

Surface text & the characterisation process: A case study of *Wyrd Sisters* by Terry Pratchett.

This research focuses on characterisation and how text and reader knowledge combine during this process. I am interested in exploring the role of surface text in the characterisation process and identifying how, or if, this is compatible with current models of characterisation. Specifically, I will extend and apply Culpeper's (2001) model of characterisation, which largely maintains the original assumption made by van Dijk & Kintsch (1983) that text processing involves the construction of propositions. In doing so, it lacks detail regarding the specifics of text-driven processing. Although Culpeper gives much detail about the features of text responsible for characterisation in his wider discussion of the model, these features are not adequately represented in the model itself. In other words, the existing model needs adapting or replacing in order to better represent how reader knowledge is triggered by texts and how this affects the application of this knowledge during the characterisation process. Like Culpeper, I am interested in the textual cues for characterisation, and the process by which readers assimilate these to form an overall idea of character. By expanding on or reworking these elements of the comprehension model, I will be able to apply the model more systematically to text analysis. In addition, focussing on the surface text and its links to reader knowledge also allows an exploration of the ways in which characterisation in these novels relies heavily on intertextual knowledge and how these links are made linguistically.

1. Applying a model of characterisation

Culpeper's model of characterisation (2001) provides significant insights into the processing involved in forming impressions of character. The model, like the original by van Dijk & Kintsch (1983) on which it is based, includes three levels. The first of these is situation, where prior knowledge about real and fictional people combines with textual cues that dictate how much knowledge of each type should be employed. The second is textbase, where prior knowledge dictates how much information about characters should be stored during reading. The third is surface, where linguistic features that combine with the situation level are processed syntactically and semantically to form textbase propositions. Both models illustrate that comprehension is cognitive, involving the construction of events based on discourse, with interpretation of these events occurring during discourse

processing, rather than after reading. Because of this, interpretation involves working hypotheses that can be adapted during processing. In addition, comprehension depends on the individual user, as their knowledge, beliefs, opinions and experience lead to the formation of inferences about the discourse events. These methods are used flexibly and take into account contextual and pragmatic factors during processing in order to create a representation of the social situation of the discourse for functional, communicative purposes.

However, the model lacks detail regarding how such contextual knowledge is triggered by surface text in the first place and how this affects the application of knowledge during processing. Culpeper (2001) does briefly address this, pointing out that van Dijk & Kintsch's original model of comprehension is too focussed on propositions and doesn't allow for other important factors in the comprehension of text. He believes that surface structure does not always dissolve into the propositions to form a textbase. This is especially true when reading literature, when relevant surface features may be included in the situation model. Culpeper provides an answer to why surface features are preserved in his discussion of explicit and implicit cues. These include important information about characters and so are likely to be preserved as it appears on the text, rather than propositionalised. What the model and subsequent discussion do not offer is an answer to how these features of surface text are preserved.

2. The role of propositions in character comprehension

2.1 What's wrong with propositions?

Consider the difference between the following representations of a line of text:

Proposition: MEET[WE, AGAIN, ?]

Surface text: "When shall we three meet again?"

(*Wyrð Sisters*: 5 & *Macbeth*: I, i: 1)

Although the pragmatic goal of arranging a meeting is still expressed in the proposition, much information essential to characterising the speaker and modelling the situation of which this question is a part is lost. Here, then, is an example of an instance in which a proposition is unlikely to be formed from the surface text. There are several theories, outlined below, that suggest similar issues with propositions.

A visual alternative

Emmott & Sanford (2012) suggest that understanding requires access not just to information represented in propositions, but to the information surrounding those propositions. They link this to the concept of mental models put forward by Johnson-Laird (1980) and use the following example as an illustration:

Andrew, Bill and Claire are sitting at a table. Andrew is to the left of Bill, and Bill is to the left of Claire. On which side of Claire is Andrew?

(Johnson Laird, 1983: 261, cited in Emmott & Sanford, 2012)

The key point here is that the answer is not in the propositional content of the text. If the table is round, then the answer is that Andrew is to the right of Claire. However, if everyone is sat along the same edge of a rectangular table, the answer is to the left. This does provide convincing evidence that a visual, or mental, model of the scene is created in order to answer the question. The following suggestion is particularly significant: 'following the mental models idea, the question becomes one of what linguistic cues help determine which mental model to adopt, and how mental models might result from the text itself' (Emmott & Sanford, 2012: 19). This supports the idea that the language of a text plays a far more prominent role in the overall impression formed by the reader during comprehension than many theories currently allow. After all, reading 'is not a mere complex of text-based propositions, it is an interpreted piece of language' (Emmott & Sanford, 2012: 20).

Propositions and literary processing

The idea that surface features are particularly important in literature and may not always be lost to propositions is well documented (Zwaan 1993, 1994 & 1996). Zwaan (1993) claims that reading literature involves creating strong surface and textbase representations, but relatively weak situation models. He goes on to explain that while literary readers are slower to construct a situation model, they do construct one. They simply do this at a later point, so as to retain more information from the surface text for use in later processing. According to Zwaan, the weak situation model that this strategy produces also explains the lack of constraint by consensus reality for literary texts, as a lack of definite situational knowledge regarding texts events means fantastic events are less likely to create a processing problem. Therefore, Zwaan's findings may be particularly useful when looking at the characterisation of fantastic characters like Pratchett's witches.

Rhetorical Focussing

Emmott & Sanford's (2012) Rhetorical Processing Framework (RPF) deliberately steps away from establishing basic meaning from text, often the focus of theories that suggest processing via proposition formation, in order to understand the comprehension of fictional narrative. The RPF states that style controls reader attention and, in doing so, controls the construction of mental models, including those for characters and their situations. Foregrounded language is key to this control of reader focus during text processing. Rhetorical focussing not only guides the reader to pay attention to specific parts of the text, it also means that some areas of text are processed in more detail than others (Emmott & Sanford, 2012). This provides a useful starting point from which to consider an enhanced role for surface text in the process of characterisation.

2.1 If not propositions, then what?

Sinclair's (2004) theory of units of meaning is helpful for exploring the role of surface language in processing. Here, he suggests that meaning is not held within a single lexical item but in commonly occurring chunks of language. This challenges the view that the meaning of words can be easily established by their definition, in isolation from their syntactic and semantic role. Idiom, phrase and collocation, rather than being confined to a linguistic 'rubbish dump' (Sinclair, 1991: 104), are central to language processing. Sinclair's ideas question the validity of models that view words as having clear propositional content that can form a knowledge base. When considering units of meaning, the message which may be encoded propositionally is only part of the overall meaning of a text. He argues that to fully understand a text, contextual and inferential information needs to be considered. Thus 'the propositional content of an utterance may on occasion be of minor interest compared with other aspects of delivery of the utterance - its relation to the previous utterance, for example, or its prospection of the next one' (Sinclair, 2004: 158).

Sinclair also rejects the usefulness of parsing as a primary method of linguistic comprehension with the suggestion that 'the realisation of meaning is much more explicit than is suggested by abstract grammars' (1991, 108). He uses the term 'open choice principle' to refer to approaches like this in which grammaticality is the only factor in language use and comprehension. Conversely, the 'idiom principle' refers to Sinclair's own approach of integrated and dependent word meaning in which 'a language user has available to him or her large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments' (Sinclair, 1991: 109). The conclusion is that it is not useful to use grammatical analysis for sections of a text

which seem to be constructed via the idiom principle. The likely scenario is that readers default to the idiom principle, only switching to rule-based analysis when necessary (for example in cases of ambiguity).

An example: 'not to mention'

As an example of the idiom principle, consider the phrase 'not to mention'. Data from the British National Corpus (BNC) shows 4457 instances of the word 'mention', in which 'mention' is used as a verb 285 times (6.4% of the total number of uses). Of these 285 instances, 35 appear with 'not' in the two slots before the node word 'mention' (12.3% of the total number of verb uses). This suggests that while use of 'mention' as a verb is relatively infrequent, its collocation as a verb with 'not' is more common. The only items which appear more frequently in the search are 'no' and 'to'. I discount the first of these because all 48 instances of 'no' are found in some form of the phrase 'no mention is made', in which 'mention' is not performing the role of the verb within the sentence. There are 42 instances of 'to', 28 of which are in the structure 'not to mention', giving further support for the collocation of these items.

Sinclair insists that the open choice and idiom principles are oppositional. The phrase 'not to mention' can also be used to illustrate this opposition. Idiomatically, it adds emphasis to additional information, as in the following example from the BNC corpus:

To rebuild after the devastation, will require huge construction projects. Not to mention re-equipping their armed forces.

However, when analysed grammatically a very different meaning is achieved, prohibiting someone from speaking about a subject. Only one example of this features in the BNC data:

The only thing you have to be careful not to mention is...

Both meanings are equally valid, although the idiomatic use is clearly more common than the grammatical in this particular data set. What is key to Sinclair's theory is the idea that both meanings cannot be applied to the phrase at the same time. Either the idiomatic or the grammatical meaning is understood, but not both simultaneously. The idiomatic use of 'not to mention' is, in fact, opposite in meaning to the literal. The latter forbids speaking of something, whereas the former actively emphasises it. The meaning of the lexical item 'mention' (defined in the Oxford English Dictionary as 'to refer to briefly and without entering into detail; to remark upon incidentally') is therefore lost in the idiomatic use. This illustrates Sinclair's claim that word meaning is often delexicalised and indeterminate, as it is difficult to pinpoint where in the comprehension process an accurate meaning is arrived at.

2.3 How can units of meaning be identified?

Most words, Sinclair (2004) claims, gain meaning from the structures in which they appear, meaning that lexical and syntactic patterns are often inseparable. Specifically, lexical meaning is derived from the words with which an item appears (collocation), the grammatical structures of which they are part (colligation) and associated meanings (semantic prosody). Partington (2004) provides clarification on the latter of these terms, pointing out that semantic prosody applies to units of meaning that frequently co-occur with items that carry a particular evaluative meaning. In addition to semantic prosody, he uses the term semantic preference to refer to units that share non-attitudinal semantic links. This is closer to lexical collocation, Partington claims, and simply denotes words that often appear with those from the same semantic set. Semantic prosody is therefore a more abstract term describing additional meaning achieved through an item's cumulative preferences for semantic sets with a particular type of prosody (often, but not always, positive or negative). He summarises his ideas:

semantic preference is a “narrower” phenomenon – relating the node item to another item from a particular semantic set – than prosody which can affect wider stretches of text. [analysis reveals] how they interact: the former, preference, contributes powerfully to building the latter, prosody; conversely, the latter dictates the general environment which constrains the preferential choices of the node item (Partington, 2004: 151).

Partington emphasises that prosody is a cline, with some units of meaning having stronger prosody than others.

The mental lexicon

Several researchers have expanded the idiom principle in order to explain variation within seemingly idiomatic units. Taylor (2012) argues that idioms, far from being merely fixed expressions, can be subject to syntactic and even lexical variation. This allows for more linguistic flexibility than many definitions of the idiomatic would allow, and so expands the concept of units of meaning. The need for such expansion can again be seen in data from the BNC, which shows that ‘not to mention’ can be changed, using a semantically related verb, to ‘not to speak of’. There are 28 instances of idiomatic use of this new variation in the first 100 of 557 results. This suggests that idiomatic phrases may allow more variation than at first thought. Taylor himself uses idioms that refer to being frightened as an example of such flexibility. The structure is commonly *to V the N out of someone*, but these verb and noun slots can be filled by a variety of lexical items. For example, *scare the life...*, *scare the*

shit..., *frighten the life...* are all common usages in the BNC corpus (Taylor, 2012: 77). Such structures are best understood as schematic in nature, having a generic structure on which to build varying content. This suggests that we learn language by encountering examples of it in use. Rules are then generalised once enough instances of a particular use have been acquired. This creates what Taylor labels a mental lexicon:

Knowledge of a language can be conceptualized in terms of the metaphor of the mental corpus. Language is acquired by a strictly bottom-up process, through exposure to usage events, and knowing a language consists, not in knowing a battery of rules, but in accumulated memories of previously encountered utterances and the generalizations which arise from them (Taylor, 2012: 263).

Lexical priming

Like Sinclair and Taylor, Hoey also emphasises the ‘pervasiveness of collocation’ (Hoey, 2005: 1). He claims that this can only be explained using the notion of lexical priming. Specifically, words are acquired through encounters with their use, which also include the contexts in which those words are used and the words with which they frequently co-occur. Priming is strengthened or weakened with each use of a word. Partington (2004) points out that it is through lexical priming that units achieve semantic prosody and preference, as well as pragmatic associations (a term introduced by Hoey). In other words, ‘language users have a set of mental rules derived from the priming process, alongside or integrated with the mental lexicon, of how items should collocate.’ (Hoey, 2005: 132).

Priming can be illustrated using the phrase ‘never mind’, which can be used as a less formal alternative to the idiomatic use of the ‘not to mention’. Of the first 100 concordance lines in the BNC search data for the phrase, 14 use ‘never mind’ in the idiomatic sense of ‘not to mention’. Therefore, ‘never mind’ can be said to be likely to be primed for some, perhaps most, people in this idiomatic sense. Hoey refers to changes in priming and language use as ‘drift’. In this context, ‘never mind’ to add emphasis to further examples, as in example 5 above, would be seen as a relatively traditional use. A drift in priming means that ‘never mind’ is now often used less formally in a dismissive or derogatory way either to disregard what has been said or to signal that the speaker thinks the hearer is too stupid to understand further explanation. Urbandictionary.com gives the following example:

Desi: Hey, do you know the real meaning of nevermind is?
Jesse: What?
Desi: You were too stupid to understand me the first time, so I gave up trying.
Jesse: What?
Desi: *sigh* Nevermind

There are fewer examples of this usage in the BNC data. The reason for this is likely to be that this is a more recent use of the phrase and, as the BNC does not contain data from the 21st century, it is less likely to feature more modern usages. What the BNC data can support regarding this drift in usage is that 'never mind' colligates with discourse markers and conjunctions ('well', 'oh', 'but') that signal pause or concession. This could also be framed as a semantic association with these discourse markers suggesting frustration or disappointment. For example

1. I wasn't too happy about that, but never mind.
2. Got that? Well, never mind.

This semantic association is exaggerated in example 1, above, with a sigh. The prosody, or pragmatic association, of the phrase can thus be said to be largely negative, as it frequently performs a dismissive function (either of an addressee's understanding of a situation or concept, or of the speaker's feelings regarding a situation).

Pattern grammar

Hunston & Francis (2000) provide the final development of Sinclair's ideas with their theory of pattern grammar. As with the theories outlined above, pattern grammar looks at the mutual dependence of lexis and grammar, claiming that 'all words can be described in terms of their patterns' (Hunston & Francis, 2000: 3). Unlike the other developments of the idiom principle discussed in this chapter, the focus of pattern grammar is on Sinclair's notion of colligation. A grammatical pattern is, the authors claim, associated with a particular sense of a word, meaning that patterns and lexical meaning are mutually dependent, restricting each other. While focussing on colligation may seem like a narrow approach, the authors claim that lexical phrases only account for part of the language, whereas pattern grammar attempts to describe the whole language. They realign their theory with Sinclair's ideas by pointing out that 'it is not patterns and words that are selected, but phrases, or phraseologies, that have both a single form and a single meaning. The outcome of this view is the idiom principle.' (Hunston & Francis, 2000:21). Using analogy, language users may replace a word in a pattern with one of similar meaning, even though the replacement word is not usually associated with the pattern being used. In this way, variability and novel use within this system is accounted for. This leads to a set of words with fuzzy boundaries being associated with any given pattern, and aligns the theory with Taylor's notion of the mental lexicon. Finally, the authors describe semantic prosody in terms of pattern grammar, pointing out that the 'tension between the meaning of the pattern and the meaning of the word used with the pattern can be exploited to imply meaning not explicitly stated' (Hunston & Francis, 2000: 108).

The phrase 'never mind the' + noun (a 'nested priming', to use Hoey's term for phrase development) illustrates the concept of pattern grammar. The final word in the phrase is subject to variation, but within the grammatical boundaries of the phrase (it must be a noun). Corpus data is not as helpful in analysing this particular use of the phrase 'never mind', as relevant examples are usually titles or names of products. Therefore, they are unlikely to be found contextualised within concordance lines with any frequency, even within a more up-to-date corpus (such as the NOW corpus). However, the prominence of this pattern in modern culture can be seen using internet search engines. When the phrase 'never mind' is typed into Google, the examples 'Never mind the bollox' and 'Never mind the buzzcocks' are the first two options given. This suggests that, at least in the contexts that inform popular internet searches, the phrase colligates strongly with 'the' + noun. The nested nature of this pattern is confirmed when a search for 'never mind the...' is performed. Again, 'buzzcocks' and 'bollox' are the top two results. After further results related to these two items, 'never mind the danger' and 'never mind the bike shops' are the next suggestions. Regardless of the products these names represent (a sporting venue and a bike shop, respectively), the choice of name for all three provides evidence of the pattern grammar associated with 'never mind' that stems from its original use as an album title. As a punk album, this automatically carries semantic and pragmatic preferences of rebellion and popular culture. This web of meaning and use can be accessed by any products that use the pattern 'never mind the' + noun, supporting Hunston and Francis's claim that patterns carry implicit meaning.

3. Text analysis

3.1 Units of meaning

The following short extract from Terry Pratchett's *Wyrd Sisters* will be used to apply the ideas outlined so far to characterisation. The section highlighted in bold will undergo linguistic analysis, with the whole extract given for context:

In the middle of this elemental storm a fire gleamed among the dripping furze bushes like the madness in a weasel's eye. It illuminated three hunched figures.
As the cauldron bubbled an eldritch voice shrieked: "When shall we three meet again?"
There was a pause.
Finally another voice said, in far more ordinary tones: "Well, I can do next Tuesday."
(*Wyrd Sisters*, 5)

Figure 1 presents the text in both propositional and unit form.

Proposition (P)/ Unit (U) number	Proposition content	Surface text content
1	BUBBLE[CAULDRON]	as the cauldron bubbled
2	SHRIEK[VOICE], P3]	an eldritch voice shrieked
3	MEET[WE, AGAIN, ?]	when shall we three meet again?
4	BE[PAUSE]	there was a pause
5	SAID[[ANOTHER]VOICE, P6, 7]	Finally another voice said
6	IN[[ORDINARY]TONES]*	in far more ordinary tones
7	DO[I, NEXT[TUESDAY]]	well , I can do next Tuesday

**When following strict rules of proposition formation, any descriptive language would be removed. This would remove P6 entirely.*

Figure 1: Propositional & surface text content

This section is concerned with justifying why the text has been split into the units it has in Figure 1. I have used my own intuition supported by corpus data in order to identify the meaning units present in the surface text. As Hoey (2005) points out, any intuitive analysis will reflect lexical priming, and so can be a useful tool for exploring the idea of priming itself. Such analysis will only be valid, however, if it can be verified by more objective means. According to Sinclair (1997), while intuition is an important asset regarding word meaning and sentence well-formedness in isolation, 'it is not, however, reliable about the way words and sentences are combined in actual communication' (Sinclair, 1997: 32). Therefore, any intuitive analysis needs to be checked against corpus data to compare it with actual language use. These checks will serve to address the potential problems of a purely intuitive analysis, noted by McEnery et al (2006), who agree with Sinclair that intuition can be a poor guide to collocation due to differing linguistic knowledge, preferences, memory and imagination of language users. In addition, they claim that most people overlook ordinary structures but notice unusual ones, giving a highly subjective picture of a text's language and meaning. Corpus analysis is therefore necessary as 'the corpus represents both a resource against which to test such intuitions and a motor which can help to generate them' (Partington: 1998: 2).

For this analysis, a slightly different initial approach to applying the idiom principle is required. Specifically, Sinclair (2004) outlines corpus analyses that begin with node words and work outwards through collocation, colligation, semantic preference and semantic prosody. However, the processing divisions outlined above are, arguably, not yet focussed on specific phrases. Therefore, perhaps the most obvious means by which to justify the division of the surface text content here is with reference to pattern grammar, rather than

collocation. For example, Units 1 and 2 are formed primarily as they occupy the two halves of the pattern 'as N + V, N + V'. In other words, we expect a second event to be presented after the first due to the use of the initial conjunction *as*, and these two events therefore form units of meaning. Similar justifications based on the pattern grammar of each unit can be made. Although this is arguably similar to parsing, the pattern identified will be processed by the reader as the structural basis for a unit of meaning, in the manner suggested by pattern grammar theory.

3.2 Pattern grammar & collocation

Collocation and pattern grammar (or lack of) also plays an important role in focussing attention in the extract. Intuitively, 'eldritch' and 'shrieked' seem like uncommon choices within these units. This is confirmed by data from the BNC. 'Shriek' appears 154 times in the data, 26 times as a verb, 128 as a noun. Of the overall instances, 153 are found in the fiction section of the corpus. This suggests that 'shriek' is more common as a noun, though not uncommon as a verb. The verb form 'shrieked' appears 177 times in the data, but collocates with 'voice' only once ('inner voice shrieked') and never with 'eldritch' in the two slots before the verb. Only four adjectives appear in the two slots before 'shrieked' (possessed, locked, inner & fat), each appearing only once, so no significant collocational or semantic associations can be made based on this. The data also contains 16 instances of 'eldritch', 5 of which are proper nouns (a surname). Of the 11 remaining instances, 10 appear directly before a noun, as in the *Wyrd Sisters* extract. Of these, three are semantically associated with 'voice': 'shriek', 'screech' and 'cry'. One collocates with the noun 'shriek', hinting at some (uncommon) collocation, but none appear with 'voice'. However, it is common for the noun 'voice' to be modified with an adjective (2298 instances of this are cited in the BNC). This suggests that the *Wyrd Sisters* extract uses common pattern grammar, foregrounding with more unusual collocation and relying on language almost entirely associated with fictional writing to do so. In doing so, Pratchett plays on the reader's literary competence for his characterisation.

Conversely, 'cauldron bubbled' does not seem intuitively unusual. However, of the 153 instances of 'cauldron' in BNC, only 5 collocate with 'bubble' in any form (all 'bubbling'). Four of these instances are an adjective + noun structure ('bubbling cauldron'), with only one using the noun + verb of the text example. There are also 10 instances of semantically related adjectives (boiling, steaming, suppurating, seething, foaming). Of these instances, three of 'bubbling' appear in the fiction section of the corpus, along with six of the semantically related adjectives. This indicates no strong literary association with the collocates used. This is surprising when we acknowledge that, rather than being foregrounded due to unusual language use or drawing on semantic associations of literary

description, the unit is foregrounded due to the intertextual knowledge it draws on, as it echoes the line 'fire burn and cauldron bubble' from *Macbeth*, Act IV, scene I. Such intertextual references are explored in more detail below.

3.3 Pragmatic context

Unit 7 ('well, I can do next Tuesday') sees attention focussed using more pragmatic means. This is firstly due to the formulaic use of language signalled by the discourse marker, which also expresses informality, likely to be interpreted as an attitudinal response to the formality of the question in Unit 3. This is supported by a sample of BNC data, which shows 74 of 100 uses of 'well' in the spoken corpus used as a discourse marker in this way, as opposed to only 62 in the fiction corpus (all of these found in representations of speech). It also carries the semantic and/or pragmatic associations of pausing for thought. This may be either to consider a response or, possibly in this case, to consider the style of the question. The narration in Unit 6, which introduces the reply as 'far more ordinary', and the use of 'do' instead of a more situation-specific verb (for example, *meet* or *see you*), adds to this effect. Long's (1994) idea that verbatim memory of surface text increases in oral contexts, which also see an increased use of formulaic language, is useful here. As spoken discourse carries information about reader goals, attitudes and emotions (Long, 1994), the oral style of this unit, along with its relatively formulaic use of language, lead to increased memory of this section of text. Whoever is speaking in this extract is in a battle for control over the style of the interaction, giving information about the goals and attitudes of the speakers. The significance of formulaic language in this extract can also be applied to the use of intertextual quotations, explored in the next section.

3.4 Intertextual units of meaning

Common patterns, collocations and pragmatic meaning have so far proved useful. However, further explanation is needed for some of the units featured in this analysis. Here, cultural and intertextual knowledge is also key. As Stubbs explains, 'our knowledge of language is not only a knowledge of individual words, but of their predictable combinations, and of the cultural knowledge which these combinations often encapsulate' (Stubbs, 2001: 3). The importance of cultural knowledge of character type is emphasised in Unit 3, which is foregrounded in a different way to those units already discussed. It is not unusual collocation or colligation, or the signalling of a particular register, that focusses attention here. Rather, Unit 3 exists regardless of its linguistic collocation or colligational content as it

is a direct quotation from *Macbeth*. The likelihood of readers making such intertextual connections is highlighted in the number of Goodreads.com reviews of *Wyrd Sisters* that mention Shakespeare in general, and *Macbeth* specifically. For example, the novel is summarised neatly, if superficially, in Susanna G.'s comment (Goodreads.com 2.10.2016): 'Hamlet and Macbeth had a baby, and it was a comedy'.

In triggering such associations with other texts and cultural knowledge, the quotation in Unit 3 confirms the character clues we are given earlier in the extract regarding the 'three hunched figures'. Three figures around a cauldron are likely to be witches, based on cultural knowledge; in *Macbeth*, the characters who are presented in such a setting are confirmed as such. These units are relevant to the activation of associated character knowledge using intertextual knowledge of character type. The construction of propositions for such extracts would be unhelpful as it is the stylistic choice of the phrase itself that creates the association. Intertextual associations with *Macbeth* are reinforced throughout *Wyrd Sisters* with such stylistic choices: "When shall we three meet again?" (5), "Something comes," she said. "Can you tell by the pricking of your thumbs?" (14), "Your peasant magic is for fools, mother of the night" (15). The first extract here is a direct lift from *Macbeth* (Act I, scene I, line 1). The second is a rephrasing of the statement 'by the pricking of my thumbs, something wicked this way comes' (Act IV, scene I, lines 44-45). The last echoes Macbeth's greeting of the witches with 'How now, you secret, black, and midnight hags' (Act IV, scene I, line 47). Given even a rudimentary knowledge of the play or other cultural references to and parodies of it, these extracts will activate knowledge linking these characters and the text as a whole to *Macbeth*, or perhaps to Shakespearean plays in general if knowledge is less precise. These meaning units in the surface text trigger intertextual associations that cannot be accounted for by following the formation of propositions. In fact, modelling the comprehension of such units using propositions would completely lose the intertextual references necessary for characterisation.

Conclusion

This analysis suggests that propositions, the traditional way of representing how information passes from text to brain via comprehension, are not a useful way of modelling the comprehension of character. The units of meaning proposed in the Idiom Principle provide a more accurate way of visualising characterisation. This is because they reflect natural language processing and production more closely using the notion of lexical priming, which attaches meaning to groups of words regarding lexical association (collocation), grammar (colligation), semantics (preference and prosody) and pragmatic functions. Units of meaning also allow a better approach to explaining how cultural and intertextual knowledge about character is triggered by text – an element often essential to characterisation.

References

The British National Corpus, version 3 (BNC XML Edition) (2007) Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>

Culpeper, J. (2001). *Language and characterisation: People in plays and other texts*. Harlow: Longman.

Emmott, C. and Sanford, A. J. (2012). *Mind, brain & narrative*. Cambridge: CUP.

Goodreads. Goodreads Inc. https://www.goodreads.com/book/show/34504.Wyrd_Sisters?

Hoey, M. (2005). *Lexical priming: a new theory of words and language*. Retrieved from <https://ebookcentral.proquest.com>

Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, 12(2), 249–268. <https://doi.org/10.1075/ijcl.12.2.09hun>

Hunston, S. and Francis, G. (2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Retrieved from <https://ebookcentral.proquest.com>

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science: A Multidisciplinary Journal*, 4(1), 71–115. http://doi.org/10.1207/s15516709cog0401_4

Kintsch, W., and van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394. <http://doi.org/10.1037//0033-295X.85.5.363>

Long, D. (1994). The effects of pragmatics and discourse style on recognition memory for sentences. *Discourse Processes*, 17, 213-234.

McEnery, T., Xiao, R., and Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.

Partington, A. (1998). *Patterns and meanings: using corpora for English language research and teaching*. Retrieved from <https://ebookcentral.proquest.com>

Perfetti, C. A. & Britt, M. A. (2012) Where do propositions come from?, in Weaver, C., Mannes, S. and Fletcher, C. (eds). *Discourse Comprehension*, pp. 11-34. Florence: Taylor and Francis. Retrieved from <http://ebookcentral.proquest.com/lib/hud/detail.action?docID=1099340>

Pratchett, T. (1994). *Wyrd Sisters*. London: Corgi.

Sinclair, J. (1991) *Corpus, concordance & collocation*. Oxford; OUP.

Sinclair, J. M. (2004). *Trust the text: Language, corpus and discourse*. New York, N.Y: Taylor & Francis.

Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell.

Taylor, J. R. (2012). *The Mental Corpus*. Oxford: OUP.

Urban Dictionary. <https://www.urbandictionary.com/define.php?term=Nevermind>

Zwaan, R. a. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 920–933. <http://doi.org/10.1037/0278-7393.20.4.920>

Zwaan, R. A. (1996). Toward a model of literary comprehension. In Britton, B. K. & Graesser, A. C. (Eds). *Models of understanding text*. (241-255). New Jersey: Lawrence Erlbaum.