Open Research Online
The Open University's repository of research publications
and other research outputs

Corpus-assisted literary evaluation

Journal Article

oro.open.ac.uk

Corpus-assisted literary evaluation

Kieran O'Halloran[1]

Abstract

Fleur Adcock's poem, *Street Song*, is evaluated by the stylistician, Roger Fowler, as 'dynamic and disturbing'. I agree with his literary evaluation. These unsettling effects take place in initial response to the poem, effects which draw me into the work. In other words, they are experienced before proper reflection and analysis of the poem and individual interpretation of it. Implicit within Fowler's evaluation is that this is likely to apply for readers generally. The purpose of this article is to show how empirical corpus evidence can usefully provide substantiation of such initial evaluations of literary works, showing whether or not they are likely to be stereotypically experienced by readers. In drawing on both schema theory and corpus analysis to achieve this, the article makes links between cognitive stylistic and corpus stylistic foci.

1. Introduction

1.1 Fowler's evaluation of *Street Song*

The aim of this paper is to show how the use of a large corpus can assist in the evaluation of initial responses to a literary work. As data I will use the poem, *Street Song*, by Fleur Adcock. The corpus I will draw on is the Bank of English, a corpus of 450 million words. I first encountered *Street Song* in an analysis of it by the stylistician, Roger Fowler (1996: 201–4). The poem (Adcock, 2000: 141–42) can be seen on the following page.

Fowler (1996: 202–4) notes how the 'shifts of register' in the poem 'produce unsettling shifts of tone'. For him the poem 'feels dynamic and disturbing'. So, for example, he observes that the first verse has a 'four-beat pulse… associated with popular oral verse and, particularly, verse for children'. For him, the cheery child-like verse shows an ironic relation to the 'menace and perversion of vocabulary that follows'. This creates, as Fowler notes, double meanings for 'waiting', 'hiding' and 'games' since

---

[1] Centre for Language and Communication, Faculty of Education and Language Studies, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
 Correspondence to: Kieran O'Halloran, *e-mail*: k.a.ohalloran@open.ac.uk

the game is not only 'hide and seek' but the 'perverse ("peculiar") sexual play of the adult predator'.

Street Song

Pink Lane, Strawberry Lane, Pudding Chare:
someone is waiting, I don't know where;
hiding among the nursery names,
he wants to play peculiar games.

In Leazes Terrace or Leazes Park
someone is loitering in the dark,
feeling the giggles rise in his throat
and fingering something under his coat.

He could be sidling along Forth Lane
to stop some girl from catching her train,
or stalking the grounds of the RVI
to see if a student nurse goes by.

In Belle Grove Terrace or Fountain Row
or Hunter's Road he's raring to go –
unless he's the quiet shape you'll meet
on the cobbles in Back Stowell Street.

Monk Street, Friars Street, Gallowgate
are better avoided when it's late.
Even in Sandhill and the Side
there are shadows where a man could hide.

So don't go lightly along Darn Crook
because the Ripper's been brought to book.
Wear flat shoes, and be ready to run:
remember, sisters, there's more than one.

© Fleur Adcock, 2000

When Fowler makes these observations, he is not offering an *interpretation* of the poem, his own singular reading of it, which could well differ from what the poem means to other readers in different contexts. Instead, he is providing an *evaluation* that the poem 'feels dynamic and disturbing'. These are effects experienced in the first instance. By 'in the first instance', I mean the initial stage of literary response before a reader goes to the effort of making an individual interpretation of the literary

work, but has formulated some impressions – has noticed something about the work which attracts them into it. As with much literary evaluation, Fowler leaves it implicit that this judgement of the poem could apply to readers generally. While we can read what Fowler points to in the text to substantiate his evaluation, ultimately, the judgement relates to his own cognition – the effects of the poem on his mind. Since, when we examine such an evaluation, we can have no direct access to a literary critic's mind, there is an inevitable 'leap of faith' for the examiner of this kind of literary judgement. This leap is from the textual evidence, which is observable, to what the critic says is happening in their minds, which is not directly observable.

1.2 Aims

I agree with Fowler's evaluation that *Street Song* 'feels dynamic and disturbing'. For me, it feels this way because, in part, it is unclear whether intention to act malignly is being expressed. This initial effect is a little like walking down quiet streets at night and not knowing whether the footsteps behind are those of someone who is merely walking or someone who may also have malign intentions. What I want to do in this paper is to show how the investigation of a large corpus can have a useful role to play in:

i)   helping to substantiate whether or not such initial evaluations of a literary work are likely to be experienced by readers generally; in this case, substantiating that the 'dynamic and disturbing' effects of *Street Song* are likely to be stereotypical; and,

ii)   showing, where possible, where responses by a critic to a literary work are likely to be idiosyncratic and where others are likely to be shared by readers generally; in this case, where Fowler's responses to *Street Song* are likely to be stereotypical and where they are not.

The first aim is the main concern of this paper. Work in corpus stylistics has on the whole focussed on showing the power of large corpora in providing a systematic description of a literary work's salient features (e.g., Stubbs, 2005) or assisting in the interpretation of a literary work (e.g., Adolphs and Carter, 2002; Louw, 1997; O'Halloran, 2007; Starcke, 2006), that is, what the literary text means to an individual stylistician after some reflection. There has been some focus, too, on how corpora might provide evidence of underlying cognitive processes (e.g., Deignan, 2005; Stefanowitsch, 2006). However, the focus has not been on using corpora to support initial literary impressions. Nor has it been to evaluate how a literary text works in the first instance to establish itself in readers' minds, by making us want to ponder it or to draw us in. (In other words, *before* we

make an effort to provide an individual interpretation based on a fair degree of analysis.) To help me in this focus, I will draw on the corpus-informed concept of formulaic sequence.

1.3 Formulaic sequence and language cognition

For Sinclair (1991), much language use is in line with the 'idiom principle': the hearer or reader understands language in chunks, rather than as individual words in a grammatical sequence. Such chunks are difficult to define because they range from the long, 'You can lead a horse to water; but you can't make him drink', to the short, 'Oh no!' (Schmitt and Carter, 2004: 3). To capture this variability, Wray (2002) refers to such chunks as 'formulaic sequences': '*formulaic* carries with it some associations of "unity" and of "custom" and "habit", while *sequence* indicates that there is more than one discernible unit, of whatever kind' (Wray, 2002: 9). On this broad definition, formulaic sequences can include both collocation and phraseology – phenomena that feature in the corpus analysis later in this paper.

Underwood, Schmitt and Galpin (2004) use measurements of eye movements to assess processing of formulaic sequences. They find that terminal words of sequences are processed more quickly than the same words in non-formulaic contexts. This is taken to indicate holistic storage and processing of formulaic sequences. Moreover, Wray (2002) finds both pausing and errors to be much less frequent inside formulaic sequences than outside them. There is some contention over whether or not formulaic sequences are always stored in a holistic way (see Schmitt, Grandage and Adolphs, 2004). Nevertheless, the evidence points to the fact that language processing takes place holistically. In other words, formulaic sequence meaning has cognitive reality. This has significant ramifications for the implicit or explicit evaluations that stylisticians make about how readers will perceive a literary work, such as Fowler's when he describes *Street Song* as 'dynamic and disturbing'. This is because such judgements are based on text-focussed analytical practices (e.g., Jakobson, 1960), which do not take account of formulaic sequence meaning in relation to the processing of literary works (see Section 3.2).

Since the link between formulaic sequence meaning and cognition is important in this paper, I will need to ground corpus-derived formulaic sequence evidence, in relation to *Street Song*, in cognitive theory. This will help us to see (where it is possible) whether what is activated in Fowler's mind in reading the poem (see Section 2.2) is likely to be stereotypical or not. A cognitive theory that has been used in stylistics is schema theory (e.g., Cook, 1994; Semino, 1997). A schema (plural schemata) is a packet of knowledge that is needed for processing of language and other (e.g., visual) types of data. It is to schema theory that I now turn, linking it to the concept of formulaic sequence.

2. Schema analysis and formulaic sequence

2.1 Scripts, plans, goals and themes

Cook (1994) makes a distinction between three types of schema: language, text and world. Language schemata refer to typical knowledge of a particular language. So to understand a piece of text written in English, one requires schemata for English grammar, *etc*. But one will also need text schemata – knowledge of how a language is shaped for particular register purposes (e.g., we need a text schema of a menu to understand what the waiter places in our hands when we are in a restaurant). And we also need world schemata, knowledge of the world, to help us to make sense of a text (e.g., by selecting from a menu in a restaurant, we understand that the meal we choose will then be cooked for us). In common with the schema theoretical perspective of Schank and Abelson (1977), Cook (1994) separates world schemata into four sub-types: scripts, plans, goals and themes. A *script* refers to knowledge of a stereotypical situation or activity, e.g., we choose food from a menu in a restaurant. A *goal* relates to stereotypical purposes, e.g., we want food when we are hungry. Another type of stereotypical schema is known as a *plan* and is often activated in advance of a goal schema. A plan is something that needs to happen so that a goal can be achieved, e.g., for a cordon bleu restaurant we usually need to make a booking (plan schema) so as to eat there (goal schema). Lastly, other more abstract and evaluative schemata may be activated in a particular situation. If one hears that a friend spent 150 euros on a meal at a cordon bleu restaurant, one might evaluate this politically ('most of the world is impoverished – how can spending 150 euros on one meal be justified?'), aesthetically ('it's good for one's quality of life to experience the best') and so on. Such schemata, which are less tied to specific situations but derive from our evaluations of our experience, are known as *themes*. A theme thus carries a stronger element of subjectivity by contrast with scripts, plans and goals, which are more stereotypical.

2.2 Schemata in Fowler's response to the poem

Below are extracts from Fowler (1996: 202–4) which provide other examples of his responses to *Street Song*.

(a)   '"Loitering" is a uniaccentual word from the register of police observation. A person can only loiter with bad intent; and in this context the intent is sexual assault on children and young women…'

(verse 2, line 2)

(b) 'gratification watching [children and young women]: "fingering something under his coat" is clear enough.'

(verse 2, line 4)

(c) 'vernacular, colloquial mode…as if some local people are talking about a voyeur or rapist...'

(verse 3)

(d) '"stalking", "Hunter's Road", and "raring to go" have connotations of animals and hunting.'

(verse 3, line 3 and verse 4, line 2)

(e) '"Quiet shape", "cobbles", "Back", "Monk", "Friars", "Gallowgate" are sinister or medieval in their connotations; Chaucer's readers know all about the bad morals of the monks of olden times.'

(verse 4, lines 3 and 4; verse 5, line 1)

Fowler does not draw on schema theory or any other type of cognitive theory in his analysis.[2]   I have indicated below how these responses in Fowler's reading could be related to respective schematal types. Having Fowler's response discriminated into these schematal types will, when I look later at corpus evidence in relation to *Street Song*, allow us to see, where possible, whether these are likely to be stereotypical schematal activations or not (i.e., in line with my second aim, stated in Section 1.2):

World schemata

Script	*(b)*
Plan	*(a)*, *(b)*
Goal	*(a)*, *(b)*
Theme	*(a)*, *(e)*

Language schemata

Historical connotation	*(e)*
Hunting animals connotation	*(d)*
Intention	*(a)*
Negative connotation	*(a)*, *(e)*

Text schemata	*(a)*, *(c)*

---

[2] However, Fowler (1996: 203) moves towards a schema-theoretical account when he states that, 'the last stanza… would allow us to construct a *situation* [my emphasis] of, say, someone addressing a local support or self-defence group'. 'Situation' seems to me to be akin to a script schema.

Let me deal with the world schemata first. I place (a) and (e) into the schema category of theme because of the evaluation ('bad', 'sexual assault'; 'bad morals'). (a) and (b) imply the activation of plan and goal schemata in Fowler's mind, and in the case of (b) a script, too. Aside from the world schemata, Fowler also implies language schemata which were activated in his reading. Fowler notes this with *loiter* (a) since he associates it with an intention. This is also regarded as having negative connotations and the items detailed in (e) are regarded as having negative connotations or historical ones. The items detailed in (d) have connotations of hunting animals. There are text schemata for (a) and (c) as well. This is because in (a), for Fowler, *loitering* is from a particular linguistic register and for (c) he associates verse three with the vernacular and colloquial.[3]

2.3 Stereotypical world schemata from reading *Street Song*

Looking at the whole poem, it would be difficult to imagine that *Street Song* will not for most adult readers activate stereotypical world schemata, which include plans, goals, scripts and themes about voyeurs (e.g., from 'loitering in the dark'), male masturbation (e.g., from 'fingering something under his coat') or sexual assaulters (e.g., from 'stalking the grounds of the RVI to see if a student nurse goes by') and so on. These schemata will no doubt be starting points for readers generally, as they are for Fowler (see also Section 1.1). However, on two occasions in my own reading of *Street Song*, I did not experience similar world schemata activations to those of Fowler. Instead of the Chaucerian theme schema for 'Monks' and 'Friars', I experienced a more contemporary script/theme with regard to members of the Catholic clergy who were indicted for child molestation in the 1990s. 'Raring to go' as associated with animals and hunting did not chime with my intuitions. This is not, of course, to say Fowler's schematal activations here are wrong, but that questions might be raised about how stereotypical they are likely to be. Chaucer, for example, is usually only read these days as part of some university courses in English literature.

2.4 Using a corpus to find evidence of stereotypical language schemata

Let me return to how Section 1 was rounded off: in what ways, then, might formulaic sequences as revealed through corpus analysis relate to schemata? Consider, firstly, the following from Cook (1994: 201):

---

[3] I only analyse here what Fowler presents in his analysis. Naturally, other schemata will have been activated in his reading but for which there is no trace in what he details. For example, (a) 'loitering with (bad) intent' will also involve scripts (e.g., a lone man in a park looking shifty).

If we accept the existence of the three schemata types: world schemata, text schemata and language schemata (represented respectively by: S(W), S(T), S(L)) we can assume that all of these are present in the mind of any reader. A reader's feeling that the text structure or linguistic choices of a given discourse are normal or deviant derives from a comparison of its text structure (T) and its language (L) with the reader's pre-existing text schemata S(T) and language schemata S(L). The interaction of these interactions creates the illusion of a 'world' in the discourse (W), which can then be compared with the world schemata of the reader, yielding a judgement as to the normality or deviance of that illusory world.

A large corpus provides evidence of typical formulaic sequences. It thus provides a form of evidence for typical language usage, i.e., S(L). Typical language schemata, naturally enough, are associated with typical world schemata, i.e., S(W). Indeed, for Moon (1998: 166) typical phrases trigger 'agglomerates of cultural information'. And, for Stubbs (2001: 211), common phrases such as:

'accosted by a stranger'; 'lurking in the shadows'; 'loitering on street corners'; 'fighting one's way to the top'; 'forced to undergo a serious operation' activate stored scenarios of the things which typically happen to people. We know how the world works, and given such a phrase, we can predict other components of the stories in which they occur. These ideas are also compatible with a theory of social cognition which sees linguistic repertoires (ways of talking) as sustaining certain views of social reality.

It should be stressed that saying typical world schemata are associated with typical formulaic sequences is not the same as saying that typical world schemata can only be activated by typical formulaic sequences. An abnormal form can, of course, also trigger a stereotypical world schema. 'Combating one's route to the zenith' is intuitively less typical than 'fighting one's way to the top' but, of course, could also activate stereotypical world schemata for career ambition, job hierarchies, *etc*. It must also be stressed that a large corpus only provides evidence for typical language use (S(L)) and not typical world schemata (S(W)). So a large corpus cannot be used to substantiate whether the world schemata activated in one person's head, in reading a poem, are stereotypical or not. (So I will not be able to use the Bank of English to substantiate whether the world schemata triggered by Fowler, e.g., with regard to Chaucer, are likely to be stereotypical). A large corpus can, however, substantiate the 'reader's feeling that…linguistic choices' of a literary work are 'normal or deviant' in some way from S(L). This is an important aspect of this paper.

Having discussed the relationship between formulaic sequences and language schemata, let me now set out the theory and method for my

corpus-informed analysis of *Street Song* which I use to show why it is likely to be stereotypically 'dynamic and disturbing'.

3. Theory and method

3.1 Using a corpus to substantiate 'dynamic and disturbing' effects of a literary work

3.1.1 Deviation versus non-prototypicality

In this paper, departure from typical formulaic sequence is articulated through two notions: *deviation* and *non-prototypicality*. First, consider the title of an e.e. cummings poem, *love is more thicker than forget*. Intuitively, this would seem to be a straightforward case of deviation from typical formulaic sequence. Nevertheless, investigation of a large corpus is useful for providing good empirical grounds for such deviation, i.e., if there is little or no evidence for, say, this grammatical pattern in the corpus, then this provides empirical grounds for deviation from typical formulaic sequence. By contrast, other patterns in a literary work may be examples of non-prototypical formulaic sequences rather than instances of deviation from typical usage. For instance, evidence from a large corpus can give us grounds for deciding whether '*x* is waiting by something' or '*x* is waiting for something' is the less prototypical phraseology. (*cf.* '*x* is waiting of something' which is a deviant rather than non-prototypical form).

3.1.2 Corpus evidence for potential tensions in reading

Consider a situation where stereotypical world schemata (S(W)) that would be expected to be activated in reading a particular literary work are triggered through the following: deviation from typical formulaic sequence or non-prototypical formulaic sequence, which has been identified through comparison with a large corpus. In such a situation, S(W) would not correlate with (L). Moreover, because the identification of such deviation and non-prototypicality is corpus-informed, there would be empirical grounds for supposing that a tension, between S(W) and (L), in the reading of the literary work could occur. Let us assume there were many such tensions identified for the particular literary work, through the use of a large corpus. All the evidence could, then, empirically substantiate the evaluation that 'dynamic and disturbing' effects in the reading of a literary work are likely to be stereotypical.

3.2 Equivalence and deviation / non-prototypicality

Given the influence of Jakobson on stylistics, looking for equivalences (grammatical, semantic, *etc*.)[4] is a starting point for many stylistic analyses of literary texts, as it is for Fowler (1996: 203) in his analysis of *Street Song*:

> The Jakobsonian principle of equivalence should lead the experienced reader of poetry to link together the series of words and their meanings: 'waiting', 'hiding', 'loitering', 'feeling', 'fingering', 'sidling', 'stalking', 'raring to go'. The poem is unified by this series…

A Jakobsonian focus on equivalences is a text-inherent one only. So the series which Fowler isolates and then judges to unify the poem does not take account of:

- reader awareness (conscious or subconscious) of typical formulaic sequences (S(L)) in which the above *–ing* lexical verb forms appear;
- reader awareness (conscious or subconscious) of any non-prototypical / deviant collocations and phraseologies in the poem containing the *–ing* forms; and,
- equivalences in the poem established from repeated non-prototypical / deviant collocations / phraseologies containing the *–ing* forms.

My focus here is on the *–ing* forms of the poem, since, like Fowler, I agree that they unify (most of) the text. In Section 4, through investigation of the Bank of English corpus, I establish whether the *–ing* collocations and phraseologies in the poem are non-prototypical or deviant. On the basis of this evidence, in Section 5, I explore possible tensions between:

i)   activations of likely stereotypical world schemata (see Section 2.3) *and* non-prototypicality / deviation in the *–ing* collocations / phraseologies of the poem;

ii)   *–ing* equivalences which Fowler refers to *and* equivalences in the poem established from repetition of *–ing* collocations / phraseologies which are non-prototypical / deviant from typical usage.

---

[4] One of the most famous quotations in stylistics is the following by Jakobson: 'The poetic function projects the principle of equivalence from the axis of selection into the axis of combination. Equivalence is promoted to the constitutive device of the sequence' (Jakobson, 1996 [1960]: 17).

Taken together, while the poem 'on the page' is mostly unified by *–ing* forms, I explore to what extent it is likely to lead to a non-unified reading, which could account for its 'dynamic and disturbing' effects.

To enable comparability between corpus results for the different *–ing* forms, one needs analytical consistency, as far as it is possible to achieve it. Semantic criteria are also needed to facilitate this, as well as to institute relevance for analytical focus. I move now to the three semantic criteria which guide my analysis.

3.3 Semantic criteria for evidence focus

3.3.1 Place

Place is a key topic in *Street Song*, reflected in its title as well as the number of roads, *etc.,* mentioned. The *–ing* forms, *waiting*, *hiding*, *loitering*, *sidling*, *stalking*, *raring to go*, all relate to place in the poem and, apart from *waiting* and *stalking*, co-occur with locative-functional prepositions. My method of analysis here is to see how typically in the corpus the *–ing* forms collocate with locative-functional prepositions generally and those they co-occur with in the poem. I look immediately one place to the right of the node word (n+1) since this is the typical position for locative-functional prepositions, as indeed is the case in *hiding among*, *loitering in* and *sidling along* in the poem. In the case of *stalking*, and once again for consistency's sake, I look to see whether or not it typically collocates with a place immediately after the node word, as it does in the poem ('stalking the grounds'). 'In Belle Grove Terrace or Fountain Row or Hunter's Road' (verse four, lines one to two) precedes 'he's raring to go' (verse four, line two). Front-weighting, as with this long locative-functional prepositional phrase, is a marked phenomenon; it is much more common in literary texts than in other registers (Biber *et al*., 1999: 954). My investigation is concerned with whether 'raring to go' co-occurs with a locative-functional preposition in *typical* usage. This is why I examine collocation after the node word. Once again, to facilitate comparability, I focus on n+1 collocates.

The realisation of the present progressive is not consistent in the poem. Sometimes subject and auxiliary are present ('someone is waiting'; 'someone is loitering'; 'he's raring to go'); at other times subject and auxiliary are ellipted ('hiding among the nursery names'; 'feeling the giggles rise in his throat'; 'fingering something under his coat'; 'stalking the grounds of the RVI') or there is a modal verb ('he could be sidling along Forth Lane'). To enable comparability of collocational information at n+1, the corpus search form needs, as far as is possible, to be consistent. Since all the *–ing* present progressives are in the third person singular, one candidate that recommends itself for investigation is the 'is+*–ing*' form. However, there is a problem here since there is not enough data in the Bank

of English for 'is loitering' (three instances). Much more data can be provided using lemma forms, of course, but then again such a focus would be moving away from one which is lexico-grammatically sensitive. In view of this, I decided the best solution is a search for –*ing* forms without the auxiliary. Where it would be appropriate to look at lemma forms so as to provide further substantiation, I do so.

3.3.2 Intention to act in a place

Another semantic criterion which guides analysis is 'intention to act in a place' (of a male sexual assaulter, voyeur, *etc*.). Since I intuit, in part, that the dynamic and disturbing power of the poem comes from it being unclear whether (male) intention to act in a place is being expressed or not, I also examine the corpus data, where appropriate, to see to what extent intention is typically expressed around these –*ing* forms. Such a semantic focus may entail going well beyond investigating n+1 collocates since much more co-text may be needed for this type of examination. For this focus, I use the co-textual facility of the Bank of English. (The Bank of English allows concordance lines to be expanded to five lines of co-text.) Since, as Jones and Sinclair (1974) argue, significant collocates are usually found within spans of only four words, *t*-score calculation (see Section 3.4) for collocation, then, may not be appropriate for this semantic focus.

3.3.3 Action around a male body

Rather than intention to act, or otherwise, in a place, the –*ing* forms of the second couplet of verse two, *feeling* (in 'feeling the giggles rise in his throat') and *fingering* ('fingering something under his coat') relate to actions around a male body. So I treat these two lines with a different semantic focus. I intuit *giggles* as being more associated with girls than with boys. Guided by this intuition, I look to see to what extent this is the case in the corpus. Since I had this intuition for *giggles*, and these two lines form a semantically-related couplet, I also wondered whether in some way *fingering* in typical usage might be in tension with maleness, in this case with a stereotypical script (i.e., a world schema) for male masturbation. So, this is something I also investigate.

As the reader can see, these three criteria are different, which in turn leads to differences in the way in which formulaic sequence evidence is investigated for each of them (see Section 4).

3.4 Collocation and *t*-score

Comparing raw frequencies is useful, initially, in identifying which collocates are recurrent. However, it is difficult with raw frequencies to attach a precise level of attraction between a collocate and a node word. The statistical measure, *t*-score, provides this information.[5] More precisely, it measures the 'certainty of collocation' (Hunston, 2002: 73) because it takes into account the size of the corpus used. The Bank of English software automatically generates *t*-scores in collocate searches. A *t*-score of more than two is 'normally taken to be significant' (Hunston, 2002: 72) but a *t*-score in double figures is very significant (Hunston, 2001: 16).

      Unless otherwise stated, the whole of the Bank of English is used in the analysis. Where individual subcorpora of the Bank of English are used, this will be indicated and justified.

4. Corpus analysis of –*ing* forms in *Street Song*

4.1 Place and intention to act

4.1.1 (someone is) waiting (, I don't know where) (verse one)

There are 49,852 instances of *waiting*. The first prepositions which occur at n+1, and which can be locative-functional, are *in* (1,834 instances; *t*-score 21.7) and then *at* (804 instances; 19.1 *t*-score). However, a locative-functional preposition is not the most common preposition. *For* and *to* (both to indicate purpose) are by far the most typical prepositions. *Waiting* collocates at n+1 with *for* 19,149 times and with *to* 7,748 times. The *t*-scores are well over ten for both *for* (135.2) and *to* (73.8) and are thus extremely significant. The corpus evidence tells us that, habitually, *waiting* occurs in the phraseologies 'is waiting for+something / someone' and 'is waiting to do something', and much more so than in phraseologies with a locative-functional preposition. Indeed, corpus evidence (through expansion of co-text) overwhelmingly communicates that intention is indicated around *waiting* where there is a human subject.

---

[5] *T*-score depends on a number of calculations. The first is the number of instances of the co-occuring word in the specified span. This value is known as 'the Observed'. The second calculation is based on the null hypothesis: the co-occurring word has no effect at all on its lexical environment. In other words, its relative frequency of co-occurrence with the node word in the specified span is the same as its relative frequency in the entire corpus being investigated. This value is known as 'the Expected'. The final calculation that *t*-score depends on is 'standard deviation'. This calculation involves the probability of co-occurrence of the node and the collocate and the number of words in the specified span in all concordance lines. *T*-score is calculated by subtracting 'the Expected' from 'the Observed' and dividing this number by the standard deviation value.

From this corpus evidence, we can infer that in, 'someone is waiting, I don't know where', there is *phraseological deviation* because nothing in the phraseology indicates intention being expressed. This is a form of deviation different to that which Mukařovský (1932) highlighted, since deviation from phraseological norms is not immediately obvious, as compared with other forms of deviation such as grammatical deviation, e.g., 'someone are waiting'. But let us say I could have intuited that 'someone is waiting, I don't know where' includes a phraseological deviation. With the statistical evidence gained through empirical exploration, I am in a stronger position: I am not just supposing that this is a phraseological deviation; I have shown that it is the case using data from the Bank of English corpus. Lastly, I have chosen 'deviation' as a description here rather than 'non-prototypicality' since expression of intention around 'waiting' with a human subject is the overwhelming norm.

4.1.2 hiding among (the nursery names) (verse one)

There are 9,461 instances of *hiding*. Fifty-eight instances of *hiding* have *among* as an n+1 collocate (*t*-score 5.0). For other prepositions which can be locative-functional, and which are n+1 collocates, the frequencies and *t*-scores are much higher: *in* (3,575 instances; *t*-score 36.8), *behind* (702 instances; *t*-score 25.8), *under* (283 instances; *t*-score 14.2). So *among* as an n+1 collocate of *hiding* is less common than other prepositions of place, and much less so in comparison with *in*. 'Hiding among' in verse one is thus an example of *non-prototypical collocation*. And when people are hiding among others, on the evidence in the Bank of English, there is a mini-pattern: the intention to act on people is not indicated. 'Hiding among' often means lying low, being dormant for a while, e.g:

```
…Britain is likely in future to be involved in
conflicts where there has been no declaration of
war and where the job of British forces will be to
locate and destroy an enemy hiding among a civilian
population.
```

4.1.3 (someone is) loitering (in the dark) (verse two)

There are 361 instances of *loitering*. There are seventy-two instances of *in* as an n+1 collocate and with a *t*-score of 7.6. In fact, *in* has the highest *t*-score of all n+1 collocates. All instances of *in* relate to place so 'loitering+in+the dark' in the poem is not deviant from a typical formulaic sequence (structurally or semantically), nor a non-prototypical formulaic sequence.

In Section 1.2, I indicated a second aim of the paper is to test, where possible, whether Fowler's responses to the poem are likely to be stereotypical schematal activations or not. Recall that *loitering* activated in Fowler (1996) a text schema, i.e., the crime of 'loitering with intent'. For Fowler (1996: 203), *loitering* involves intention to act and 'a person can only loiter with bad intent', i.e., Fowler's schemata also include language schemata for negative intention. There is evidence for this take on *loitering* in the Bank of English, e.g.:

```
Mrs de Rosnay told the Old Bailey that she had seen
a well-built man in a dark suit loitering near Miss
Dando's house two hours before the killing.
```

There are also seventeen instances of the expression 'loitering with intent' (*with* and *intent* have *t*-scores of 3.9) though all of these are ironic and not related to criminal activity. The following, for example, is from a football report:

```
Instead, Gordon Marshall claimed the ball and
launched his kicks at Hearts' left-back area, where
Craig Dargo and Kris Boyd were loitering with
intent.
```

But the corpus search also tells us that usage of *loitering* is more complicated. People may loiter with an intention to act, but an action that does not have (criminal) 'bad intent':

```
When he arrived he had to make his way through a
loitering group of journalists. They regarded him
with brief interest, until they concluded he was
neither a doctor not a policeman, and they ignored
him.
```

Alternatively, people can loiter in the sense of just 'hanging about' with no clear and specific intention to act. Here is an example:

```
Should you be loitering around Hyde Park Corner
over the next three weeks, pop into Pizza on the
Park for a comical crash course in the lost art of
cabaret.
```

Indeed, around 40 percent of instances of *loitering* in the Bank of English occur without obvious intentions being indicated or being readily inferable. Thus, the corpus evidence usefully tells us that, stereotypically, *loitering* is sometimes associated with intention to act (sometimes negatively) and sometimes not associated with an intention to act.

In analysing corpus evidence, one must be careful to distinguish quantitative frequency evidence from qualitative evidence about the

salience of a phenomenon in a culture. The crime 'loitering with intent' may be salient across a culture without it necessarily being talked or written about very much. Just because a phenomenon is not reflected by frequency of instances in a corpus does not mean it is not salient. The corpus evidence does not tell us that Fowler's generation of a language schema associated with bad intention for *loitering* is wrong. But it does tell us that *loitering* is not always associated with (bad) intention and thus, *pace* Fowler's (1996: 203) text schema, *loitering* is not then 'uniaccentual' (see Section 2.2). The corpus evidence tells us that *loitering* does carry a negative semantic prosody[6] (Louw, 1993), unsurprisingly, when used in relation to crime.

4.1.4 (He could be) sidling along (Forth Lane) (verse three)

There are eighty-nine instances of *sidling* in the whole corpus. *Up* is the most common n+1 collocate featuring forty-two times with a *t*-score of 6.4. By contrast, *along* only features six times with a *t*-score of 2.4. However, given there were only eighty-nine instances of *sidling*, I went on to explore the lemma SIDLE (434 instances) to seek possible corroboration of this pattern. There are 211 instances (*t*-score, 14.5) of 'SIDLE *up*+preposition+ someone' but only twelve instances (*t*-score, 3.44) of 'SIDLE *along* somewhere'. The discrepancy between these respective *t*-scores is significant; secondly, only 5 percent of usage is for the lemma SIDLE+along. The corpus evidence shows that 'sidling' / 'SIDLE' collocate much more with *up* or *over* and with people (e.g., 'sidling up behind her') than with a place. Thus, the corpus evidence reveals that 'sidling along' exhibits non-prototypical collocation.

      Let me shift focus to the second aim of the paper as set out in Section 1.2. For Fowler (1996: 203), verse three evokes the, 'vernacular, colloquial mode … as if some local people are talking about a voyeur', and so Fowler has a text schema activated here. To investigate whether this could be substantiated, I went on to look at the spoken corpora of the Bank of English (sixty-eight million words) only. *Sidling* occurs once and the lemma, SIDLE, seven times. On the evidence of the spoken corpora, it

---

[6] The concept of 'semantic prosody' has had wide currency in corpus-based linguistics (e.g., Sinclair, 1991; Louw, 1993; Hunston, 1995; Stubbs, 1996; Channell, 2000; Sinclair, 2004). Here is a recent definition from Sinclair (2003: 178):

> A corpus enables us to see words grouping together to make special meanings that relate not so much to their dictionary meanings as to the reasons why they were chosen together. This kind of meaning is called a semantic prosody; it has been recognised in part as connotation, pragmatic meaning and attitudinal meaning.

Sinclair (2004: 30–35) gives the example of the seemingly 'neutral' phrase, 'the naked eye'. Corpus investigation reveals a common phraseology, 'visibility + preposition + the + naked + eye', which in turn reveals a negative semantic prosody such as in 'too faint to be seen with the naked eye' or 'it is not really visible to the naked eye'.

would not seem to be associated with colloquial usage, which is thus in tension with Fowler's text schema activation.

4.1.5 (or) stalking (the grounds of the RVI) (verse three)

Analysis is a little different with this *–ing* form since there is no preposition at n+1. Also, consistent with the above line, where place features at n+2, I explore collocates at n+2 as well as n+1.

There are 1,788 instances of *stalking*. The top *t*-scores for stalking of place are *streets* (twenty instances, *t*-score, 4.4), and *corridors* (thirteen instances, *t*-score 3.6). (These are n+2 collocates preceded by *the*.) These may be significant *t*-scores but, in line with my semantic focus on place, it is important to semantically group the n+1 / n+2 collocates so as to see the extent to which *place* collocates with *stalking*. Taken together 'stalking+place' only makes up around 10 percent of collocation in these positions. By contrast, around 80 percent of instances of *stalking* collocate, in these positions, with people as AFFECTED semantic roles, for example, 'a psychopathic serial killer stalking a *woman*'. Indeed, there are sixty-three instances of the n+1 collocate *her*, which has a *t*-score of 7.4. 'Stalking+place', as we have in the poem, is thus another non-prototypical collocation.

4.1.6 (or Hunter's Road he's) raring to go (– unless) (verse four)

There are 520 instances of *raring* and 445 instances of *raring to go* with a *t*-score of 21.1. There seemed, then, to be evidence that *raring to go* is a formulaic sequence. Hence, I went on to investigate what its common collocates are at n+1, rather than just looking at *raring*. *Unless* immediately follows *raring to go-* in verse four of the poem. However, there are no instances of *unless* as an n+1 collocate of *raring to go* in the corpus. Its most common n+1 collocate is <p> (seventy-eight instances; 8.3 *t*-score). This is mark-up in the corpus which indicates a new paragraph. The only n+1 preposition is *at* with eight instances and a low *t*-score of 1.9; only one of these instances refers to a place ('at St James Park'). Other prepositions which can be locative-functional have non-significant *t*-scores, e.g., *on* (six instances, 1.2 *t*-score). The evidence suggests, then, that it is much more common for *raring to go* to finish a sentence rather than to co-occur with an n+1 locative-functional preposition. To sum up: (i) we can say that the collocation of *raring to go* with *unless* in verse four of the poem is deviant from a language schematic perspective; (ii) the co-occurrence of *raring to go* with a locative-functional prepositional phrase in the poem ('In Belle Grove Terrace…') is non-prototypical from a language schematic perspective; and, (iii) since *raring*

to go does not end a sentence in the poem, its use in *Street Song* is non-prototypical from a text schematic perspective.

Interestingly, *raring to go* commonly occurs in the sports report register, and particularly football news. <p>, the most common n+1 collocate of *raring to go*, is commonly followed by commentary by a sportsperson on their readiness for sport activity, e.g.:

```
But despite his trials and tribulations the winger
insists he's fit and raring to go.
<p> He said: 'The pre-season training has been
excellent so far.  From my own point of view things
have been going great and I feel fit and mentally
ready as well.
```

Fowler experiences language schemata around 'hunting and animals' for line two of verse four (see Section 2.2). On the basis of corpus evidence, we can at least go beyond Fowler's personal commentary in saying that many people are likely to have a text schema for sports reporting in relation to *raring to go*.

4.2 –*ing* forms relating to male body

4.2.1 feeling the giggles (rise in his throat) (verse two)

In Section 4.1, I mainly explored n+1 collocates of –*ing* forms because of the focus on prepositions which are typically locative-functional in this position. In contrast, in Section 4.2.1, I look for evidence as to whether *giggles* is typically associated with females or males. My focus is less positionally-specific in relation to collocation and so this is why I expand my collocate focus and use both an n–4 and an n+4 span. I choose up to four places in line with Jones and Sinclair's (1974) judgement that significant collocates are usually found within spans of four.

There are 979 instances of *giggles* in the corpus. The most common collocates for *giggles* can be seen in Figure 1. Notice the number of times females are referred to, e.g., *Anna, Gerti, she, her, girls, girl* and *girlish*. In contrast, males are only referred to three times through pronouns; there are no instances of the lemma, BOY, though there are two male names, *David* and *Gary*. While *she* at n–1 has a significant *t*-score of 9.2 and collocates ninety times, *he* at n–1 has a lower *t*-score of 4.7 and collocates thirty-four times. Overall, the evidence seems to indicate that *giggles* is more likely to be language-schematically associated with girls (perhaps females generally) rather than boys (perhaps males generally). This evidence thus contrasts with the information relating to the male gender in the line, 'feeling the giggles rise in *his* throat.'

n–4	n–3	n–2	n–1	NODE	n+1	n+2	n+3	n+4
a	fit	fit	**she**	NODE	and	<u>he</u>	**her**	i
into	a	fits	of	NODE	<p>	the	says	**her**
you	into	of	the	NODE	at	i	**she**	**she**
i	**her**	got	and	NODE	when	**her**	audience	so
then	fits	me	into	NODE	as	<u>he</u>	memory	hand
her	gave	into	<u>he</u>	NODE	but	s	**she**	my
fun	both	get	few	NODE	from	lot	me	my
she	gary	getting	nervous	NODE	i	then	i	little
leary	**girl**	stifle	**her**	NODE	about	<u>his</u>	then	got
conversati	attack	collapsed	**girlish**	NODE	again	they	s	knows
sweet	**she**	collapse	uncontroll	NODE	over	it	**girls**	both
had	still	few	with	NODE	it	ve	were	at
we	to	high	stifled	NODE	**she**	some	looks	t
like	my	then	helpless	NODE	what	from	**girl**	but
says	burst	back	pitched	NODE	during	says	added	then
david	were	gerti	gerti	NODE	nervously	when	began	bistro
face	taste	<p>	between	NODE	then	drools	as	backstage
away	reduced	gerti	baby	NODE	or	cackles	doesn	toy
in	your	smirks	their	NODE	friendship	gurgles	you	jokes
were	bit	shrieks	more	NODE	anna	wiggles	down	drives

Figure 1: Collocation grid for *giggles*, for collocates within n–4 and n+4, with the twenty highest *t*-scores in descending order
(Note: Word-forms of the lemma *GIRL*, *her* and *she* are shown in boldface. Instances of masculine pronouns are underlined)

4.2.2 (and) fingering (something under his coat) (verse two)

There are 400 instances of *fingering* in the corpus. These are all verb forms. The stereotypical world schema likely to be activated from 'fingering something under his coat' is male masturbation. The poem uses *something* as the object of *fingering*. I am interested in seeing what kinds of phenomena are fingered, that is, what would typically fill the 'something' slot. So I look only to the right of the node word (see Figure 2). For comparability with Section 4.2.1, I use a span of n+4.

```
NODE      the        beads      of          his
NODE      his        gold       as          the
NODE      a          wound      and         her
NODE      her        little     beads       goatee
NODE      my         way        collar      paper
NODE      their      keys       gold        if
NODE      its        diamond    sprouted    of
NODE      through    wire       mustache    she
NODE      herself    tie        violin      cashmere
NODE      koba       small      buttons     keys
NODE      rolls      lacy       fence       bow
NODE      prayer     chords     cap         button
NODE      technique  folds      strike      reward
NODE      patterns   moustache  dark        actual
NODE      some       strand     trying      tie
NODE      them       dresses    on          ring
NODE      hand       sword      again       that
NODE      it         trigger    same        hair
NODE      one        stem       little      he
NODE      amorously  collar     man         didn
```

Figure 2: Collocation grid for *fingering* for collocates within n+4.
Collocates for the twenty highest *t*-scores in descending order

The kinds of thing which are fingered in the corpus are, for example, *beads*, *cashmere*, *diamond*, *goatee* and *moustache*. The only things which are fingered which have a *t*-score higher than two are *beads* (five instances; *t*-score 2.3) and *gold* (five instances; *t*-score 2.2). So there is no collocate which has a pronounced relationship with *fingering*. Having said that, by viewing the co-textual information, we can see that many instances have something in common: many involve light touching. I will address the implications of this in Section 5.1.

4.3 (to stop some girl from) catching (her train) (verse three)

Catching, lines one and two of verse three, 'He could be sidling along Forth lane / to stop some girl from catching her train', relates to the

behaviour of a *female* human being in relation to a train. This *–ing* form does not fit with my semantic foci (see Section 3.3). In the interests of consistency, I thus ignore this *–ing* form (as indeed, interestingly, does Fowler).

I have highlighted departures from collocational and phraseological norms in a 450 million word corpus with regard to *–ing* forms in *Street Song*. On the basis of the corpus evidence, in the next section I will indicate the various potential equivalences, or tensions between equivalences, that *–ing* forms contract into, which in turn can count as support for evaluation of the poem as one likely to be stereotypically 'dynamic and disturbing'.

5. Using corpus evidence in literary evaluation

5.1 Tensions between phraseology and stereotypical world schemata

5.1.1 Stereotypical plan / goal schemata

Stereotypically, people will activate plan and goal world schemata for (child) sex offenders in their reading of verse one (i.e., schemata with specific intentions). So, the phraseological deviation in 'someone is waiting, I don't know where' (verse one, line two), where intention is not indicated, is in tension with stereotypical world schemata. It could be argued, of course, that 'someone is waiting' is elliptical where, for example, 'to do something or other' or 'for something' is implicit. But if this is the case, I would expect on the basis of corpus evidence as a follow-up 'I don't know why' or 'I don't know what for' rather than 'I don't know *where*'. Indeed, the 'where' in the poem would seem to have been already communicated in the first line of verse one with the locations of 'Pink Lane, Strawberry Lane, Pudding Chare'. A related tension emerges in verse two. From the corpus evidence, 'someone is loitering' in verse two may or may not be associated with intention. But, one would expect a reader to have stereotypical plan and goal schemata for a sex offender from verse two, that is, activation of specific intentions.

5.1.2 Stereotypical script schemata

Fingering (verse two, line four) has the semantic prosody of 'light touching'. So, the corpus evidence is in tension with the stereotypical world schema of male masturbation, which includes the script that it is a vigorous activity.

5.2 Equivalences

5.2.1 Non-prototypical collocation

We know from corpus evidence (Sections 4.1.4 and 4.1.5) that: (i) 'sidling along (something)' is much less common than 'sidling up (to someone)'; and, (ii) 'stalking a place' is much less common than 'stalking a person'. Thus, in verse three, 'sidling along Forth Lane' and 'stalking the grounds of the RVI' are not only equivalent grammatically but equivalent because they are both instances of non-prototypical collocation in relating to places rather than human beings. A potentially 'disturbing' tension is created, then, between actions in a place ('Forth Lane' and 'the grounds of the RVI') and typical language schemata where *stalking* (especially in relation to voyeurism) and *sidling* associate with human beings. Lastly, it should be noted that 'hiding among' (verse one), as another instance of non-prototypical collocation, is then also equivalent to 'sidling along' and 'stalking the grounds of the RVI' in verse three. This latter equivalence is not so visible on a text-focussed stylistic analysis only.

5.2.2 Gender

In verse two, *feeling* (line three) and *fingering* (line four) are equivalent morphologically. What a corpus-informed perspective illuminates is that 'feeling the giggles rise in his throat' and 'fingering something under his coat' are equivalent in another sense: because of stereotypical language schemata around gender which are *not* met. In turn, this equivalence also contributes to the 'dynamic and disturbing' effects of the poem.

5.2.3 Phraseological fragment

'Raring to go' has a high *t*-score for *and* (251 instances; *t*-score 9.6) at n–1. The evidence tells us that the formulaic sequence 'raring to go' can regularly be longer before the node words, e.g., 'fit and raring to go', 'relaxed and raring to go', 'refreshed and raring to go'. This can be seen in Figure 3. There is a relatively high *t*-score (7.43) for *fit* (fifty-six instances). The semantic preference[7] before the node words is commonly:

---

[7] The concept of 'semantic preference' is another common one in corpus-based linguistics. It refers to a set of different, frequently-occurring collocates which are from the same semantic field. Here is more information on semantic preference from Sinclair (2003: 178):

> Sometimes in the structure of a phrase there is a clear preference for words of a particular meaning. The word class is not important, and any word with the appropriate meaning will do (though there are often collocational patterns within semantic preference). While the majority of the choices will show the preference clearly, there may be a small number of marginal cases where the preferred meaning has to be interpreted in a rather elastic fashion, and some which appear to be exceptions. For this reason, we do not use a word like "restriction" instead of "preference".

recovering from injury / illness; and refreshing oneself before engaging with the particular sport, usually football (see extract in Section 4.1.6). In other words, 'raring to go' is commonly bound up with this particular text schema.

On the basis of the Bank of English corpus, 'raring to go' (verse four, line two) could be treated as a phraseological fragment. We also know that 'someone is waiting' (verse one, line two) is a phraseological fragment since corpus evidence indicates the norm that intention is expressed with, for example, an ensuing *for* or *to*. So we have equivalence between these lines in verses one and four. This is a different form of equivalence from what we have observed so far: this might be called *phraseological fragment equivalence*. This equivalence is reinforced by the structure of the poem; each of the above lines referred to is a second one in the verses. Again, this is an equivalence which is not so visible on a text-focussed stylistic analysis only.

```
others      and        i          and         NODE
he          indian     fit        is          NODE
and         now        he         be          NODE
<p>         he         will       m           NODE
i           is         ll         was         NODE
fit         i          it         are         NODE
</dt>       said       ready      firms       NODE
but         fully      they       s           NODE
friday      fit        should     were        NODE
now         m          back       am          NODE
sun         be         up         still       NODE
cup         <hl>       refreshed  re          NODE
feel        punters    relaxed    just        NODE
week        they       fresh      him         NODE
being       are        d          team        NODE
<hl>        feel       were       been        NODE
m           season     everybody  absolutely  NODE
back        but        everyone   now         NODE
raring      revved     strong     clearly     NODE
batteries   pumped     club       mendonca    NODE
```

Figure 3: Collocation grid for 'raring to go' for collocates within n–4 (in line with Jones and Sinclair, 1974). Collocates are for the twenty highest *t*-scores in descending order
(Note: Instances of *fit* are emboldened, as is 'and: n–1')

5.3 Non-equivalences

5.3.1 Phraseological

There is quite a complicated relationship between 'someone is waiting' (verse one, line two) and 'someone is loitering' (verse two, line two). While they are grammatically and linearly equivalent in the poem, they are not phraseologically equivalent: 'someone is waiting, I don't know where' is phraseologically deviant whereas 'someone is loitering in the dark' is not.

There are infinitives in verse three expressing intentions to act: 'sidling…*to stop* some girl…', 'stalking…*to see* if a student nurse…' There is a tension between the presence of these infinitive-of-purpose phraseologies in verse three and the lack of (infinitive of) purpose following 'someone is waiting' in verse one, which from the corpus evidence we know regularly follows 'waiting' in English usage. So, on the basis of this phraseological evidence, *waiting* and *sidling / stalking* can be seen as non-equivalent in the poem.

The lines 'Pink Lane, Strawberry Lane, Pudding Chare' (verse one, line one) and 'Monk Street, Friars Street, Gallowgate' (verse five, line one) are semantically equivalent in the poem since they list road names and are linearly equivalent, too, being first lines of verses. Verse five, line one is a subject of a verb and as such 'Monk Street, Friars Street, Gallowgate are better avoided when it's late' is grammatically prototypical. However, from the corpus evidence, we know that the absence of a locative-functional preposition, such as *in*, from verse one line one is non-prototypical in relation to *waiting*. Phraseologically, the lines are thus non-equivalent.

A last phraseological non-equivalence is between verse two, lines one to two, 'In Leazes Terrace or Leazes Park / someone is loitering in the dark' and verse four, lines one to two, 'In Belle Grove Terrace or Fountain Row/ or Hunter's Road he's raring to go –'. These lines have linear and grammatical equivalence with the front-weighting of the locative-functional prepositional phrases in lines one and the positions of the –*ing* forms in lines two. However, we know from corpus evidence that while *loitering* collocates typically with locative-functional prepositions such as *in*, this is not the case for 'raring to go'.

5.3.2 Intention to act

From corpus evidence, 'HUMAN SUBJECT+(is)+waiting' (verse one, line two) is overwhelmingly associated with intention, but this is not expressed in verse one through, for example, use of the prepositions *to* or *for*. In verse one, line three, intention to act is not so apparent in 'hiding among the nursery names'; the corpus evidence for 'hiding among' provides some

corroboration of this. However, in verse one, line four, 'he wants to play peculiar games', intention to act is expressed.   There is, then, non-equivalence in the first verse in the expressing of intention, which corpus evidence helps to reveal / substantiate. Furthermore, whereas from corpus evidence 'HUMAN SUBJECT+(is)+waiting' is overwhelmingly associated with intention, corpus evidence also shows that 'HUMAN SUBJECT+(is)+loitering' (verse two, line two) may or may not be associated with intention. Despite the fact that 'someone is waiting' and 'someone is loitering' are grammatically equivalent, with regard to expression of intention they are not equivalent.

5.4 Higher resolution of dynamic and disturbing effects

In verse four, implicitly there are two kinds of male behaviour associated with the preparation for sexual assault: one (type of) man who is 'raring to go' and one who is a 'quiet shape…on the cobbles'. In Sections 4.1.6 and 5.2.3, I found that 'raring to go' has a semantic preference for an athlete's recovery from injury, an arduous sport season, *etc.*, in the use of the phrase 'fit / refreshed, *etc.*, and raring to go'. So this 'raring to go' man can be given a 'higher resolution' by the corpus evidence as one who is young and athletic.

Let me now move on to the 'quiet shape…on the cobbles' man. In the Bank of English, there are only sixty-one instances of 'on the cobbles'. However, significantly, around two thirds of these instances appear in book corpora (105 million words) which consist of historical fiction. Figure 4 shows concordance lines for these thirty-nine instances. In the book corpora, 'on the cobbles' has, to a reasonable degree, a semantic preference: *cart*, *cloven*, *coach*, *hay*, *hoof, mule*, etc., belong to the same semantic field. Here is one example of 'on the cobbles' with *horses* in 'historical fiction':

```
…but her horse broke into a trot at the sight of
the castle gate, and the moment was lost.  As they
entered the courtyard, the scrape of their horses'
hooves  echoing  hollowly  on  the  cobbles,  Teidez
burst  from  a  side  door,  crying  'Iselle!  Iselle!'.
Cazaril's  hand  leapt  to  his  sword  hilt  in  shock –
the  boy's  tunic  and  trousers  were  bespattered  with
blood – then fell away again…
```

The activation in Fowler's (1996: 203) mind of 'medieval' (see Section 2.2), and thus an historical association, is coincident with much of the collocation of 'on the cobbles' (though Fowler only cites 'cobbles'). So, the corpus evidence provides some substantiation of Fowler's language schema. Another point to make is that 'on the cobbles' also has some semantic preference for expressions of sound. In other words, 'sound'

could well be included in a stereotypical language schema for 'on the cobbles'. This is interesting because the man 'on the cobbles' referred to in verse four is a '*quiet* shape'.

```
at the top was Anna Abraham. On the cobbles beside her were two baskets. frost
          covered the town and on the cobbles of the Town Quay were scattered
     the clatter of hoof-beats on the cobbles and, going to the doorway, saw a
      his hair. He heard the boots on the cobbles behind him, put his head down,
         the sound of their boots on the cobbles, their eyes glinting behind the
hooves, the trundle of carts on the cobbles, the shouts of ostlers and the
Sulieman's hooves clattering on the cobbles. Two grooms came running but
      the burning building. Down on the cobbles, her face lit by the garish glow,
    joined the rattling of feet on the cobbles. A woman's voice started up, sang
      another. The sun gleamed on the cobbles as though they were metallic, the
         hooves echoing hollowly on the cobbles, Teidez burst from a side door,
on him, the sound of hoofs on the cobbles almost louder than the rain. He
clang of iron-shod hooves on the cobbles of the forecourt covered Sebastian
        of black feathers lay on the cobbles already, one still, one twitching.
Cloven feet clacking loudly on the cobbles, he cantered to the steps, then,
     <p> Kids playing marbles on the cobbles. There were no birds, or animals.
clacking of their boot-nails on the cobbles by the yard's arched entrance. The
horses trampled nervously on the cobbles of the courtyard, then looked up
    the mangled torso laid out on the cobbles. The Zangre's castle warder,
    of a mule and lay it out on the cobbles right up to the end of Parliament
      White lines were painted on the cobbles while he cleaned his tack.
     bound shock of hay, Rhodry on the cobbles. The boy waited until he had
        his heavy boots ringing on the cobbles. Her reading of the situation was
      to the yard, heels ringing on the cobbles in indecision, and then the door
    door bang, shoes scraping on the cobbles; the largest of the sailors, the
     people abroad. Feet shifted on the cobbles behind them, and he swung round
      was the scrape of a shoe on the cobbles of the stable yard, staring at her
<p> They found Trish sitting on the cobbles and were woken by the rain. They
   the gallows, were sleeping on the cobbles and he fell full length. 'I
   but his club foot slipped on the cobbles. 'Well done,' Lavisser said
  as he jerked spasmodically on the cobbles far below. Bag jerking in his
   suicide, wasted, spattered on the cobbles as the rope's cut end flicked
legs of the chair splintered on the cobbles beneath, but she was quite simply
   a small gallery breaking up on the cobbles. Devil and Honoria had already
     both already waiting on the cobbles, though the buckets were gone.
coach still glistened wetly on the cobbles and the shouting and cursing of
hooves or carriage wheels on the cobbles. She recognized Phyllida and came
at the sound of their wheels on the cobbles.
```

Figure 4: Concordance lines for 'on the cobbles' from the book corpora of the Bank of English
(Note: Highlighted in bold are things which belong to the semantic field of equine animals; expressions which indicate sound are underlined)

It is clear that, without the corpus evidence, there are two types of behaviour being referred to in verse four. However, the corpus evidence throws the two types of behaviour into starker contrast since the types of

men are given a higher resolution: young, sporty athletic man in the first two lines versus a man who has some 'historical' associations in the last two lines – two very different men. Use of corpus evidence, then, helps to substantiate the 'dynamic' element of the poem which Fowler mentions, since it firms up the quick shift from one very different type of man to another.   Lastly, corpus exploration provides more evidence for the 'disturbing' tensions set up in reading the poem, i.e., between the '*quiet shape*' man and sonorous associations of 'on the cobbles'.

5.5 Textual unity versus disunity in reading

Recall from Section 3.2 that Fowler (1996: 203) argues that the poem is 'unified' by the *–ing* series. On a Jakobsonian text-focussed stylistic analysis only, there is 'unity' in Fowler's sense: the *–ing* forms are morphologically and grammatically equivalent. On the basis of the corpus-informed analysis, however, there is likely to be disunity in reading the *–ing* verb forms of *Street Song*. This is primarily due to:

- tensions between stereotypical world schemata that are likely to be activated in reading and non-prototypical / deviant formulaic sequences in which the *–ing* forms occur in the poem; and,
- the existence of different, and thus non-unified, patterns of equivalence and non-equivalence for the *–ing* forms in the poem in relation to non-prototypical / deviant formulaic sequences.

Since regular collocations and phraseologies are bound up with stereotypical language schemata, the tensions and non-unified patterns identified provide some corroboration of how the poem is likely to be stereotypically 'dynamic and disturbing' for readers in the first instance.

6. Conclusion

While the patterns of corpus-informed equivalence or non-equivalence I have highlighted are not so readily detectable on the text-focussed approach of Jakobsonian stylistics, this is not to say I think that stylistics should abandon a Jakobsonian approach. Rather, I would argue that large corpus analysis should complement a Jakobsonian text-focussed approach for the following: if the analyst is seeking corroboration for their evaluation of a literary work's capacity to establish itself, in the first instance, in readers' minds generally.   With this evaluative focus, then, I have not been conjecturing the kinds of interpretations that might be made of *Street Song*. Clearly, it is impossible to forecast what individual interpretations different readers in different contexts will make in their sustained reflections on this poem.

It must be borne in mind that individual readers, like Fowler, will approach a literary text with a mix of stereotypical schemata and not so stereotypical schemata such as 'Chaucer' in Fowler's reading of verse five, line one, of *Street Song*, 'Monk Street, Friars Street, Gallowgate'. This does not affect the method of this paper since I have provided an explanation of the poem *only* where readers have stereotypical activations of (world, language and text) schemata in their reading of it.

With my focus on evaluating the capacity of a poem to establish itself in the minds of readers in the first instance, I have in essence been discussing a particular aspect of literary creativity. In literary studies in the 1970s and 1980s, investigations of the nature of creativity largely lay dormant while oppositional critique (deconstructionist, feminist, neo-Marxist, post-colonial) became ascendant. This was understandable in times of social change when it became readily apparent that what was regarded as canonically creative was to some extent a reflection of mainstream socio-political forces as well as literary merit. However, since oppositional critique has now become fairly established, this is allowing a revisiting of issues of creativity (e.g., Attridge, 2004; Carter, 2004; Cook, 2000; Goodman and O'Halloran, 2006; Maybin and Swann, 2006; Pope, 2005). On the basis of this paper, I would argue that corpus investigation is valuable for:

- showing 'jarring' between stereotypical world schemata activated and the 'surface' form of a literary work; and,
- investigating the relationships between equivalences of forms in a literary work and the formulaic sequences S(L) in which those forms habitually occur.

Both can give insight into how the creativity in a literary text works to establish the text in readers' minds in the first instance by creating 'dynamic and disturbing' effects. I would also argue that the above method is valuable for other reasons:

- to help reduce individual reader-relative, speculative analysis of schemata through using a corpus to investigate the kinds of stereotypical language and text schemata likely to be activated more generally by a literary work; and,
- pedagogically: using the corpus as a tool, students could examine literary critics' evaluations of creativity in a literary text to see if these evaluations can be verified or falsified. This can also benefit students (especially students of a second language or foreign literature) in testing their own evaluations of literary texts and thus contribute to their learning autonomy.

Literary reading is of course a highly complex phenomenon. 'Literary schemata' will also affect our evaluation of whether something

succeeds or not in its creation of effects. In their reception of a text, readers are expected by writers to take account of the form of a poem (or novel or play) and its associated expectations, which are often established over time. So what also needs to be measured is something generic as well as what exists in the language as a whole. This is not, however, something that can easily be done by corpus means. For a comprehensive exploration of literary evaluation, mixed-method interdisciplinary research would be needed – something of an ideal which would only really be possible as part of a well-resourced research project. While large corpus exploration in relation to a poem, like any method, ultimately only offers partial insights, it is both convenient and powerful for indicating the following: the degree to which what is activated in a literary critic's reading is likely to be shared by readers generally, as well as being useful in substantiating (or not) evaluation of the cognitive effects of a literary work.

Acknowledgements

References

Adcock, F. 2000. Poems 1960–2000. Tarset, Northumberland: Bloodaxe Books Ltd.

Adolphs, S. and R. Carter. 2002. 'Corpus stylistics: point of view and semantic prosodies in *To The Lighthouse*', Poetica 58, pp. 7–20.

Attridge, D. 2004. The Singularity of Literature. London: Routledge.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. Longman Grammar of Spoken and Written English. Harlow, Essex: Pearson Education.

Carter, R. 2004. Language and Creativity: The Art of Common Talk. London: Routledge.

Channell, J. 2000. 'Corpus-based analysis of evaluative lexis' in S. Hunston and G. Thompson (eds) Evaluation in Text: Authorial Stance and the Construction of Discourse, pp. 38–55. Oxford: Oxford University Press.

Cook, G. 1994. Discourse and Literature: The Interplay of Form and Mind. Oxford: Oxford University Press.

Cook, G. 2000. Language Play, Language Learning. Oxford: Oxford University Press.

Deignan, A. 2005. Metaphor and Corpus Linguistics. Amsterdam: John Benjamins.

Fowler, R. 1996. Linguistic Criticism (Second edition). Oxford: Oxford University Press.

Goodman, S. and K.A. O'Halloran (eds). 2006. The Art of English: Literary Creativity. Basingstoke: Palgrave.

Hunston, S. 1995. 'A corpus study of some English verbs of attribution', Functions of Language 2, pp. 133–58.

Hunston, S. 2001. 'Colligation, Lexis, Pattern, and Text' in M. Scott and G. Thompson (eds) Patterns of Text: In Honour of Michael Hoey, pp. 13–33. Amsterdam: John Benjamins.

Hunston, S. 2002. Corpora in Applied Linguistics. Cambridge: Cambridge University Press.

Jakobson, R. 1960. 'Closing statement: linguistics and poetics' in J.J. Weber (ed.). 1996. The Stylistics Reader: From Roman Jakobson to the Present, pp. 10–35. London: Routledge.

Jones, S. and J. Sinclair. 1974. 'English lexical collocations', Cahiers de Lexicologie 24, pp. 15–61.

Louw, B. 1993. 'Irony in the text or insincerity in the writer?' in M. Baker, G. Francis and E. Tognini-Bonelli (eds) Text and Technology. In Honour of John M. Sinclair, pp. 157–76. Philadelphia and Amsterdam: John Benjamins.

Louw, B. 1997. 'The role of corpora in critical literary appreciation' in A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (eds) Teaching and Language Corpora, pp. 240–51. London: Longman.

Maybin, J. and J. Swann (eds). 2006. The Art of English: Everyday Creativity. Basingstoke: Palgrave.

Moon, R. 1998. Fixed Expressions and Idioms in English. Oxford: Clarendon Press.

Mukařovský, J. 1932. 'Standard language and poetic language' in D.C. Freeman (ed.). 1970. Linguistics and Literary Style. New York: Holt, Rinehart and Winston.

O'Halloran, K.A. 2007. 'The subconscious in James Joyce's *Eveline*: a corpus stylistic analysis that chews on the 'Fish hook'', Language and Literature 16 (3), pp. 227–44.

Pope, R. 2005. Creativity. London: Routledge.

Schank, R. and R. Abelson. 1977. Scripts, Plans, Goals and Understanding. Hillsdale, NJ: Lawrence Erlbaum.

Schmitt, N. and R. Carter. 2004. 'Formulaic sequences in action: an introduction' in N. Schmitt (ed.) Formulaic Sequences: Acquisition, Processing and Use, pp. 1–22. Amsterdam: John Benjamins.

Schmitt, N., S. Grandage and S. Adolphs. 2004. 'Are corpus-derived recurrent clusters psycholinguistically valid?' in N. Schmitt (ed.) Formulaic Sequences: Acquisition, Processing and Use, pp. 127–51. Amsterdam: John Benjamins.

Semino, E. 1997. Language and World Creation in Poems and Other Texts. London: Longman.

Sinclair, J. 1991. Corpus, Concordance, Collocation. Oxford: Oxford University Press.

Sinclair, J. 2003. Reading Concordances. Harlow: Longman.

Sinclair, J. 2004. Trust the Text: Language, Corpus and Discourse. London: Routledge.

Starcke, B. 2006. 'The phraseology of Jane Austen's *Persuasion*: Phraseological units as carriers of meaning', ICAME Journal 30, pp. 87–104.

Stefanowitsch, A. 2006. 'Corpus-based approaches to metaphor and metonymy' in A. Stefanowitsch and S. Th. Gries (eds), Corpus-Based Approaches to Metaphor and Metonymy, Trends in Linguistics 171, pp. 1–16. Berlin and New York: Mouton de Gruyter.

Stubbs, M. 1996. Text and Corpus Analysis. Oxford: Blackwell.

Stubbs, M. 2001. Words and Phrases: Corpus Studies of Lexical Semantics. Oxford: Blackwell.

Stubbs, M. 2005. 'Conrad in the Computer: Examples of Quantitative Stylistic Methods', Language and Literature 14 (1), pp. 5–24.

Underwood, G., N. Schmitt and A. Galpin. 2004. 'The eyes have it: an eye-movement study into the processing of formulaic sequences in N. Schmitt (ed.) Formulaic Sequences: Acquisition, Processing and Use, pp. 153–72. Amsterdam: John Benjamins.

Wray, A. 2002. Formulaic Language and the Lexicon. Cambridge: Cambridge University Press.