and one which comes about as a result of the underlying conceptual metaphor (see Lakoff and Johnson 1980) CHILDREN ARE COMMODITIES, with all the attendant mappings between source and target domain (e.g. having a child is taking on a financial risk). What makes this doubly strange is the fact that both Mr and Mrs Darling appear to share this mind style, suggesting that it is the reader's mind style that is abnormal.

Mind style as a result of odd logical reasoning is illustrated by McIntyre (2005), in an analysis of the speech of the character Miss Shepherd, in Alan Bennett's play *The Lady in the Van*:

> [Context: Miss Shepherd is painting her van.]
> **Alan Bennett 1** What kind of paint are you using?
> **Miss Shepherd** The shade is crushed mimosa.
> **Alan Bennett 1** But it's gloss paint. You want car enamel.

Miss Shepherd Don't tell me about paint. I was in the infants' school. I won a prize for painting.

(Bennett 2000: 24)

Miss Shepherd's speech involves a premise (*I won a prize for painting while at infants' school*) and a conclusion (*therefore I know about paint*), but the problem, of course, is that this conclusion does not logically follow from the premise and is consequently inductively invalid (see McIntyre [2005: 28–32] for a fuller analysis of this extract from the play). McIntyre (2005) suggests that the logical leaps that Miss Shepherd makes throughout the play combine to convey a distinctly odd mind style that comes about as a result of guilt and paranoia, owing to her involvement (and culpability) in a fatal road accident many years before.

2. Investigating mind style using a quantitative approach

As the examples discussed above illustrate, mind style can be conveyed in a variety of different ways in a text. What most research into mind style has in common though is that it is qualitative in nature. This is true of most existing studies of the phenomenon (e.g. Halliday 1971; Fowler 1977; Leech and Short 1981; Bockting 1990; Black 1993; Bockting 1994a, 1994b; Semino and Swindlehurst 1996; Hoover 1999; Semino 2002; and McIntyre 2005), though Toolan (1990) discusses the issue from a quantitative perspective. However, if we return to Fowler's (1977) original discussion of mind style, we find that there are several constraints on his definition that would seem to belie a purely qualitative analytical approach. Fowler's (1977) definition of mind style makes the point that "[c]umulatively, consistent structural options, agreeing in cutting the presented world to one pattern or another, give rise to an impression of a world-view". The key point here is that for a mind style to be recognised as such, the techniques by which it is brought into being must be deployed consistently. The issue of consistency is something that purely qualitative analyses of mind style miss. This is unsurprising since in any such analyses there is the inherent problem of how to measure consistency. The difficulty, of course, is the issue of what to measure. Do we count the number of instances of a particular linguistic feature (say, instances of transitive verbs being used intransitively), and if so, against what norm would we compare such a figure? Clearly, such a method of measuring consistency would be both unsatisfactory and methodologically unsound. Measuring consistency then is a thorny problem in the study of mind style. We propose in this paper that it is possible to gain some measure of consistency by looking not at the number of instances of a particular indicator of mind style, but at the statistical significance of its occurrence within a text. We concentrate here on the potential for semantic domains to indicate mind style, and as a means of exploring the issue of consistency we

investigated the distribution of semantic domains across the whole text of Alan Bennett's play *The Lady in the Van*, using the UCREL Semantic Annotation System (henceforth USAS) developed at Lancaster University (see Rayson et al. 2004), which is described in Section 2.2. We paid particular attention to the semantic domains utilised by the character of Miss Shepherd. In doing this we were able to gain some measure of the usefulness of applying semantic domain analysis in the quantitative study of mind style, as well as being able to test some of the claims about Miss Shepherd's mind style made in McIntyre's (2005) qualitative analysis. In order to give some context to our analysis in Section 3, in the next section we provide a brief plot synopsis of *The Lady in the Van*, along with a summary of McIntyre's (2005) conclusions regarding the character of Miss Shepherd and her deviant mind style.

2.1. *The deviant mind style of Miss Shepherd*

Alan Bennett's highly acclaimed stage play *The Lady in the Van* (2000) is an autobiographical story of the playwright's somewhat odd relationship with Miss Shepherd, "the lady in the van" of the play's title. Bennett met Miss Shepherd in the late 1960s in Camden Town, London, where he was living at the time. Miss Shepherd, an eccentric by any definition, was living in a dilapidated van which she had parked in the street where Bennett lived. Following a warning from the police about being illegally parked, Miss Shepherd was invited by Bennett to park her van in his garden. There she stayed for the next 15 years, always maintaining a strangely formal and often antagonistic relationship with the playwright, and never directly expressing any gratitude for his generosity.

In the play (see McIntyre [2006] for a fuller synopsis) Miss Shepherd displays some markedly odd character traits, including paranoia, extreme indirectness and often surreal behaviour (though she herself perceives this as perfectly normal). Examples of her strange activities include writing to the College of Cardinals in Rome to suggest that a taller Pope might be more suited to the job, and suggesting that Bennett persuade the BBC to give her a radio phone-in programme on which she would dispense advice whilst sitting behind a curtain. McIntyre (2005) shows how such strange behaviour betrays a deviant mind style, and how that mind style is manifested linguistically as an inability to form inductively valid logical arguments. As the play progresses, it becomes apparent that Miss Shepherd's odd behaviour is due in part to the guilt she feels over her part in causing the death of a motorcyclist in a hit-and-run accident.

In a qualitative analysis of Miss Shepherd's mind style, McIntyre characterises her as (i) reluctant to reveal too much of her personal life (2005: 34), and (ii) unwilling to commit herself to any proposition, or to answer any questions directly (2005: 35). Our analysis of key semantic domains in Miss Shepherd's speech, and in *The Lady in the Van* more generally, was designed in part to see

if there was quantitative support for these two conclusions reached by McIntyre about Miss Shepherd's character as a result of this qualitative analysis of her mind style.

2.2. *USAS and Wmatrix*

Our analysis was carried out using Wmatrix (Rayson 2003, 2008), a web-based text analysis tool developed at Lancaster University. Wmatrix has most of the functionality of other text analysis tools, such as WordSmith Tools (Scott 2004), and enables the analyst to generate frequency lists, key word lists, n-grams and collocates for whatever text is uploaded. What sets Wmatrix apart from most other tools, however, is its capacity for the automatic semantic analysis of contemporary written and transcribed spoken British English texts (some testing has also been done on applying such analysis to older forms of English, with varying degrees of success). This is achieved using USAS (UCREL Semantic Analysis System; see Rayson et al. 2004), developed at Lancaster's University Centre for Computer Corpus Research on Language (formerly the University Centre for Corpus Research on the English Language, hence UCREL). USAS uses a semantic tag-set based originally on McArthur's (1981) *Longman Lexicon of Contemporary English* and further developed in later projects. The automatic semantic analysis of uploaded texts involves two stages. First, a part-of-speech tag is assigned to every lexical item or multiword expression (MWE), using probabilistic Markov models of likely part-of-speech sequences. This stage of the process has an accuracy rating of 97%. Once this has been achieved, the output is fed into SEMTAG, which assigns tags on the basis of pattern matching between the text and two computer dictionaries developed for use with the program. This stage of the process has a 92% accuracy rating, which necessitates some manual checks and revisions following an initial scan of the results. (Wmatrix offers the facility to create one's own 'dictionary'). Once a text is fully tagged, it is then possible to generate key word and key semantic domain lists by comparing the target corpus against a variety of reference corpora (samples from the BNC) stored in Wmatrix, or against any other corpus that has been uploaded to Wmatrix and tagged.

To analyse *The Lady in the Van* we began by creating a master electronic version of the whole play-text of *The Lady in the Van*. From this we created separate files for each individual character's speech. These files were then uploaded to Wmatrix and tagged for parts-of-speech and semantic domains. Following the automatic tagging, the texts were checked manually for inconsistencies and any tagging errors were rectified. There is, of course, a degree of subjectivity to this process – particularly in deciding what alternative semantic domain to shift an incorrectly assigned item to. Nonetheless, we would argue that since manually assigning incorrectly tagged items to new semantic

domains is done *before* generating initial results, this does not bias the analytical process. Using Wmatrix we were then able to undertake a semantic analysis of the speech of Miss Shepherd, focusing on frequent and key semantic domains as well as key words for her character.

3. Analysis of results

In order to see whether there was any quantitative evidence of Miss Shepherd demonstrating a deviant mind style we looked at a number of different aspects of her speech. We began by compiling a list of the most frequent semantic domains in Miss Shepherd's speech.

The point of this exercise was simply to see whether potential areas of analytical interest might be indicated in such a frequency list, and whether the frequent domains might suggest any support for the qualitative claims made in McIntyre (2005).

Table 1. *Most frequent semantic domains in Miss Shepherd's speech*

Domain	Examples
1. GRAMMAR BIN	*the, a, to, of, and, in*
2. PRONOUNS	*I, it, you, that, me, they*
3. EXISTING	*was, be, is, 'm, were*
4. LIKELY	*could, possibly, may*
5. NEGATION	*n't, not, nothing*
6. POSSESSION	*have, get, take, keep*
7. ACTIONS	*make, doing, banging*
8. COMING/GOING	*go, left, went, pilgrimage*
9. LOCATION	*there, here, outside*
10. PERSONAL NAMES	*Mr Bennett, Susie Wong*
11. SPEECH ACTS	*say, tell, name, says*
12. WANTING, PLANNING	*want, wanting, hope*
13. RELIGION/SUPERNATURAL	*catholic, soul, priest*
14. PLACE NAMES	*Holloway, France*
15. DISCOURSAL	*no, yes, god, oh*
16. MINIMISERS	*only, just, alone*
17. TRANSPORTATION (LAND)	*van, tyres, car*
18. TIME: FUTURE	*will, going to, shall*
19. OBLIGATION/NECESSITY	*should, have to, need*
20. PULLING/PUSHING	*put, push, taken away*
21. COMMUNICATION	*speak, voice, reeled off*
22. UNKNOWN TERMS	*fidelis, guerre, j'ai*
23. OBJECTS	*bell, stick, wheels, lid*
24. QUANTITY	*some, any, half*
25. CLOTHES/BELONGINGS	*coat, clothes, towel*
26. KNOWLEDGE	*know, knew, knowledge*

The table above gives some indication of specific topics that appear to be important to Miss Shepherd. The fact that the domain LIKELY appears high on the list is particularly interesting and suggests that it may indeed be worth looking at this domain in more detail, to see whether it reveals quantitative support for the notion of Miss Shepherd as a character often unwilling to commit to any proposition. Of course, a simple frequency list is not enough to count as evidence for such a claim, and so to gain a more reliable measure of the statistical significance of frequent semantic domains in Miss Shepherd's speech, we undertook an analysis of the key words and key domains; that is, those words and semantic domains which are over- or under-represented in the target data when compared against a reference corpus.

3.1. *Key words and key semantic domains in Miss Shepherd's speech*

In order to generate a list of those words and semantic domains which are key in Miss Shepherd's speech, we compared her dialogue against the BNC Written Imaginative sampler, an extract from the British National Corpus consisting of 222,541 words of fiction. There are, of course, always issues in choosing an appropriate reference corpus. The BNC Written Imaginative corpus is appropriate in the sense that it is composed of fiction texts, and in this sense it fulfils Culpeper's (2009) criterion that the closer the reference corpus to the target corpus in terms of content, the more likely it is that the key items generated will reveal issues specific to the target text/s. However, it is also the case that the BNC Written Imaginative sampler is not composed solely of dramatic fiction (this, of course, would be the ideal reference corpus and for this reason, we also experimented with comparing Miss Shepherd's speech against that of all the other characters in the play; see Section 3.2 for a discussion of these results).

Comparing Miss Shepherd's speech against the BNC Written Imaginative sampler produced a key word list in which 83 items had a statistical significance above the log-likelihood critical value of 15.13 ($p < 0.0001$, indicating 99.99% confidence of significance). As a further test we also compared Miss Shepherd's dialogue against that of the rest of the characters in the play, aggregated. This produced a key word list in which 26 items had a statistical significance above the log-likelihood critical value of 6.63 ($p < 0.01$, indicating 99% confidence of significance). The key word list generated through comparison with the BNC Written Imaginative sampler is presented in Table 2 below. Italicised words are those which are also key when Miss Shepherd's speech is compared against that of the rest of the characters in the play. The number in parentheses after such words indicate their position in this second key word list.

Before we move on to examining key semantic domains, there are a number of observations that we can make about the key words themselves. At the top of the list is the proper noun *Mr Bennett*, suggesting that this is a character who

Table 2. *Miss Shepherd's key words compared against BNC written imaginative sampler*

Key word	LL (critical value = 15.13)	Key word	LL (critical value = 15.13)
1. *Mr_Bennett (1)*	209.00	43. residents	18.77
2. *possibly (2)*	186.18	44. be	18.32
3. *I (3)*	167.28	45. *could (10)*	17.34
4. van	123.91	46. write	17.19
5. *may (4)*	77.34	47. curtain	16.58
6. *want (9)*	74.73	48. rubber	16.58
7. *'ve (5)*	56.15	49. sin	16.58
8. *catholic (6)*	54.18	50. soul	16.33
9. *it (26)*	49.76	51. 1964	15.48
10. do	38.98	52. Albany_Street	15.48
11. *pilgrimage (13)*	38.70	53. Downing_Street	15.48
12. you	34.99	54. Khrushchev	15.48
13. *in_question (11)*	33.34	55. Liebestraum	15.48
14. *parking (12)*	33.34	56. St_Albans	15.48
15. 'm	32.52	57. Susie_Wong	15.48
16. *thee (23)*	30.96	58. Tunbridge_Wells	15.48
17. n't	30.20	59. ambulances	15.48
18. *manure (21)*	26.00	60. amendment	15.48
19. *got (16)*	25.08	61. battery	15.48
20. priest	25.05	62. bona_fide	15.48
21. *rights (14)*	25.05	63. cardinals	15.48
22. have	24.04	64. college	15.48
23. *pencil (22)*	23.41	65. communist	15.48
24. Parkway	23.22	66. elderly	15.48
25. St_John	23.22	67. guerre	15.48
26. disabled	23.22	68. heaven	15.48
27. electricity	23.22	69. ignorance	15.48
28. en	23.22	70. next_of_kin	15.48
29. fidelis	23.22	71. nomination	15.48
30. incognito	23.22	72. papal	15.48
31. lady-seller	23.22	73. pencils	15.48
32. novices	23.22	74. phone-in	15.48
33. petrol	23.22	75. re	15.48
34. reliant	23.22	76. recipe	15.48
35. only	22.91	77. scruples	15.48
36. anonymous	21.53	78. this_minute	15.48
37. *justice (19)*	21.53	79. times	15.48
38. besides	20.83	80. towels	15.48
39. key	19.29	81. vaccinated	15.48
40. convenient	18.77	82. wheelchair	15.48
41. ivy	18.77	83. yellow_lines	15.48
42. nuns	18.77		

is important to Miss Shepherd in some sense. What is noteworthy here is Miss Shepherd's use of title plus surname as opposed to first name to refer to Alan Bennett. In terms of social deixis, this is perhaps indicative that the relationship Miss Shepherd has with Alan is not a close one. With regard to the other proper nouns in the list, one refers to a Saint (*St John*), while the other (*Susie Wong*) turns out not to be participant-related at all, but is Miss Shepherd's rhyming slang for *pong* (i.e. smell). There is, however, a participant-related phrase (*next of kin*) that also has formal connotations, and we might hypothesise that formality as a value is important to Miss Shepherd. We might link this indirectly to the claim made in McIntyre (2005) that Miss Shepherd is a character reluctant to reveal too much about herself. For such a character, a degree of formality seems appropriate.

With further regard to McIntyre's (2005) claims about Miss Shepherd, it is interesting to note that *possibly*, *may* and *could* appear in both key word lists and perhaps relate to the notion that Miss Shepherd is a character unwilling to commit unconditionally to any proposition.

Other observations about the key word list include the fact that a number of words relate to crime and law and order (*Holloway* [prison], *justice* and *rights*), and religion (*God*, *Catholic*, *priest*, *pilgrimage*, *soul*). This is noteworthy when we consider that at the heart of Miss Shepherd's torment is the guilt that she feels at having abandoned the scene of the accident in which she inadvertently killed a motorcyclist. McIntyre (2005) argues that this guilt is the source of Miss Shepherd's odd mind style, and the statistical significance of these related key words confirms that these are important issues for Miss Shepherd.

If we turn now to an analysis of key domains, we find that the following semantic fields are key in Miss Shepherd's speech when compared against the BNC Written Imaginative sampler:

Table 3. *Miss Shepherd's key semantic domains compared against BNC written imaginative sampler*

Domain	LL (critical value = 15.13)
1. LIKELY	110.72
2. RELIGION AND THE SUPERNATURAL	60.29
3. VEHICLES AND TRANSPORT ON LAND	54.89
4. WANTED	40.12
5. EXISTING	27.18
6. PRONOUNS	26.75
7. EXCLUSIVIZERS/PARTICULARIZERS	21.87
8. NEGATIVE	19.56
9. ETHICAL	17.84
10. GETTING AND POSSESSION	17.76
11. LIGHT	15.48

The first key domain suggests clear support for the claim made in McIntyre (2005) that part of Miss Shepherd's odd mind style is a reluctance to commit to the propositions she makes. The domain LIKELY contains 156 instances of modal expressions (including verbs, adverbs and nouns). These are *can*, *can't*, *could*, *make sure*, *may*, *might*, *possibility*, *possible*, *possibly*, *probably* and *would*. Of these, *possibly*, *may* and *could* are also key words. What this suggests is that modality is indeed a key feature of Miss Shepherd's speech, to the extent that she over-uses modal expressions. We can relate the statistical significance of Miss Shepherd's use of modality to Fowler's (1977) claim that consistency is a defining feature of mind style. For example, one deviant modal usage by a character may not in and of itself indicate an unusual mind style, but in qualitative analysis there is the danger that such an instance of foregrounding would be seized on by the analyst as just such an indicator. Fowler's (1977) argument is that such a usage would not be indicative of an aberrant mind style because it is not a consistent usage. While it may be difficult to measure the extent to which such a usage is spread across a complete text, the fact that it is statistically over-used is indicative of it being abnormal; and this is clearly an important aspect of the construction of deviant mind styles. (In the case of keywords, it is possible, using some concordancing packages (e.g. WordSmith Tools), to generate dispersal plots to gain a better sense of how the target word is spread across the corpus as a whole. Currently it is not possible to do this for semantic domains though this would undoubtedly be a useful additional function for Wmatrix).

Second on the list of key domains is RELIGION AND THE SUPERNATURAL. This contains 48 items, included in which are the key words *God*, *Catholic*, *priest*, *pilgrimage*, *soul*. The keyness of this domain provides support for the notion that Miss Shepherd is obsessed with religion (to the extent that she over-uses words from this semantic field), and this seems to provide appropriate support for the idea that guilt (note its importance in the popular notion of Catholicism) is the source of her deviant linguistic behaviour.

The third item on the key domain list is VEHICLES AND TRANSPORT ON LAND, which supports the notion that Miss Shepherd is obsessed with motoring. The significance of this is not simply related to the fact that Miss Shepherd herself drives a van. A look at a sample of concordance lines from this domain reveals her to be obsessed with the rules of road. It is not difficult to argue that this is a direct result of the hit-and-run accident which is the source of her odd behaviour:

through being stationary on the	carriageway	. Where do you live ? I could pa
easure the distance between the	tyres	and the kerb . One and a half in
tance between the tyres and the	kerb	. One and a half inches is the i
thes that did it . I was on the	station	and the policeman thought I was
? I 'm thinking of painting the	van	. One of those little mop things
hild . And mumps . What is it ?	Yellow lines	. In the street . They wanted me
t is it ? Yellow lines . In the	street	. They wanted me to shift the va

t . They wanted me to shift the	van	so that they could make the line
I am . So now they ‘ve put the	yellow lines	as far as the van and started th
the yellow lines as far as the	van	and started them the other side
ts ‘ parking and kept it in the	street	I ‘m liable to be attacked again
hearing of it and seeing of the	bodywork	. The word accident was mentione
f justice would jump heavily on	drivers	making a noise in the evening .
need insuring . It ‘s like the	van	. I ‘m insured in heaven . I sha
anything . I ‘m an experienced	driver	. I drove ambulances in the blac
I ‘m an experienced driver . I	drove	ambulances in the blackout . You
dy-seller was stationary in her	vehicle	and scrupulous as thy servant ha
, contrives to collide with the	van	. As was claimed , fatally . The
uld either of you object if the	van	were to become a place of pilgri
possibly ? I am thinking of the	car	that Catholic priest was murdere
en murdered sometimes . Say the	van	were left on site , that would e

Of the remaining key domains, none appear to provide explicit support for the two claims under specific investigation here: namely, that Miss Shepherd is unwilling to commit to the propositions she expresses, and that she is reluctant to reveal anything about her personal life. Nonetheless, some are revealing of other aspects of Miss Shepherd’s odd character which are indirectly related to the deviant mind style that she conveys in the play. WANTING consists primarily of the lexical verb *want* in various forms, the most frequent subject of which is *I*. One effect of this is to characterise Miss Shepherd as unduly demanding. EXISTING is composed primarily of copular verbs, and again most of these have *I* as a subject, which might lead us to an interpretation of Miss Shepherd as self-obsessed (cf. self-revealing). This is supported by the preponderance of first-person pronouns in the PRONOUNS domain (326 of 835 instances, or 39%, though we should view this as tentative support rather than confirmation, since the make-up of the reference corpus, being not solely composed of direct speech, may well be responsible for PRONOUNS being key in Miss Shepherd’s speech). EXCLUSIVERS consists primarily of *only* and *just*, which we might see as related to Miss Shepherd’s propensity for making excuses about her behaviour. NEGATIVE consists entirely of instances of *not* and *nothing*, and again might be seen as characterising Miss Shepherd as pessimistic, which is appropriate given her other characteristics. Finally, GETTING AND POSSESSION includes verbs such as *have*, *take* and *keep*, and while there may not be a clear link to aspects of Miss Shepherd’s mind style, there is perhaps a connection to aspects of her personality when we note that many of these instances again have *I* as subject.

3.2. *Comparison of Miss Shepherd’s speech with that of other characters*

In determining the source of Miss Shepherd’s deviant mind style quantitatively we have inevitably focused primarily on the dialogue of her character. Nonetheless, other insights can be gained by comparing her speech against that of the other major characters in the play. Table 4 below details the top ten most key domains ($p < 0.0001$) for the characters Alan Bennett 1 (abbreviated to AB1; this is Alan Bennett the character in the text world that Miss Shepherd

Table 4. *Key domains for the major characters in the play (log-likelihood values decrease moving from left to right across the table)*

Character	Key semantic domains (reference corpus = BNC Spoken DG sampler									
Miss Shepherd	RELIGION AND THE SUPERNATURAL	LIKELY	VEHICLES AND TRANSPORT ON LAND	WANTED	MUSIC AND RELATED ACTIVITIES	SPEECH ACTS	POLITICS	ETHICAL	LAWFUL	GRAMMATICAL BIN
Alan Bennett 1	VEHICLES AND TRANSPORT ON LAND	SPEECH ACTS	CAUSE & EFFECT: CONNECTION	ALIVE	FARMING & HORTICULTURE	MUSIC & RELATED ACTIVITIES	TIME: GENERAL	DISLIKE	MEDICAINES AND MEDICAL TREATMENT	DEAD
Alan Bennett 2	ALIVE	PEOPLE: FEMALE	GENERAL APPEARANCE AND PHYSICAL PROPERTIES	GRAMMATICAL BIN	SENSORY: SMELL	DEGREE	FARMING & HORTICULTURE	DEAD	VEHICLES AND TRANSPORT ON LAND	EVALUATION: FALSE
Social Worker	SPEECH ACTS	MEDICAINES AND MEDICAL TREATMENT	FORMAL/ UNFRIENDLY							

inhabits), AB2 (Alan Bennett, the narrator of the play who exists in a framing fictional world), Mam (Alan Bennett's mother), the Social Worker, Rufus (a neighbour), Underwood (who is blackmailing Miss Shepherd) and Leo Fairchild (Miss Shepherd's brother). The domains with the highest log-likelihood values are towards the left-hand side of the table and decrease in value towards the right (though all are statistically significant at $p < 0.0001$). Here we have compared each character's speech against the BNC Spoken Demographic sampler of 501,953 words. This gives us an alternative perspective on the key domains for these characters. The lists of key items discussed in Section 3.1 were generated using the BNC Written Imaginative sampler as a reference corpus, since there we were concerned with determining the extent to which the playwright had manipulated language in order to create a foregrounding effect. Here, we are more concerned with how the characters' speech compares with that of real speakers, hence our use of this larger reference corpus. (Note that this changes the rank ordering of some of Miss Shepherd's domains, and brings in some which were not included on the initial list. As it transpires, these do not affect the status of our arguments so far; Culpeper 2009 discusses the pros and cons of using different reference corpora, and Scott 2009 suggests that there is no significant advantage in using one over another).

In terms of results, we can note that for the Social Worker, the key domains are quite typical of what we would expect given her profession. What stands out, however, is the FORMAL/UNFRIENDLY domain. On further investigation, it turns out that most of these occurrences are the Social Worker complaining about the hostility she feels from Alan. This provides an interesting counterpoint to the notion that Alan Bennett is the moral centre of the text world, against whose normal behaviour Miss Shepherd's is judged as abnormal.

AB1 (Alan Bennett the character) and AB2 (Alan Bennett the narrator) share some similarities, though there are also differences between them. Both discuss life and death (in the domains ALIVE and DEAD which may potentially relate to thematic issues in the play, a further literary issue that Wmatrix can assist in uncovering). AB2, however, is particularly preoccupied with smell (SENSORY: SMELL) and appearance (GENERAL APPEARANCE AND PHYSICAL CHARACTERISTICS). The latter category is explained by AB2's narratorial function, which incorporates a substantial amount of scene-setting ('telling' as opposed to 'showing'), while the former is constituted entirely of references to Miss Shepherd's unpleasant odour. The PEOPLE: FEMALE category highlights the significance of women in Alan Bennett's life (his mother, Miss Shepherd and the Social Worker), and particularly the fact that these are all women who cause him problems of various kinds. This inevitably impacts on our perception of Miss Shepherd. EVALUATION: FALSE relates entirely to Alan's scepticism about Miss Shepherd, and since the audience of the play is set up to view events from Alan's ideological perspective (see McIntyre 2006 for this

argument), this further reinforces our sense of Miss Shepherd as being somehow abnormal. While the other categories are often character-related (DEGREE, for instance, is potentially character-related in that it is constituted entirely of *as*, which might be seen as a style marker for AB2), we are focusing here on those categories which pertain to the indirect characterisation of Miss Shepherd.

Of AB1's domains that are indicative of Miss Shepherd's character, the most important appears to be MUSIC AND RELATED ACTIVITIES. The contents of this domain occur when Alan engages Miss Shepherd in a conversation about music, at which juncture she becomes irrationally angry. The significance of this is that Miss Shepherd has in effect led two lives. Prior to the hit-and-run accident that she was responsible for, she had trained as a classical pianist. After the accident, she has no possibility of returning to that life and cannot even bear to be reminded of it, such is the pain that this causes her. Here again we see an abnormal reaction which is consistent with her abnormal mind style.

4. Conclusions and future work

In this article we set out to explore how a computational semantic approach might be used to measure the consistency of mind style within a text. In discussing the conclusions that might be drawn from such research it is worth first stating a caveat that should be borne in mind before undertaking such work. In and of itself, a computational semantic analysis cannot uncover features of mind style in a text. An analysis of key domains exhibited in a character's speech or in narration may indicate a potentially deviant mind style, but in order to confirm this, close textual analysis of a qualitative nature is needed. The kind of quantitative analysis we have reported on here, then, is best seen as a support to a qualitative analysis. The other advantage of the approach we have described in this paper is that it can provide a way to look at the relationships and differences between characters in a text (be this a play, as we have analysed, or a prose text), which may lead, in turn, to the uncovering of deviant linguistic behaviour and the subsequent appraisal of such behaviour as constituting a mind style. The notion of keyness is of particular importance when we consider one of the main problems with identifying mind styles in texts. Leech and Short (1981) summarise this when they say that, despite the notion of mind style existing on a cline (with normal linguistic behaviour at one end and deviant linguistic behaviour at the other), it is necessarily the case that only those mind styles which are markedly odd or which we notice as being strikingly different from our own that are picked up on in qualitative analysis. In theory, the computational approach to mind style described in this paper could provide a means of uncovering those mind styles which would be less noticeable in a qualitative analysis. This is because the keyness or log-likelihood score of a

particular domain gives an indication of the extent to which the domain in question is "abnormal" (be this as a result of being overused or underused) and, by so doing, allows us to identify nuances of linguistic behaviour that it is not always possible to recognise in a purely qualitative analysis.

Our analysis of key words and key semantic domains in the speech of Miss Shepherd does provide quantitative support for some of McIntyre's (2005) claims about her mind style. The fact that the semantic domain LIKELY rated so highly in terms of its log-likelihood value confirms and provides quantitative evidence for McIntyre's (2005) conclusion that Miss Shepherd is a character very unwilling to commit firmly to any given proposition. The contents of this domain are all modal expressions that indicate some lack of confidence in the degree of commitment to the assertion being made. This is clearly one way in which her intensely secretive nature is exhibited. Similarly, McIntyre's claim about the influence of religion (namely her Catholic guilt) on Miss Shepherd's mind style is borne out by the fact that RELIGION is the most key semantic domain in Miss Shepherd's speech. We can thus be reasonably confident in stating that the mind style exhibited by Miss Shepherd is consistent, in that those elements of Miss Shepherd's speech identified by McIntyre (2005) as contributing to the creation of her mind style are, in fact, statistically significant within the whole text of *The Lady in the Van*.

There remains, of course, much more work to be done in order to develop the use of quantitative analysis for the study of mind style. We are, for example, currently investigating how an analysis of key semantic domains can be used to assess the cumulative nature of mind style. Testing across a greater range of texts would hopefully provide further insights into how such a quantitative approach as that described in this paper might be usefully employed in the study of mind style.

University of Huddersfield
University of Central Lancashire

Note

Correspondence addresses: d.mcIntyre@hud.ac.uk; dearcher@uclan.ac.uk

References

Barrie, J. M. 1911. *Peter Pan*. London: Hodder and Stoughton.
Bennett, Alan. 2000. *The lady in the van*. London: Faber and Faber.
Black, Elizabeth. 1993. Metaphor, simile and cognition in Golding's *The inheritors*. *Language and Literature* 2. 37–48.

Bockting, Ineke. 1990. The impossible world of the schizophrenic: William Faulkner's Quentin Compson. *Style* 24(3). 484–497.

Bockting, Ineke. 1994a. Mind style as an interdisciplinary approach to characterisation in Faulkner. *Language and Literature* 3(3). 157–174.

Bockting, Ineke. 1994b. *Character and personality in the novels of William Faulkner: A study in psychostylistics*. Amsterdam: University of Amsterdam Press.

Culpeper, Jonathan. 2009. Keyness: words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet. International Journal of Corpus Linguistics* 14(1). 29–59.

Faulkner, William. 1989 [1931]. *The sound and the fury*. London: Pan.

Fowler, Roger. 1977. *Linguistics and the novel*. London: Methuen.

Hoover, David. 1999. *Language and style in 'The inheritors'*. Lanham: University Press of America.

Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: Chicago University Press.

Leech, Geoffrey N. & Michael H. Short. 1981. *Style in fiction*. London: Longman.

McArthur, Tom. 1981. *Longman lexicon of contemporary English*. London: Longman.

McIntyre, Dan. 2005. Logic, reality and mind style in Alan Bennett's *The Lady in the van*. *Journal of Literary Semantics* 34(1). 21–40.

McIntyre, Dan. 2006. *Point of view in plays: A cognitive stylistic approach to viewpoint in drama and other text-types*. Amsterdam: John Benjamins.

Rayson, Paul. 2003. *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. Lancaster: Lancaster University PhD thesis.

Rayson, Paul. 2008. *Wmatrix: A web-based corpus processing environment*. Computing Department: Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/ (accessed 07/08/2010).

Rayson, Paul, Dawn Archer, Scott Piao & Tony McEnery. 2004. The UCREL semantic analysis system. *Proceedings of the Workshop on Beyond Named Entity Recognition Semantic Labelling for NLP Tasks, in association with the 4th International Conference on Language Resources and Evaluation* (LREC 2004), Lisbon, Portugal, 7–12.

Scott, Mike. 2009. In search of a bad reference corpus. In Dawn Archer (ed.), *What's in a word list? Investigating word frequency and keyword extraction*, 79–91. Aldershot: Ashgate.

Semino, Elena. 2002. A cognitive stylistic approach to mind style in narrative fiction. In Elena Semino & Jonathan Culpeper (eds.), *Cognitive stylistics: Language and cognition in text analysis*, 95–122. Amsterdam: John Benjamins.

Semino, Elena & Kate Swindlehurst. 1996. Metaphor and mind style in Ken Kesey's *One flew over the cuckoo's nest*. *Style* 30(1). 143–166.