CONRAD, CONCORDANCE, COLLOCATION:
HEART OF DARKNESS OR LIGHT AT THE END OF THE TUNNEL?

Michael Stubbs

FB2 Anglistik, University of Trier, 54286 Trier, Germany

ABSTRACT

Since the 1960s, linguistic stylistics has passed through several stages, from Roman Jakobson's optimism that literary texts could be objectively analysed, followed by Michael Halliday's detailed grammatical analysis of a novel by William Golding, followed by Stanley Fish's criticisms that such analyses are circular and arbitrary, followed by responses to these criticisms. In the mid-1960s, John Sinclair argued that a literary text has meaning only "as a sample of an enormously large body of text". This observation still awaits detailed study, but it can now be investigated with the corpus techniques which he has more recently developed. I will discuss some linguistic features of Joseph Conrad's *Heart of Darkness* (1899/1902), including patterns of phraseology in the text itself, and their relation to patterns in the language as sampled in large computer-readable corpora.

The title of the lecture is a reference to John Sinclair's book *Corpus, Concordance, Collocation* (1991), which showed how computer-assisted corpus analysis reveals the central place of phraseology in language use.

PRESENTATION CONVENTIONS

Page references to *Heart of Darkness* are to the (identically paginated) Penguin Modern Classics (1973) and Penguin Popular Classics (1994) edition. All linguistic examples which are italicized or presented on separate lines are from *Heart of Darkness*, unless otherwise stated. Frequencies of words and phrases are given in diamond brackets, e.g. <freq 10>. Lemmas (lexemes) are in upper case, and word-

forms are in lower case: e.g. the lemma SEEM consists of the word-forms *seem*, *seems*, *seemed*, *seeming*. It is sometimes convenient to treat different parts of speech as members of the same lemma (e.g. *silence*, *silent*), and sometimes revealing to separate them (e.g. *knowledge* [noun] versus KNOW [verb]). This question is unresolved in lexical theory.

> "I did not see the real significance ... I fancy I see it now, but I am not sure - not at all. ... I had to keep a look-out for the signs ... I've been telling you what we said - repeating the phrases ... They were common everyday words." (*Heart of Darkness*, pp.30, 49, 95.)

1. THE "SINCLAIR" LECTURE

Since the 1960s, John Sinclair has worked on several aspects of text and corpus analysis: on individual short texts (mainly stylistic analyses of poems), on text-types (such as classroom discourse), and on large computer-readable corpora. But he has done relatively little work which relates these areas to each other - that is, until recently. In April 2002 - a year before I gave the Sinclair lecture - I was being pressed to provide a title for my talk. I had more or less decided to discuss how corpus analysis could contribute to stylistics. A text that I had been using in teaching stylistics was Joseph Conrad's *Heart of Darkness*. It was John who first pointed out to me how interesting the text is, and I had a vague memory of discussing the book with him. So, it seemed a good choice for a lecture in his honour. I e-mailed him, to ask if he had ever published anything on the book - without telling him why I was asking - and he e-mailed back immediately that he had the book on his desk, and was at that very moment starting to prepare a talk on *Heart of Darkness* for a stylistics conference. I discovered, that, as in many other areas, John has been there before us.

When I had recovered from this setback - and had wondered whether to abandon my plan and do something completely different - I decided that the exercise could be interesting. If we worked independently, with comparable methods, and said something similar, then that would be interesting. But then it would also be interesting if we said different things. And there is a deep problem here. A traditional aim of stylistics is to make objective descriptive linguistic statements about texts. But as Michael Toolan (2002) recently pointed out, if ten stylisticians are invited to write about the same text, say for a special issue of a journal, they are most unlikely to discuss the same things. So, there are deep questions here about selective attention to data, objectivity, replicability and interpretation. Some months later - in November 2002 - John and I compared notes on the planned content of our lectures, and discovered that there was indeed rather little overlap in the stylistic features we had chosen to discuss.

Anyway, possibly in a slightly reckless mood, I decided to go ahead and discuss whether some rather simple quantitative methods could say anything interesting about Conrad's book.

2. THE STORY

First, here is a brief reminder of the story and some general comments about the book. The main narrative is embedded in different frames:

1. The book starts with an unnamed narrator on a boat on the Thames.
    2. Marlow becomes the narrator, and talks about the Thames in Roman times.
        3. Marlow tells of his visit to a European city.
            4. Marlow tells the main story, which takes up most of the book: he travels up a river in Africa in search of an ivory trader called Kurtz. He finds him, but Kurtz dies on the trip back down river.
        5. Marlow tells of his visit to Kurtz's fiancée back in the European city.
    6. [There is nothing corresponding to frame 2. But some vocabulary from frame 2 is repeated in frame 7.]
7. The book ends with a paragraph from the unnamed narrator back on the Thames.

One feature of the book is that major places are never named. We know that it is based on Conrad's own experiences of going to Brussels, where he arranged to travel to the Belgian Congo, and up the River Congo. Even the word *Africa* appears only once, when Marlow is discussing maps which he looked at as a child. It is not difficult to infer that the city is Brussels (it is a city on the continent where they speak French), and it can be inferred from encyclopedic knowledge (the references to yellow patches on a world map) that Marlow is talking about Belgian territory in Africa. I will sometimes talk of "Brussels" and the "Congo", but please bear in mind that these words never appear in the book. Of course, if words don't appear in the text, then the reader has to supply them (and the computer can't find them).

In the main story, Marlow takes a boat up a river. The original purpose is never entirely clear, but the journey turns into Marlow's search for - and obsession with - a man called Kurtz, a trader who has been stealing ivory from the inhabitants of the region. He has apparently gone mad, is worshipped as a god by the native population, seems to have an African mistress, and may have been implicated in cannibalism.

3. A LOT OF LITERARY CRITICISM ...

Heart of Darkness was published just over 100 years ago (in 1899 in magazine instalments, then in 1902 in book-form). It is a very short novel, really a long short story of less than 40,000 words. It is well known, and easily accessible, in both printed and computer-readable versions (which means that you can check my findings). There are carefully edited versions which identify references to contemporary texts and the autobiographical sources in Conrad's *Congo Diary*, and it has been studied in detail by literary critics.

Indeed it is a key twentieth century text, which is still constantly referred to after 100 years. It is possibly "the most commonly prescribed novel in twentieth century literature courses ... in American universities" (Achebe 1975/1988). Though, given recent American foreign policy, it seems that not many Americans have read it very carefully: nor do they seem to have learned anything from the film *Apocalypse Now*, which is based loosely on the book, but moves the action to Viet Nam.

One reason I was interested in the book was that two of the professors of English literature in my university in Germany had just written articles on it (Breuer 1999, Klooss 2002). This all raises the intriguing puzzle of why the text has been so popular for so long. But it also means that it looks like one of the worst possible books to try formal linguistic analysis on, since it seems unlikely that I could say anything new about a text which has been intensively studied for 100 years.

It also looks like a poor choice for formal linguistic analysis in so far as it is an overtly political text about late Victorian ideas of Empire, which has been read from different political positions. It was famously attacked by Chinua Achebe (1975) for presenting stereotyped and racist views of Africans, and defended by others, who point out that the narrator Marlow is at least very critical of the more unpleasant sides of colonial exploitation, which he describes as a *sordid farce* and *senseless delusion* (p.19). All this is a demonstration, if we need one, that books can be interpreted in different ways, and that stylistic analysis will not provide a definitive reading. (For a more recent interview with Achebe, see Phillips 2003.)

4. ... BUT VERY LITTLE LINGUISTIC ANALYSIS

However, despite a hundred years of critical discussion, there is surprisingly little on the book's linguistic style. More generally, few attempts have been made to relate stylistics and statistics. So I decided - again possibly slightly recklessly - to give it a try.

It is often argued (e.g. Widdowson 1975: 159) that literature uses language in ways which cannot be accounted for by the theories developed by linguists. I am certainly not going to argue that a purely automatic stylistic analysis is possible. The linguist selects which features to study, the corpus linguist is restricted to features which the software can find, and these features still have to be given a literary interpretation. However, Conrad expressed his ideas through the language of the text. There must therefore be things in the text which the computer can find, and I will claim that methods of formal analysis can identify textual features which critics seem not to have noticed. In addition, as John Sinclair (1975) has said, there must be something wrong with linguistic theories if they cannot explain the language of those texts which are regarded as the highest achievement of literary expression.

5. MAJOR THEMES

Given the huge amount of criticism of the book, we cannot approach it as naive readers (though the computer can ...), and I assume there is consensus about some of its major leitmotifs and themes, including the hypocrisy of the colonisers, and breakdown: as a symbol of the unreliability of technological progress. Marlow's boat keeps breaking down; the colonial outposts are littered with *decaying machinery* (p.22); Kurtz has a mental breakdown (so does a Dane called Fresleven). There are breakdowns in communication: people speak different languages and dialects; Marlow tells a lie about Kurtz to Kurtz's fiancée; and amongst the most frequent content words in the book is the lemma *SILENCE* <freq 37> (*silence* 26, *silent* 8, *silenced* 1, *silencing* 1, *silently* 1).

Other major themes of the book are conveyed by repeated contrasts, especially light and dark, restraint and frenzy, and appearance and reality. There are frequent references to dreams (p.48), nightmares (p.100), trances (p.56), phantoms, apparitions and visions (pp.85, 87, 105, 110); Marlow has trouble maintaining *contact with reality* (pp.19, 54), and asks himself *what it all mean[s]* (p.33).

Critics have pointed out that Conrad uses these contrasts ironically. For example, he is questioning whether *heart of darkness* refers to "darkest Africa", as the stereotype has it, or rather to the immorality of the white colonialists. The phrases *Through the Dark Continent* and *In Darkest Africa* were the titles of books published just a few years before Conrad was writing (in 1878 and 1890), by Henry Morton Stanley (the Stanley who "found" David Livingston). *Heart of Darkness* is based around a series of problematic and false contrasts: Europe and Africa; the Thames in Roman times and the Congo under Belgian imperialism; Marlow and Kurtz; Kurtz's black mistress and his white fiancée; lies and truth (Reeves 1985: 301).

6. ANTONYMS AND CONTRASTS

These major themes are often identified by literary critics. I now move on to linguistic observations about the text, under three main headings: first, some observations about different linguistic patterns; second, some quantitative observations about linguistic patterns; and, finally, some quantitative observations about phraseological patterns. These quantitative observations can then relate this particular text to general language use.

First, patterns of antonymy are much more pervasive than the few contrasts which I have listed so far. The opening three pages, where the speaker is the anonymous narrator, have a particularly large number of lexical contrasts. Just a few examples (pp.5-7) are

* the sea and the sky
* from glowing white ... to a dull red
* a short day that comes and departs
* borne to the rest of home or to the battles of the sea
* knights all, titled and untitled
* the ships and the men ... the adventurers and the settlers ...hunters for gold or pursuers of fame ... the sword and ... the torch

The episode (pp.14-16) in which Marlow visits Brussels to arrange his contract has many antonymous pairs:

* whited – black, right – left, fat – slim, plain – all the colours of the rainbow, skinny – plumpness, younger – old, feverishly – placidity

The last episode (pp.106-10), in which Marlow meets Kurtz's fiancée, provides another burst of antonymous pairs, all within close proximity

* floor – ceiling, opened – closed, black – pale, young – mature, dark – pale, rich – pauper, glow – darkness, happy – unhappy, died – lived

Antonyms are a formal feature of the text, but it is difficult to state their exact literary effect. First, they can be interpreted differently in the case of the two narrators. The anonymous narrator is presented as naively in favour of British imperialism: *What greatness had not floated on the ebb of that river [= the Thames]!* (p.7). His antonyms perhaps express inclusiveness: *the sea and the sky*. Marlow expresses a very different view of imperialism in his very first words (p.7): *And this [= the Thames] also has been one of the dark places of the earth*. Perhaps his antonyms suggest his confusion between reality and appearance? that everything implies its opposite? that nothing is itself? In addition, antonyms are a feature of many texts: several corpus studies (Justeson & Katz 1991, Fellbaum 1995, Jones 2002) show that antonymous lexical contrasts within sentences are much more frequent than could be expected by chance, and are an important mechanism of textual cohesion. I return shortly to the unsolved stylistic problem of relating formal textual features to literary interpretations.

7. VAGUE IMPRESSIONS AND UNRELIABLE KNOWLEDGE

A major theme of the book is Marlow's unreliable and distorted knowledge. He is a notoriously unreliable narrator. His journey is an anti-climax: we never quite find out what has happened, or what Kurtz has done, what *monstrous passions* (p.95) and *vile desires* (p.105) he has indulged. Marlow himself never quite understands what is going on: he complains that his experiences seem like a dream, and Kurtz never tells him anything very specific before he (that is, Kurtz) dies. Kurtz has talked at length to a young Russian, but he, in turn, is unable to tell Marlow what Kurtz has said. Kurtz dies uttering the words *The horror! The horror!*, but we never find out what that refers to: for example, is Kurtz now horrified by what he himself has done?

When I discussed the book briefly in e-mails with John Sinclair, he described it as "the longest shaggy dog story in the canon of English literature". The unnamed narrator in the outside frame is pretty ironic about this, and comments that *we knew we were fated ... to hear about one of Marlow's inconclusive experiences* (p.10). Critics have taken different views of all this: F R Leavis (1962: 180) thought that Conrad simply didn't know what he wanted to say. (So did E M Forster 1936: cited by Leavis, and in Kimbrough ed 1988: 240.)

But Ian Watt (1979) argues that it is sophisticated early modernism, and that lack of clarity and impressionism are part of the point of the story. He points out (Watt 1979) that "mist or haze is a very persistent image in Conrad", and words from this lexical field are frequent in the book (in total almost 150, well over one per page on average):

- blurred 2, dark/ly/ness 52, dusk 7, fog 9, gloom/y 14; haze 2, mist/misty 7, murky 2, shadow/s/y 21, shade 8, shape/s/d 13, smoke 10; vapour 1

(This is s*moke* in the relevant sense, excluding s*moking my pipe*, etc.) *Shade* and *shadow* are also used in the sense of ghosts and the dead, sometimes quite explicitly: for example, Marlow refers to Kurtz as *that Shadow - this wandering and tormented thing* (p.94), and to Kurtz's Intended as *a tragic and familiar Shade* (p.110). *Shades* is also an old term for Hades, the Underworld. For readers around

1900, Biblical references were probably also evident: in the Bible the phrase *shadow of death* is frequent <20>, and *shadow* collocates with *darkness* <8>.

Marlow is frequently looking into a fog, uncertain of what he is seeing. Unexplained things happen: a man hangs himself for no apparent reason (p.21). Things are *improbable, inexplicable, and altogether bewildering* (p.78). As he approaches Kurtz, Marlow is confronted with a series of signs that he cannot interpret or misinterprets: a message on a board (p.53), a book with strange writing (p.54), a figure dressed as a harlequin (p.75), and round carved balls on posts (which turn out to be human skulls, p.75). And then Conrad did have a problem (especially in Victorian England), in that whatever *unspeakable* things (p.71) it is that Kurtz has done, they must remain unspoken, because they would probably be ridiculous or blasphemous or pornographic, or all three.

Literary critics tend to identify content words, such as *fog* and *mist*, *vague* <frequency 5> and *indistinct* <4>:

- I saw <u>vague</u> forms of men
- Marlow ceased, and sat apart, <u>indistinct</u> and silent

However, they tend to ignore the many grammatical words denoting extreme vagueness and uncertainty. So, here is where I start to discuss quantitative aspects of style. The word *something* occurs over 50 times, in expressions such as:

- I don't know - <u>something</u> not quite right
- <u>something</u> like a lower sort of apostle
- reminded me of <u>something</u> I had seen - <u>something</u> funny

There are over 200 occurrences of *something*, *somebody*, *sometimes*, *somewhere*, *somehow* and *some*, plus around 100 occurrences of *like* (as preposition), plus over 25 occurrences of *kind of* and *sort of*, all of these often collocated with other expressions of vagueness:

- the <u>outlines</u> of <u>some sort of</u> building
- <u>seemed</u> <u>somehow</u> to throw <u>a kind of</u> light
- I <u>thought</u> I could see <u>a kind of</u> motion
- <u>indistinct</u>, <u>like a</u> <u>vapour</u> exhaled by the earth ... <u>misty</u> and silent before me

Things are frequently described, not as themselves, but as *something* that looks *like* something else:

- a <u>thing</u> <u>like</u> calipers
- <u>something</u> that <u>looked like</u> a dried gourd

If we add all this to frequent occurrences of *seemed* <ca 50, see below>, there is a high frequency of words expressing vagueness <ca 385>: well over three per page. The frequency of these vague words is consistently higher in *Heart of Darkness* than in a reference corpus of fictional texts, and much higher again than in a corpus of mixed written texts (see Appendix). I will comment later on the grammar of the many extremely vague place expressions.

8. STYLISTIC OPTIMISM AND PESSIMISM

In the 1960s, Roman Jakobson (1960) optimistically proposed "an objective scholarly analysis of verbal art". In the 1970s, Michael Halliday (1971) proposed an analysis of "stylistically relevant" linguistic regularities in a novel by William Golding. Stanley Fish (1973) attacked what he called "Halliday's machine", and accused such procedures of being "circular" and "arbitrary". In the 1990s, Michael Toolan (1990) and others responded to Stanley Fish and defended even simple quantitative data such as word-frequencies. Also back in the 1960s, John Sinclair (1965) argued that a literary text has meaning only "as a sample of an enormously large body of text". This observation still awaits detailed study, but it can now be investigated with the corpus techniques which he has more recently developed.

I will try and bear in mind what is a serious problem for stylistics: selective attention to data. Stanley Fish (1973) points to a logical dilemma in which he thinks stylistics is caught. Either it is arbitrary or circular or both. Either we select a few linguistic features, which we know how to describe, and ignore the rest. Or we select features which we already know are important, describe them, and then claim they are important. Since a comprehensive description is impossible, and since there is no way to attach definitive meanings to specific formal features, there is no way out. It is a genuine logical fork, ... which I will call the Fish Fork. ... Actually, I think there is a very simple response to the Fish Fork: namely that it applies to any study of anything. I'll come back to this later.

In the meantime, I'll just say that I will identify formal features which all express, in one way or another, the major theme of uncertainty and distorted knowledge, which Marlow makes quite explicit in one of his comments (p.39) to his listeners:

- Of course [...] you fellows see more than I could then.

9. SOME (VERY) SIMPLE FREQUENCY STUFF

One thing which corpus linguists know how to describe is word frequency. I will show in a moment that looking at individual words and their frequencies has severe limitations, but a word-frequency list is an essential starting point, since there must be some relation between frequent vocabulary and important themes, even if the relation is indirect.

Admittedly, a list of the most frequent content words in *Heart of Darkness* does not initially look very promising. Here are the top words:

- said 131, like 122, man 111, Kurtz 100, see 92, know 87, time 77, seemed 69, made 65, river 65, came 63, little 62, looked 56, men 51, Mr 51, long 50

Frequent names and nouns tend to indicate just the superficial themes of a text (*Kurtz*, *river*), and the most frequent word of all (*said*) is of course very frequent in fiction in general. But the list does capture other words which are of more interest: around 100 occurrences of *like* are in vague expressions such as "x was like y", and around 25 occurrences of *looked* are in vague expressions such as "x looked like y" or "it looked as though". I comment later on SEEM.

We have to be careful in interpreting such lists. Words such as *eyes* and *head* are frequent in many fictional texts, and also frequent in this book <*eye* 4, *eyes* 49, *head* 48, *heads* 12>, where they have a particular significance: Marlow is constantly being observed by eyes in the jungle, and there are the heads on the posts. Verbs are often a better candidate for stylistically relevant words: but again, the psychological verbs and verbs of perception (such as *see*, *know*, *looked*) are also usually frequent in fictional texts. (For frequency data, see Stubbs and Barth in press 2004.) In addition, non-fiction texts are usually "about" a given topic and this may be signalled quite accurately by the frequent vocabulary. In literary texts, frequent vocabulary will, at best, identify frequent characters and themes, but will not tell you what the text is about. The river is an important theme in *Heart of Darkness*, but it is not a book "about" a river.

We can also check which words are significantly more frequent in *Heart of Darkness* than in reference corpora. We can, for example, compare *Heart of Darkness* with other texts, using the "keywords" software in Scott (1997). If we compare *Heart of Darkness* with a reference corpus of mixed English fiction, then the top content words which are both frequent in *Heart of Darkness* <20 or more>, and also significantly more frequent than in the reference corpus are

- Kurtz 100, <u>seemed</u> 69, river 65, station 48, great 46, manager 42, earth 39, ivory 31, pilgrims 31, darkness 25, forest 23, bank 25, wilderness 22, Kurtz's 21, cried 20

The reference corpus was the 706,600 words of the "imaginative prose" categories K, L, M and N in Brown, LOB, Frown and FLOB (four million words of written British and American English).

We can also use software to lemmatize the text and look again at the most frequent verb lemmas. The top ten are:

- SAY, SEE, LOOK, KNOW, COME, MAKE, SEEM, HEAR, TAKE, THINK

The nouns in these lists still seem only to provide the basis for the crudest kind of content summary. But note for the present that the verb SEEM is amongst the top words in all of these lists, and several other verbs in the lists concern knowledge, perception and appearances. So, even very simple frequency lists do identify words which express major themes of the book.

10. WORD DISTRIBUTION AND TEXT STRUCTURE

One reason why counting individual words is certainly not sufficient is that the interesting content words cannot (by definition) be evenly distributed across the text, but are bound to be clustered at different places. Some do not occur until quite a long way into the book. For example, *Kurtz* is not named until about a quarter of the way through (p.27), and this simple fact about word distribution expresses part of the mystery about him.

Other words (e.g. *Buddha*) are only in the opening and closing frames, and in this way word distribution signals text structure. For example, the narrative frames are marked formally by repeated words and phrases. The book starts on the

Thames, describes a journey up the River Congo (unnamed), and ends back on the Thames. At the beginning and at the end we have:

- the Thames ... a <u>waterway</u> <u>leading to the uttermost ends of the earth</u> (pp.5-6) ... <u>Marlow</u> <u>sat</u> cross-legged ... he had the pose of a <u>Buddha</u> ... we felt <u>meditative</u> (pp.6, 10).
- <u>Marlow</u> ... <u>sat</u> apart ... in the pose of a <u>meditating</u> <u>Buddha</u> (p.111) ... the tranquil <u>waterway</u> <u>leading to the uttermost ends of the earth</u> (p.111, last sentence).

There are several other examples. At the beginning, Marlow visits Brussels, a city like a *whited sepulchre* (p.14). He enters offices, past *high houses* in a *narrow and deserted street*, through *doors standing ponderously ajar*. In the offices *a door open[s]* (p.14). The building is *as still as a house in the city of the dead* (p.16). At the end, Marlow returns to the *sepulchral city* (p.102). He visits Kurtz's Intended through a *ponderous door, between tall houses* in a street as quiet as *a cemetery*. In her house is a piano like a *sarcophagus*. A *high door open[s]* and the Intended comes in (p.105).

Such facts start to say something about the structure of the whole text, and we can track some of these distributions with a simple program. For example, the words *heart*, *dark* and *darkness* occur throughout the book, but they increase in frequency at the very end, when the story almost becomes *too dark - too dark altogether* (p.111).

Similarly, the lemmas DREAM <15> and NIGHTMARE <6> occur throughout the book, but are very differently distributed. DREAM occurs twice at the very beginning, then several times in a cluster on p.39 - Marlow is *trying to tell* his dream - then fairly regularly throughout the rest of the story. NIGHTMARE occurs once at the beginning, where there are *hints for nightmares* (p.21), and then again only in a cluster towards the end (ca p.95): all in collocations with Kurtz and/or wilderness. In terms of word distribution, Marlow's dream turns into a nightmare.

A more complex lexico-grammatical example is provided by the verb lemma KNOW. This is frequent <122>, and fairly evenly distributed throughout the text. Many instances are negative, either grammatically (*I don't know*, *he did not know*) or by implication (*he wanted to know*, *if only he had known*). This is a novel about the fallibility and distortions of human knowledge. Right at the end, there is a little burst of positive examples: Kurtz's Intended repeats over and over that

- I alone <u>know</u> how to mourn for him ... I am proud to <u>know</u> I understood him better than any one ... You have heard him! You <u>know</u>! ... You <u>know</u> what vast plans he had.

The irony is of course that she knows nothing of what Kurtz has done. (On how patterns of lexico-grammar can express the "fallibility of human knowledge", see also Hardy and Durian 2000.)

11. TEXT AND INTER-TEXT

Another reason why looking at individual words is not good enough is that any text refers to other texts. *Heart of Darkness* is a mixture of fragments of other texts. It is

a story of a journey, based explicitly on the metaphor of life as a journey: *the farthest point of navigation and the culminating point of my experience* (p.11), as Marlow calls it. I suppose we would not classify it as travel literature, though perhaps as a rather unlikely traveller's tale. It contains elements of a parable, an adventure story, a mystery story, and a fairy tale: characters are *bewitched* (pp.33, 39); men in the jungle are held *captive by a spell* (p.50); Marlow's journey up-river is *beset by [...] dangers as though [Kurtz] had been an enchanted princess sleeping in a fabulous castle* (p.61) ... not to mention dark forests, figures of devils, dreams and nightmares, and a reference to the *dance of death* (p.20).

There is no single coherent set of symbols. On the contrary, there are allusions to several ancient and modern myths and stories. References to Greek mythology include the figures of the Fates: two *fateful* women, not weaving, but knitting black wool (pp.15-16), and two young men in the office who are *piloted over* (p.15): a reference to the souls of the dead being ferried across the river Styx in Hades. References to Dante's depiction of hell include *the gloomy circle of some Inferno* (p.24). References to the Faust legend, in versions by Marlowe (no pun intended?) or Goethe include signing contracts (p.15), Mephistopheles (p.37), and making a bargain for one's soul with the devil (p.70), Kurtz is a figure in search of extreme forms of experience, and like Goethe's Faust, Kurtz dies blind, unable to see the light of a candle a foot in front of him (p.99). Lothe (2000, 2001) discusses these intertextual references, and also references to Virgil's *Aenead*.

There are also elements of black comedy, with fools and jesters engaged in absurd activities:

- the ridiculous clerk who claims he is *not such a fool as [he] look[s]* (p.16)
- the *harmless fool* (p.16) of a doctor who measures Marlow's skull
- the man like *a hairdresser's dummy* with a green parasol (pp.25-6)
- the brickmaker who makes no bricks (p.34)
- the *papier maché Mephistopheles* (p.37)
- and, most explicitly, the Russian (*I am a simple man* pp.84, 90) who is dressed as a *harlequin* or court jester (pp.75, 78).

There are frequent references to fools <*fool*, *foolish* 18> and to madness and insanity <*mad*, *madness*, *insanity* 11>, and therefore allusions to the ship of fools, the medieval satire on vices and follies. Marlow's steamboat has absurd passengers, a *little fat man* with *red whiskers* and *pink pyjamas* (p.57) and a *mad helmsman* (pp.63, 64, 65, 73). *Narrenschiff* by Sebastian Brandt (1494) was adapted as *The Shyp of Folys of the Worlde* (1509) by Alexander Barclay, and as *Modern Ship of Fools* by W H Ireland (1807). (Foucault 1961 describes the real *Narrenschiffe*.)

The words *fool* and *mad* occur frequently, but the linguistic allusions to Dante, Goethe and so on are sparse and not identifiable in any automatic way by computational techniques. However, there are quotes and near-quotes from specific texts which can quickly be identified by computer-assisted searches. The reference to Brussels as a *whited sepulchre* (p.14) is a quote from Matthew, 23, 27: "ye are like unto *whited sepulchres*, which appear beautiful outward, but are within full of dead men's bones, and of all uncleanness". There are frequent references to devils <20>, pilgrims <34>, disciples <2> and temptation <2>. Some Biblical references were probably more obvious to readers a hundred years ago. In the opening and closing paragraphs (pp.5-6 and 111) of the book, we have the phrase

- waterway leading to the <u>uttermost</u> ends of the earth

The word *uttermost* is not frequent in general English (or in Conrad's other writings), but the phrases *uttermost part(s) of the earth* <29> and *the ends of the earth* <30> are frequent in the King James translation of the Bible. I will come back to such abstract place terms later.

There are phrases which allude to Charles Dickens' *Tale of Two Cities* (1859). The knitters of black wool recall the women who sit at the foot of the guillotine: *In front of it (= the Guillotine) ... a number of women, busily knitting. ... the women sat knitting, knitting. Darkness encompassed them. Another darkness was closing in as surely ...the women sat knitting, knitting ... knitting, knitting, counting dropping heads*. And there are phrases which allude to Jules Verne's *Voyage au Centre de la Terre* (1864). Marlow sets off as if on a journey to *the centre of the earth* (p.18); he is a *wanderer on a prehistoric earth*, as though the earth was *an unknown planet* (p.51). This recalls phrases in Verne's novel such as: *to the very centre of the earth ... a vision of the prehistoric world ... heavy gloom of deep, thick, unfathomable darkness*. (Verne's novel was published in 1864 in French, Conrad's first foreign language, and a few years later in English translation.)

The themes of cannibalism and of atavism (the fear that apparently "civilized" humans could revert to a more primitive type) were something of an obsession in Victorian Britain, and crop up in various pseudo-anthropological books of the time and in novels such as R L Stevenson's *Jekyll and Hyde* (1886), H G Wells' *Time Machine* (1895), and Bram Stoker's *Dracula* (1897), all published just a few years before *Heart of Darkness*. (See Griffith 1995, Breuer 1999.)

12. COLLOCATIONS: WORDS AND WORDS

By now, this lecture is itself turning into a mystery story, and you may be wondering if it is going to be a shaggy dog story. I am travelling cautiously around the coast, but still hesitating at the mouth of the big river, and seem no nearer the heart of the matter. It is time to plunge into the stylistic jungle, armed only with software which can identify objective textual features which critics have at best only half-consciously noticed.

The next stage in the argument is to look not just at individual words, but at recurrent phrases. Individual words can never be more than a starting point, since it is often collocations which create connotations. Or as Sinclair (1975: 85) put it, they "create idioms within texts, by repetitions". Sometimes these connotations are idiosyncratic to the text. For example, in this book *grass* <freq 18> is usually associated with death, decay and desolation: it sprouts through the stones in the *city of the dead* (= Brussels, p.14), and through the bones of a dead man (p.13), old machinery is abandoned in it (p.22), a dead body is carried away by the river like *a wisp of grass* (p.73). Here are just some examples (emphasis added):

```
ast to meet my predecessor, the grass growing through his ribs was
enetian blinds, a dead silence, grass sprouting between the stones,
 upon a boiler wallowing in the grass, then found a path leading up
er the empty land, through long grass, through burnt grass, through
ically childish in the ruins of grass walls. Day after day, with th
layer of silver - over the rank grass, over the mud, upon the wall
```

```
as though he had been a wisp of grass, and I saw the body roll over
mit was half buried in the high grass; the large holes in the peake
 all fours - I've got him." The grass was wet with dew. I strode ra
```

The words GLITTER <freq 14>, GLEAM <8>, GLISTEN <3> and GLINT
<2> connote things which are ominous and dangerous: they collocate with *dark*,
sombre, *gloom*, *blood* and *fire*, and the *infernal stream*; people's eyes *glitter*, *glisten*
and *gleam*; arrows *glint* when they are being shot at Marlow (p.65). Kurtz's African
mistress *glitters* and *glints* (pp.87, 110) in the dusk. His Intended *glitters* and
glimmers in the *last gleams of twilight* (p.109), while her illusion of Kurtz *glows*
(p.109) in the dark. Things once even *glitter* in the *gloom*.

I am aware that such observations can be satirized. In *Small World* by David
Lodge, a computational stylistician shows a novelist that his favourite words are
greasy, *grey* and *grime*. He has not previously been aware of this, and develops
writer's block. Well, this is no longer a problem for Conrad. (Incidentally, the word
greasy occurs only twice in *Heart of Darkness*!)

13. PHRASEOLOGY: WORDS AND GRAMMAR

Words occur not only with other words, but in lexico-grammatical patterns, and
some of these patterns are very repetitive. For example, Conrad repeatedly uses
adjective pairs and long strings of adjectives and nouns:

- cruel and absurd; intolerable and appalling; appalling and excessive
- the air was warm, thick, heavy, sluggish (p.48)
- their glance was guileless, profound, confident, and trustful (p.106)
- the coast ... smiling, frowning, inviting, grand, mean, insipid, or savage, and
  always mute (p.19)
- was it superstition, disgust, patience, fear [...]? (p.60)
- joy, fear, sorrow, devotion, valour, rage - who can tell? (p.52)
- whether it meant war, peace or prayer, we could not tell (p.50)

He also repeatedly uses nominal groups which consist of an abstract noun (usually
two abstract nouns) plus an adjective with a negative prefix:

```
the aspect    of an unknown                planet
the darkness  of an impenetrable           night
the extremity of an impotent               despair
the heart     of an impenetrable           darkness
the sea       of    inexorable             time
the shape     of an unrestful and noisy    dream
the stillness of an implacable             force
the test      of an inexorable    physical necessity
```

I'll come back to these abstract noun phrases in a moment.

When F R Leavis (1962) complains about Conrad's repetitive style, he gives as
examples the individual words *inscrutable*, *inconceivable* and *unspeakable*, but
seems to miss the grammatical generalization that Conrad uses a large number of

words with negative prefixes: there are around 200, about two per page on average. The following are the most frequent:

- impossible/ity 12, uneasy/iness 8, unexpected/ness 7, impenetrable 6, inconceivable/ly 6, incredible 5, indistinct/ly 5, intolerable/ly 5, unknown 5, incomprehensible 4, inscrutable 4, unearthly 4, unsound 4

In addition to these 200 or so words, a further 50 or so end in -*less* (e.g. *colourless*, *heartless*), and there are a further 500 or so occurrences of *no*, *not*, *never*, *nothing*, *nobody* and *nowhere*, plus a further 50 occurrences of *without.* The total frequency of all these negatives is over 800: around one every 50 words of running text. This is a higher frequency than in a reference corpus of fictional texts, and higher again than in a corpus of mixed written texts (though the pattern is not quite so consistent and clear as with the frequency of vague words): see Appendix.

14. INTERPRETATION?

So, we have some further quantitative data on style. The difficulty - to which Stanley Fish correctly points - comes when we try to say what this pattern means, but here goes. A negative statement usually implies that a positive was expected, and many examples emphasize how alien Africa is, and how everything contrasts with things back home. If Marlow says the African coast is *featureless* (p.19) and *formless* (p.20), this is because we do not expect coasts to be like this. If he says *there were no villages* (p.58) along the river, then this is because we expect villages along rivers. If the river flows by *without a murmur* (p.38), then this is because we expect a river to make some sound. If he says that *nothing happened* (pp.20, 111), then this is because we expected something to happen. This theme is sometimes explicit: the earth seems *unearthly* (p.51), as opposed to what we are *accustomed to look upon* (p.51); and Marlow comments that he fails to recognize the skulls on the stakes, because he *had expected to see a knob of wood* (p.82). (EXPECT <17>.) Ian Watt (1960: 259) makes the point that "there are no negatives in nature, but only in the human consciousness".

Watt (1960) is here discussing not Conrad, but *The Ambassadors* by Henry James. Tabata (1995: 102) quotes Watt, and uses data from Dickens to argue that negatives are a measure of subjectivity. Werth (1995) argues that the negatives in the opening of E M Forster's *Passage to India* signal unexpected aspects of the scene. As other critics (e.g. Hidalgo-Downing 2000) point out, negatives are one way of implying more than is literally said (p.217). They deny expectations and challenge background assumptions (p.223), they are a way of questioning reality, and are therefore an alienation device (pp.219, 222). Non-referential noun phrases (as in *there were no villages*) are yet another way in which Conrad expresses Marlow's uncertainty and confusion.

These frequent negatives probably contribute to the representation which so irritates Chinua Achebe: that this world is represented as strange, foreign, alien, contrary to cultural expectations, incomprehensible and *impenetrable to human thought* (p.79). These pervasive morphological and grammatical patterns construct a contrast between the supposedly civilized world which Marlow has left and the supposedly primitive world which he encounters. One set of lexis which is

frequently negative is words relating to movement and change (Klooss 2002). Examples include:

- no change appeared; I did not move; not a leaf moved; leaves, boughs, festoons, motionless in the moonlight; heavy and motionless foliage

There are also frequent references to things being *always the same, monotonous, still, uniform* and *changed into stone* (plus SILENT: see above). These aspects of phraseology contribute to the theme of the timeless nature of the jungle and Marlow's journey back *to the beginnings of time* (p.58).

15. MORE ON PHRASEOLOGY: N-GRAMS

We now have some data - partly quantitative - on individual words, collocations and lexico-grammatical patterns. I move now to the final stage in my argument and discuss some quantitative data on phraseology and how the phraseology in this text relates to phraseology in general English. This also introduces a rather different sense of intertextuality: not allusions to other specific texts, but relations between this text and general language use.

One thing we can do here, with computational assistance, is to look at recurrent chains of words: recurrent two-, three-, four-, and five-word sequences. So, here is a final attempt to identify lexico-grammatical patterns which pervade the text. The most frequent two-word chains are of course pairs of grammatical words, such as *of the* <241> and *in the* <211>, which looks hardly promising, though a concordance of the top pair shows that around half of the occurrences are followed by a place term: *of the forest(s)* <9>; *of the land* <9>; *of the river* <7>; *of the earth* <6>; *of the wilderness* <5>; *of the world* <5>; *of the stream* <4>; etc. This is 3 to 4 times higher than in a large corpus of general English.

In addition, the top two-word chain which contains a content word is *seemed to* <46>: it occurs every couple of pages on average. Now the word *seemed* has risen right to the top of a frequency list. Again, we have to be careful to compare such frequencies with broader language use. The chain *seemed to* is more frequent than in a reference corpus of fiction (see Appendix): 1.2 versus 0.4 times per thousand running words. But Johansson (2001) points out that the lemma SEEM is a very common verb: almost one per thousand words of running text. It is only corpora which can provide such statistics and therefore a powerful new way of comparing individual texts and general language use.

Chains in this sense can identify repetitive phrases which relate to the central themes of Marlow's uncertainty about everything, and geographical and psychological space. For example, the complete list of four-word chains which occur more than five times each is

```
it seemed to me          7
as far as I              6
as though I had          6
with an air of           6
the depths of the        6
```

The first four of these phrases concern uncertainty and the fifth is a place expression. We can also identify more abstract phrasal frames. Here are chains which each occur individually more than once, and which differ in only one word from other chains:

```
the bottom   of the
the depths   of the
the edge     of the
the face     of the
the middle   of the
the midst    of the
the recesses of the
the rest     of the    <total freq 26>


as though I    had
as though he   had
as though it   had
as though they had     <total freq 15>


I don't know how
I don't know I
I don't know what
I don't know why       <total freq 10>
```

16. APPEARANCE AND REALITY, CENTRE AND PERIPHERY

I think we have now identified phraseological patterns which contribute to the feeling that the text is very repetitive, and which convey major themes in the text.

One major theme is appearance and reality, and Marlow's uncertainty about things is conveyed not only by words such as "fog and haze" and "dreams and nightmares", and not only by individual vague words such as *something*, but also by these recurrent phraseological units. I am not claiming that corpus methods can always produce entirely new insights into the text. Several literary critics (Senn 1980, Stampfl 1991) have observed that *Heart of Darkness* contains many negatives and many occurrences of *as if* and *as though*. However, as far as I know, the range of phrasal patterns which I discuss has not been previously identified.

Another major theme is places, geographical and psychological, Europe and Africa, here and there, centre and periphery. Kurtz is an *extremist*, who has *stepped over the edge* (p.101). This is also no news to literary critics, but we want to make explicit the whole range of linguistic means which Conrad uses to convey these themes. So, bearing in mind the title of the book and its very last phrase

- the tranquil waterway [= the Thames] ... <u>seemed to</u> lead <u>into the heart of an immense darkness</u>

let us look finally at repeated phrases which can be picked out by a mechanical method, but which still require interpretation.

The book contains a large number of phrases with the structure <u>PREP the NOUN of DET NOUN</u>, where the first noun is a place term. The following are only some examples.

```
beyond  the              bounds    of       permitted aspirations
to      the              centre    of a     continent
into    the gloomy       circle    of some  inferno
in      the uttermost    depths    of       despondency
into    the              depths    of       darkness
in      the              depths    of the   land
on      the              depths    of the   sea
towards the              depths    of the   wilderness
to      the              edge      of the   forest
at      the              edge      of the   forest
to      the uttermost    ends      of the   earth
into    the              heart     of       darkness
into    the              heart     of an    immense darkness
down    the              middle    of the   river
in      the              midst     of the   incomprehensible
from    the              recesses  of the   land
on      the              threshold of       great things
```

(Out of the five-word phrasal frames, there are 172 occurrences of the construction *the * (*) of a/an/the*. For example: *the face of the forest*; *the glitter of the reach*; *the midst of a shocking hullabaloo*; *the uttermost ends of the earth*; *the long shadows of the forest*; etc. Out of the four-word phrasal frames, the top four are: *the * of the* 106; *in the * of* 47; *the * of a* 36; *the * of his* 26. For example: *the tops of the*, *in the train of*, *the slope of a*, *the wastes of his*. There are 168 occurrences of *the NOUN of the/a/his*.)

17. ROUTINE PHRASEOLOGY

In the past, stylisticians have often argued that literary patterns deviate from everyday language use. In this book, phrases such as *in the middle of* and *on the edge of* are frequent, and significant for the themes of heart of darkness, limits of experience, and so on. I assumed initially that they would be much more frequent in the book than in general English. But my intuition was wrong: this is not the case. Although these phrases are important in *Heart of Darkness*, they are not unusually frequent in the book. On the contrary, phrases with this structure are highly frequent in general usage. Some of the most frequent five-word phrases in the 100-million word British National Corpus (with frequencies per million running words) are:

- at the end of the 45; by the end of the 18; in the middle of the 16; at the top of the 11; at the beginning of the 9; at the bottom of the 7; on the edge of the 7; towards the end of 5; in the centre of the 6

This fact itself requires an explanation. One hypothesis is that such phrases provide ways of talking about topics of great social importance - space, geographical, social and psychological, along with views of centre and periphery - and therefore ways of

expressing one of the most important kinds of pragmatic meaning: the speaker's perspective or point of view.

Many of the place expressions in *Heart of Darkness* share another characteristic with general English: they use body parts as place terms. Examples are:

```
the back   of the               house
the bottom of the               river
the bowels of the               land
the face   of the               forest
the foot   of the last          page
the head   of a   long          sandbank
the heart  of an  impenetrable  darkness
the mouth  of the big           river
the seat   of the               government
```

(There are around 50 further examples of *side*: *inside*, *outside*, *hillside*, *river-side*, *land-side*, *water-side*, etc.) This is a very widespread - probably universal - preferred metaphor (which is documented in detail in Traugott & Heine eds 1993), and it is therefore hardly surprising that the pattern also occurs in *Heart of Darkness*.

However, although the phrasal frames are frequent in general English, Conrad's place expressions do differ in several ways from their frequent use in general English:

(1) First, most of these place expressions in *Heart of Darkness* are abstract and extremely vague. That is, the final nouns in these phrasal frames do not fit the pattern in general English: they are very general (e.g. *continent*, *earth*), or mythical (e.g. *inferno*), or abstract (e.g. *darkness*), or adjectival (e.g. *the incomprehensible*). I pointed out above that real places are hardly named at all in the book. Apart from the Thames, the frequent place expressions which do occur could not be found on any map. They refer to an unnamed *forest* or *river*, to a mythical *inferno*, to an abstract *wilderness*, or to an even more abstract *darkness*.

(2) Second, the words acquire evaluative and metaphorical connotations in uses such as

- I [= Marlow] had peeped over the edge (p.101)
- he [= Kurtz] had stepped over the edge (p.101)

General corpora show that *over the edge* has clear connotations of danger. Frequent collocates are *hanging*, *dangling*, *falling*, *slipping*, *tipping*, and *dangerously*.

(3) Finally, this aspect of the recurrent phraseology is read differently because of the context, especially the title of the book, the repeated references to characters who are shades, shadows and phantoms, on the threshold between this world and the next, and the last sentence which contains two such phrases:

- to the uttermost ends of the earth
- into the heart of an immense darkness

18. OVER THE EDGE ...

Here is where I go over the edge, and provide some concluding interpretations. A published book is a public object. It is an observable selection from the potential of the language. John Sinclair's corpus work shows how patterns of co-occurrence provide observable evidence of meaning, and the hermeneutic methods he has developed can be applied both to individual texts and also to intertextual relations between texts and general language use. These methods can show overall textual patterns, particularly repeated lexico-grammatical patterns. I have given examples of several different patterns of vocabulary, phraseology and grammar: antonymous lexical contrasts, negatives (another type of contrast), lexical and grammatical expressions of uncertainty and vagueness, and highly abstract place expressions. These patterns occur throughout and all contribute to the themes of uncertainty, fallible and distorted knowledge, and faulty expectations.

So, here is where I make my final jump from description to interpretation, from forms to meanings. One of the unsolved problems of text analysis is how a close attention to the text can be reconciled with an understanding of its cultural and historical background. A literary text is not autonomous and self-contained. There are no clear boundaries between a literary text and general language use, since all texts consist of fragments of other texts. They allude to text-types (such as adventure stories and black comedy), to other stories (such as Greek myths and the Faust legend), and to specific texts (such as the Bible).

But comparative corpus methods also allow us to study a different kind of intertextuality: how far texts consist of recurrent phrasal patterns which are widespread in the language as a whole. In some ways, the language of *Heart of Darkness* deviates from the norm of everyday language use. But there are also many recurrent phrases in the book which are significant, not because they deviate from general language use, but because they use the routine phraseology of the language. An overall discourse schema in the book is "us and them", and this is conveyed by recurrent phrasal schemas which position readers - depending on who they are - in the centre, on the brink, over the edge, or beyond the pale.

Is that why the book is so popular? Not only does it fit into widely popular text-types (such as boys' own adventure story), not only does it contain repeated images from folk tales (such as getting lost in the big forest). It also uses a recurrent phraseology which is highly frequent in English as a whole, more frequent than anything else, and which reflects social obsessions of us and them, centre and periphery. Perhaps that is what is significant about the phrases. They reveal some of the metaphors we live by, to use the expression proposed by Lakoff and Johnson (1980).

Now I will point out a logical problem with my argument, for it might seem that I am here impaled on the Fish Fork. If I had discovered that this phraseology was more frequent in *Heart of Darkness* than in general English, then I would have argued that it is interesting for this reason. But since I have discovered that it is a pervasive feature of general English, then I have argued that it is interesting for that reason ... However, as I mentioned earlier, I think there are very simple responses to the Fish Fork.

First, I have made an empirical discovery about the phraseology of the text, which contradicted my intuitions. The phraseology which I have described is a formal, observable, objective feature of the book. It is only one feature, and it is

open to different interpretations, but it was not created by my analysis. If you look, you will find these patterns in the text and in the wider norms of language use. Second, the Fish Fork applies to any study of anything. Pure induction will never get you from empirical observations to interesting generalizations. You have to know where to look for interesting things. But the aim is to say systematically and explicitly what these things are: and that is where empirical, observational analysis can contribute. In the meantime ...

19. AT THE VERY END OF THE LECTURE ...

We are now at the very end of the lecture, and, like other texts, lectures often have conventional, formally marked, beginnings and ends. Conrad began *Heart of Darkness* as follows:

> "The *Nellie*, a cruising yawl, swung to her anchor without a flutter of the sails, and was at rest. ..."

Here is how Conrad might have ended this lecture:

> ... The lecture swung to its anchor, with a final flutter of its data, and was at rest. The lecturer, brooding gloomily over an extravagant problem, seems to have presented one of those inconclusive arguments, yet based on the undeniable logic and the implacable force of a computational method. It all seems somehow to throw a kind of light, and the fog seems to lift, but only to leave a bright haze gleaming like a misty halo over an insoluble enigma, inexplicable, and altogether bewildering. There is something not quite right, as if the argument is still in the making. The inscrutable logic seems to lead to the threshold of great things, to the uttermost ends of stylistics, and yet into a place of cruel and absurd mysteries, not fit for an undergraduate student to behold. The software has been sent into the depths of the text, it has raided the corpus, and in return has come a precious trickle of data. Yet, to speak plainly, it is uncertain whether the argument has stepped over the edge. The audience ask themselves whether the method is sound, and what it all means, whether the analysis has led into the heart of an immense intellectual darkness, or whether there is a glimmer of light at the end of the tunnel.

ACKNOWLEDGEMENTS

REFERENCES

Achebe, C (1975) An image of Africa: racism in Conrad's *Heart of Darkness*. Revised version in R Kimbrough ed 1988: 251-62.

Breuer, H (1999) Atavismus bei Joseph Conrad, Bram Stoker und Eugene O'Neill. *Anglia*, 117, 3: 368-94.

Fellbaum, C (1995) Co-occurrence and antonymy. *International Journal of Lexicography*, 8, 4: 281-303.

Fish, S E (1973) What is stylistics and why are they saying such terrible things about it? In Weber ed 1996: 94-116.

Fletcher, W H (2002) N-gram software. Available http://kwicfinder.com/kfNgram/.

Foucault, M (1961) *Folie et déraison: Histoire de la folie à l'âge classique*. Paris: Plon. (English translation of abridged edition: *Madness and Civilisation*. London: Tavistock, 1967.)

Griffith, J W (1964) *Joseph Conrad and the Anthropological Dilemma*. Oxford: Clarendon.

Halliday, M A K (1971) Linguistic function and literary style. In Weber ed 1996: 56-86.

Hardy, D & Durian, D (2000) The stylistics of syntactic complements: grammar and seeing in Flannery O'Connor's fiction. *Style*, 34: 92-116.

Hidalgo-Downing, L (2000) Negation in discourse: a text-world approach to Joseph Heller's *Catch 22*. *Language and Literature*, 9, 3: 215-39.

Hopper, P & Traugott, E C (1993) *Grammaticalization*. Cambridge: CUP.

Jakobson, R (1960) Closing statement: linguistics and poetics, In Weber ed 1996: 10-35.

Johansson, S (2001) The English verb *seem* and its correspondences in Norwegian. In K Aijmer ed *A Wealth of English*. Göteborg: Acta Univertsitatis Gothoburgensis. 221-45.

Jones, S (2002) *Antonymy*. London: Routledge.

Justeson, J S & Katz, S M (1991) Redefining antonymy: the textual structure of a semantic relation. In *Using Corpora*, Proceedings of 7th Annual Conference of the UW Centre for the New OED and Text Research, Oxford.

Kimbrough, R ed (1988) *Joseph Conrad: Heart of Darkness*. Critical edition, 3rd ed. NY: Norton.

Klooss, W (2002) Through the eyes of colonialism: Joseph Conrad's *Heart of Darkness* and its metaphors. University of Trier. Unpublished.

Lakoff, G & Johnson, M (1980) *Metaphors We Live By*. Chicago: Chicago UP.

Leavis, F R (1962) *The Great Tradition*. London: Chatto & Windus.

Lothe, J (2000) *Narrative in Fiction and Film*. Oxford: Oxford University Press.

Lothe, J (2001) Cumulative intertextuality in *Heart of Darkness*. In G Fincham & A de Lange eds *Conrad at the Millenium*. New York: Columbia University Press.

Phillips, C (2003) Out of Africa. (Interview with Chinua Achebe). *The Guardian*, 22 February 2003.

Reeves, C E (1985) A voice of unrest: Conrad's rhetoric of the unspeakable. *Texas Studies in Literature and Language*, 27: 284-310.

Scott, M (1997) *WordSmith Tools*. [Software.] Oxford: OUP.

Senn, W (1990) *Conrad's Narrative Voice*. Bern: Francke.

Sinclair, J M (1965) When is a poem like a sunset? *A Review of English Literature*, 6, 2: 76-91.

Sinclair, J M (1975) The linguistic basis of style. In H Ringbom ed *Style and Text*. Stockholm. 75-89.

Sinclair, J (1991) *Corpus, Concordance, Collocation*. Oxford: OUP.

Stampfl, B (1991) Marlow's rhetoric of (self-)deception in *Heart of Darkness*. *Modern Fiction Studies*, 37: 183-96.

Stubbs, M & Barth, I (in press 2004) Using recurrent phrases as text-type discriminators. *Functions of Language*, 10, 1: 65-108.

Tabata, T (1995) Narrative style and the frequencies of very common words. *English Corpus Linguistics*, 2: 91-109.

Toolan, M (1990) Stylistics and its discontents: getting off the Fish hook. In Weber ed 1996: 117-35.

Toolan, M (2002) On the centrality of stylistics. *The European English Messenger*, xi, 1: 19-25.

Traugott, E C & Heine; B eds (1991) *Approaches to Grammaticalization*. 2 vols. Amsterdam: Benjamins.

Watt, I (1960) The first paragraph of *The Ambassadors*: an explication. *Essays in Criticism*, 10: 250-74.

Watt, I (1979) Impressionism and symbolism in *Heart of Darkness*. In Kimbrough ed 1988: 311-36.

Weber, J J ed (1996) *The Stylistics Reader*. London: Arnold.

Werth, P (1995) World enough and time. In P Verdonk & J J Weber eds *Twentieth Century Fiction*. London: Routledge. 181-205.

Widdowson, H G (1975) *Stylistics and the Teaching of Literature*. London: Longman.

APPENDIX

I have said in the body of the article that vague words and negatives are more frequent in *Heart of Darkness* than in fictional texts or written texts in general. The following table provides comparisons. The figures give the numbers of different words and phrases in

(1) HEART = *Heart of Darkness*.
(2) FICTION = a corpus of fictional texts of over 710,000 words.
   The reference corpus here was the same as that used above for identifying "keywords": the "imaginative prose" categories K, L, M and N in Brown, LOB, Frown and FLOB.
(3) WRITTEN = the one million word written component of the BNC sampler.

Frequencies are given in occurrences per 1,000 running words.

	(1) HEART	(2) FICTION	(3) WRITTEN
seemed to	1.2	0.4	0.09
some	2.6	1.5	1.5
something	1.3	1.0	0.4
somebody	0.2	0.1	0.05
sometimes	0.6	0.2	0.2

```
somewhere      0.2          0.2          0.03
somehow        0.2          0.1          0.04

not            6.3          4.7          4.2
n't            4.2          6.3          1.7
no             4.1          3.2          1.9
never          1.0          1.2          0.5
nowhere        0.2          0.04         0.02
nothing        1.2          0.7          0.3
without        1.3          0.6          0.4
```

With two exceptions (*n't*, *never*), there are always more occurrences in HEART than in FICTION, and more in FICTION than in WRITTEN. Taken together, the frequency of *not*, *n't*, *no* and *never* are almost identical in HEART and FICTION, and still much higher than in WRITTEN.