Lexical Cohesion and Corpus Linguistics
Lexical Cohesion and Corpus Linguistics
Benjamins Current Topics

Special issues of established journals tend to circulate within the orbit of the subscribers of those journals. For the Benjamins Current Topics series a number of special issues have been selected containing salient topics of research with the aim to widen the readership and to give this interesting material an additional lease of life in book format.

Volume 17
Lexical Cohesion and Corpus Linguistics
Edited by John Flowerdew and Michaela Mahlberg
These materials were previously published in the International Journal of Corpus Linguistics 11:3 (2006)
Lexical Cohesion and Corpus Linguistics

Edited by
John Flowerdew
University of Leeds

Michaela Mahlberg
University of Liverpool

John Benjamins Publishing Company
Amsterdam / Philadelphia
Table of contents

Introduction 1

Lexical cohesion and rhetorical structure
  John Morley 5

Lexical bundles and discourse signalling in academic lectures
  Hilary Nesi and Helen Basturkmen 23

Cohesive chains and speakers’ choice of prominence
  Martin Warren 45

Describing the extended meanings of lexical cohesion in a corpus of SARS spoken discourse
  Winnie Cheng 65

Use of signalling nouns in a learner corpus
  John Flowerdew 85

Lexical cohesion: Corpus linguistic theory and its application in English language teaching
  Michaela Mahlberg 103

Index 123
Introduction

Lexical cohesion is about meaning in text. It concerns the way in which lexical items relate to each other and to other cohesive devices so that textual continuity is created. The seminal work on lexical cohesion is Halliday and Hasan’s (1976) *Cohesion in English*, where it is nevertheless given the shortest treatment of the five types of cohesion identified by the authors. According to Halliday and Hasan (1976), lexical cohesion concerns two distinct but related aspects: reiteration and collocation. Reiteration is “the repetition of a lexical item, or the occurrence of a synonym of some kind, in the context of reference; that is, where the two occurrences have the same referent” (Halliday & Hasan 1976:318–9), while collocation is the use of “a word that is in some way associated with another word in the preceding text, because it is a direct repetition of it, or is in some sense synonymous with it, or tends to occur in the same lexical environment” (Halliday & Hasan 1976:319). Collocations may include any words that are in some sort of semantic relationship, although Halliday and Hasan (1976) draw special attention to superordinates, hyponyms and antonyms. It is to be noted that the conception of collocation, which is exemplified as operating primarily across clauses, is different from the current understanding of the term in corpus linguistics, as is illustrated in some of the contributions to the present issue.

Halliday and Hasan’s model of lexical cohesion was developed further by Hasan, in part of her contribution to Halliday and Hasan (1985) (Chapter 5), where she developed the notions of repetition, synonymy, hyponymy and meronymy. Continuing within this tradition, Martin (1992) developed a framework for a more detailed account of lexical relations, including Hasan’s categories. In addition collocation “was factored out into various kinds of ‘nuclear’ relation” (Martin 2001:38), which consisted of elaboration, extension, and enhancement (as developed by Halliday 1994) for the clause complex. Martin used the term ‘ideation’ to refer to lexical relations “as they are deployed to construe institutional activity.” (Martin 2001:38).
The work referred to thus far is closely related to studies of lexical relations in discourse by e.g. Winter (1977) (focusing on the anaphoric function of lexis, referred to as ‘type 3’), Francis (1986) (focusing on what she called ‘anaphoric nouns’), Tadros (focusing on the cataphoric function of cohesion, including the lexical variety) and Hoey (1991) (who relates lexical patterning to how lexical cohesion operates over larger stretches of text). In addition, we can see relations to the development of an ideational semantics by Halliday and Matthiessen (1999).

The above references (which are far from exhaustive) still emphasise the inter-clausal nature of lexical cohesion and do not yet draw on corpus linguistic methodologies or descriptive tools that have only emerged fairly recently. With the development of corpus linguistic techniques, research into lexical cohesion may take new routes. Corpora provide huge amounts of real evidence and at the same time make it possible to focus on specific types of texts and on specific patterns of words. Thus they can also provide new perspectives on links between parts of texts. Corpus linguistic work that raises theoretical issues (e.g. Sinclair 2004; Hoey 2005) suggests that lexis needs more attention than it has received in traditional approaches to language. Consequently, lexical cohesion also needs more attention — or even an approach that is fundamentally different from the traditional textlinguistic one.

There are not yet many studies that apply or extend corpus linguistic concepts to a realm that has traditionally been occupied by textlinguistics. The present volume brings together different contributions that aim to have a closer look at lexical cohesion in a corpus linguistic context. However, innovative approaches to lexical cohesion do not only play a role in corpus linguistics, but also have implications for language teaching and the way in which cohesion is dealt with in the classroom. Some of these implications are addressed in the present volume.

The papers in this collection were first published in the International Journal of Corpus Linguistics. In editing that special issue we were supported by a group of anonymous reviewers who provided useful comments and advice. We would like to thank both the reviewers and the authors for taking part in this interesting project. The papers were originally presented at a colloquium at the 2004 Teaching and Language Corpora conference (TALC6) in Granada, Spain.

John Flowerdew
Michaela Mahlberg
September 2008
References


Lexical cohesion and rhetorical structure

John Morley
University of Siena

Lexical cohesion not only contributes to the texture of a text, it can help to indicate the rhetorical development of the discourse. This article looks at this argument-structuring function of lexical cohesion first by considering single texts using the techniques of classical discourse analysis and then by using the methodology of corpus linguistics to examine several million words of text. First, the nature of cohesive links within single articles is examined. Next, the link between headlines and the articles that follow them is studied. Finally, various concessive mechanisms which structure arguments are examined in detail. It is argued that an awareness of the mechanisms outlined in this article will help students to understand better the kind of argumentation presented in texts. All the texts studied are from English newspapers.

Keywords: lexical cohesion, discourse analysis, corpus linguistics, rhetoric, newspaper discourse, ESP

1. Introduction

This paper examines how different kinds of lexical cohesion can contribute to the rhetorical structure of a text. First of all, we look at the classic approach to lexical cohesion (Halliday & Hasan 1976, 1995), which identifies semantic fields and the logical relations between words within these fields and argue that even this approach can indicate the rhetorical movement of the discourse. Secondly, we will consider the kind of lexical cohesion which links newspaper headlines to the articles which follow them. In this case, the headline not only prepares us for the lexis of the article but sometimes may also indicate the kind of grammatical structures which will be used to carry forward the argument. These two sections use close analysis of individual texts and are examples of a qualitative approach to the analysis of discourse. At this point (Section 3), the methodology of corpus linguistics is used to examine some consequences of the argument-structuring use of lexical cohesion which are not immediately apparent to the naked eye.
Finally, we will look at the mechanism which I have called the Aunt Sally. Aunt Sallies are lexical items of various kinds which trigger off counter arguments later in the text. These devices are frequently used in newspaper discourse to structure the writer’s argument.

The corpus used for the second part of this article (from Section 3 onwards) is called Papers93. It contains roughly one hundred million words taken from the Guardian, Telegraph, Sunday Telegraph, The Times and the Sunday Times in 1993, which constitutes the entire verbal production of those newspapers in that period. This corpus was preferred to a corpus of general English, such as the British National Corpus or the Bank of English because the writer is particularly interested in newspaper language. A consequence of this choice is that the conclusions that are reached in this article apply only to the kinds of discourse found in newspapers. Occasionally, only parts of the corpus (Papers93) have been used. This was because it was necessary to examine much wider contexts than those given by the concordance, which is a very time-consuming process.

2. Classical lexical cohesion

In Chapter Five of Language, Context and Text (Halliday & Hasan 1995), entitled The Texture of a Text, Hasan says that lexical cohesion is based on a co-extensional tie between linguistic items. In other words, items which create lexical cohesion belong to the same semantic field. The expression ‘semantic field’ is a metaphor, as is the philosophical term ‘extension’, which is defined by the SOED as, “The range of a term or concept as measured by the number of kinds of object which it denotes or contains”. For instance, the words prison, trial and investigation, taken from the first paragraph of the newspaper text below (The prison door), all come from the semantic field of ‘law’. We can say that these words help to tie various components of the text together. Hasan, in her part of Halliday and Hasan (1995:81), describes such devices as lexical componential devices.

In The prison door article we find two major semantic fields, that of ‘law’, mentioned above, and that of ‘politics’: there are in the first paragraph expressions like prime minister, leader of the opposition and the word politics itself. Of the 553 words (tokens) in the article, 47 are from the semantic field of ‘law’ and 56 words from the semantic field of ‘politics’.

These two semantic fields do not simply create cohesive links, they are also exploited to rhetorical effect: their relative densities indicate the flow of the argument or its discourse structure. There is, for instance, a greater concentration of legal vocabulary at the beginning of the article — there are 12 items from the
semantic field of ‘law’ in the first paragraph — and on political terminology at the end — the last paragraph contains nine words from the semantic field of ‘politics’ and none from that of ‘law’. The article progresses in this way: Mr Berlusconi is identified in political terms as former prime minister of Italy and the present leader of the opposition in the first sentence of the article; his legal problems are then explained in great detail; after this, the article works towards an analysis of the political implications of the legal accusations against him, which reaches its climax in the last paragraph. Below I give the first two and the final paragraph of the article, which illustrate what I have said above. The words in bold are from the semantic field of ‘law’, while words from the semantic field of ‘politics’ are in small capitals.

<Headline>The prison door</Headline>

<First Paragraph>Silvio Berlusconi, media mogul, former prime minister of Italy and now leader of the opposition, has been sentenced to two years and nine months in prison by a Milan court. The trial, and the investigation that led to it, lasted two-and-a-half years. He has been convicted of bribing tax inspectors to soften their examination of his business activities before he went into politics. He has not denied the payments, but says the money was extorted. Mr Berlusconi has been sentenced to prison before (last December, about a film deal), but that sentence was suspended. </First Paragraph>

<Paragraph2>Will he go to jail this time? Probably not. First he will appeal, and, under Italian law, he does not have to begin serving the sentence during the appeal process. If the sentence is upheld, he can still avoid serving prison time by demanding a suspension of the sentence (normally allowed for prison terms of under three years). Better still, he is also a member of parliament, which has to waive his immunity. It is most unlikely to do so, since Mr Berlusconi is not only leader of the opposition but also founder-president of a party, Forza Italia (Go, Italy), that in the latest general election collected 8m votes. To allow Mr Berlusconi to go to jail would cause a constitutional crisis that nobody, including the prime minister, Romano Prodi, really wants. </Paragraph2>

<Last Paragraph>While the Berlusconi affair drags on, Italian politics is stuck. Badly needed constitutional reform, since it requires broad support, has proved impossible. The parliamentary opposition is ineffectual. This may suit the government, and it may not hurt the opposition all that much, since it has an excuse—and a possible martyr—in Mr Berlusconi. But it is bad for Italy. </Last Paragraph>

(The Economist)
3. **Newspaper headlines**

The next question to consider is the kind of cohesion that links headlines to the articles which follow them. This cohesion is sometimes of the classical lexical kind: a word in the head indicates the semantic field of some important number of the words which follow.\(^3\) Sometimes, however, a lexical item in the head will presage the use of expressions in the body of the article which carry out the rhetorical function described by that item: for example, the word *choice* in a headline will be followed by examples of choosing in the following text.

Writers on the print media tell us that one of the main functions of the headline is to summarize the contents of the article which it introduces, although, as Morley (1998:26) notes, the tabloids often prefer headlines which dramatize rather than give a simple summary. In such cases there is normally a summary in a sub-head. White (1997:129) sees the headline’s function in more ideological terms: “[..] the Headline/lead does not so much summarize the action or set of statements at issue as provide a particular interpretation of their significance for the social order.” According to his theory, the headline and the lead paragraph form a nucleus of which the other paragraphs are satellites, which restate, expand, explain etc. the information contained in the nucleus. From the point of view of the present argument the difference is of little importance: the head tells us what is central to the story or, in other words, what the story is about. By telling us what the story is about the headline creates expectations about the contents of the article, its style, lexis and point of view, among other things.

Through creating expectations about the lexis that will be used in the body of the article, the headline is a powerful producer of cohesion in newspaper discourse. The headline of the article *The prison door* prepares us to find a concatenation of references to the law. There is no indication that the article will also deal with politics, until we find the name *Silvio Berlusconi* as the first words of the lead paragraph. However, headlines and first paragraphs in newspapers are considered as a unit both in Bell’s model (1991), where they are referred to as the Abstract, and in White’s model (1997), where they are described as constituting the Nucleus.

An even clearer example of lexical cohesion between the headline and the first paragraph of the article is apparent in the *Guardian* (26 February 2004). The crucial words have been put in bold.

<Headline>*Choice moment*</Headline>
<Sub-head>*This is not the time for war*</Sub-head>
<First Paragraph>*The choice is now clear. It is a choice for governments but also for individuals. It is a matter of judgment and of conviction. It is a decision with momentous ramifications for the future international order. It will shape the world*
we share. It is a choice many would prefer to avoid. But after publication of the US-British draft UN resolution and the Franco-German memorandum on Iraq, it is a choice that cannot any longer be deferred. </First Paragraph>

Here we find the word choice repeated four times in 77 words, its synonym decision, the two words judgment and conviction, which refer to elements in the making of a choice and therefore have a meronymical connection with choice. This is classical lexical cohesion, as defined by Halliday and Hasan (1976, 1995).

If we look at a very similar headline and sub-head, again from the Guardian (14 February 2003), with the first paragraph which follows it, we find a rather different kind of cohesion in play. Again, the crucial words have been put in bold.

<Headline>A moment for truth</Headline>
<Sub-head>Blair and Britain face historic choices</Sub-head>
<First Paragraph> In itself, the fact that a majority of the country either opposes or is sceptical towards his policy on Iraq is not necessarily a reason for Tony Blair to change course. If a particular policy is right, then a minister is entitled to stick with it in the face of opposition, to trust to his own judgment and to take the consequences. That is how Mr Blair seems determined to see things on Iraq, fighting a lonely battle to persuade a doubting nation that George Bush has the British people's best interests at heart. But Mr Blair's view of the crisis — and of the role he is playing — remains dangerously double-edged. Mr Blair's whole approach on Iraq may be put to its ultimate test very soon. If the US decides to push ahead with a precipitate attack on Iraq in defiance both of good judgment and the proper processes of the UN, Mr Blair will face a fundamental choice. When he does, it is vital — absolutely vital — that he is compelled by all possible means to make the right one. This weekend will be a crucial opportunity — perhaps the last one — to try to save Mr Blair, and more importantly the country, from the error of supporting a misjudged US approach towards the Iraqi regime. </First Paragraph>

Once again the key word is choice, this time in the plural: the sub-head, Blair and Britain face historic choices, indicates that the article will be presenting us with a number of choices. What we find surprisingly little of, however, is lexis directly associated with the semantic field of choice. Rather, we are presented with a series of lexical and grammatical items that suggest or even embody the act of choosing. The rhetoric of the paragraph, and indeed of the article as a whole, is one of hedged statements, though in the end the reader will be left with no doubt about the side of the fence on which the Guardian stands: see the last paragraph quoted in Section 4. To indicate only the most obvious and superficial linguistic choices, we are offered in the first sentence the disjunction either/or, then there are two conditional clauses, the choice of a quintessentially dubitative verb (seems), the archetypical conjunction to introduce an opposite point of view (but), a fence-sitting adjective (double-edged), the principal modal verb of doubt (may), the repetition
of the word *choice* — the only example of lexical cohesion in the classical sense — and its co-classificatory pronoun phrase *the right one* and, to round off in hesitation, the adverb *perhaps*. Here we can see how lexis and grammar work together to create cohesion in a text.

4. Lexical cohesion, corpus linguistics and discourse structure

So far I have used examples of lexical cohesion which are discoverable through a traditional analysis of the text. Sometimes, however, the cohesive ties between the headline and the rest of the discourse may produce meaning not easily discernible on the surface of the discourse. It is in these cases that we need to use the methodology of corpus linguistics. Consider again the second headline and sub-head from the *Guardian* (14 February 2003) and, this time, the final paragraph of the article. The words I will discuss have been put in bold.

<Headline>*A moment for truth*</Headline>  
<Sub-head>*Blair and Britain face historic choices*</Sub-head>  
[…]

<Last Paragraph>*Today, when Hans Blix makes his latest report to the UN security council in New York, we will be subjected to instant propaganda pressure by Jack Straw and others to persuade us that we must jump into line behind the Bush administration response. Will we never learn? This has repeatedly been the wrong approach in the past, and it is essential not to be bounced into it again. British ministers, like the British people, must study and digest what Mr Blix says and then must place it in the bigger context of all the challenges that face us. This is a *moment for cool reason* and utmost seriousness. It is a *moment for citizens to choose, for MPs to vote, and for every minister, in public or in private, to weigh what Washington is asking in the balance of our domestic and international goals. This is one of those rare moments when threats of resignation, and even resignations, may be responsible options. Great leadership is not the same thing as opposing what your people want. Great leadership is about rising to the occasion when the country's deepest interests must be given clear voice and true direction.*</Last Paragraph>  

At first sight there seems to be nothing particularly remarkable about the lexical cohesion in this paragraph. Repetition of the phraseology *moment for*, which occurs in the headline is lexical cohesion of the most obvious kind: repetition is the easiest type of lexical cohesion to identify. What is not, perhaps, obvious to the naked eye is that the expression *moment for* has its own powerful semantic prosody. By semantic prosody I mean the *evaluative* meaning — the good or bad 'aura' associated with the lexical item — which is not exhausted within the orthographic item but spread over the surrounding co-text. Frequently the reader is not consciously

In c. 27 million words from the Guardian (Papers93) there are 84 instances of the two-word-cluster moment for. The words that appear immediately to the left of the cluster are illuminating. Leaving aside for now the indefinite and definite articles in the –L1 position (5 and 18 instances respectively) and other grammatical words (one instance each of another, his, that, this), we find a collection of adjectives which indicate that something important of either a negative or positive nature is happening at the moment mentioned: defining (8), wrong (3), critical (2), good (2), historic (2), historical (2), important (2), last (2), sweet (2), timely (2), worse (2), appropriate, big, bleak, bold, delicate, desperate, difficult, eerie, embarrassing, final, fitting, ideal, marvellous, nice, opportune, painful, perilous, poignant, proud, remarkable, salutary, special, traumatic, worst. The only other word immediately to the left of the cluster is political (2), where the context makes it clear that this word too has connotations of importance: an issue of immense environmental, economic and political moment for this country.

To see the kind of company moment for keeps when it is preceded by an article or another grammatical word, we have to look at the co-text which follows the cluster. Of the 27 examples of this type, three can be ignored because the preposition for depends on a word other than moment: There is no fashion at the moment for self-regulation — clearly for here is part of fashion for. All of the remaining 24 examples are followed by words which indicate that an important moment is being referred to: Now is the moment for the rest of the world to make it absolutely clear they support the continuation of the reform process.

It seems evident that the string moment for is primed, in Hoey’s (2004, 2005) sense, to indicate moments which are important either in a positive or a negative way, we might almost refer to them as ‘momentous moments’. It is rather unusual that the priming holds in all instances or, at least, in all the instances in the fairly large corpus which was interrogated. Hoey (2005) is insistent that priming is a tendency. For instance although the expression in winter occurs with the present tense ‘only’ 59% of the time in Hoey’s corpus, he still argues that the words are primed to occur with that tense (Hoey 2005:39).

It might be interesting to compare the behaviour of moment for with that of moment to. If we eliminate expressions like she had no time at the moment to run for office, there are 92 instances of moment to in the same sub-corpus of c. 27 million words. The adjectives immediately to the left of the string are: good (7), last (7), right (7), best (3), ideal (3), opportune (3), better (2), appropriate, apt, benighted, crucial, fitting, great, inopportune, interesting, media, perfect, possible, sensitive, tiny, uncanny,
All these adjectives except *media*, *tiny*, *uncanny* and perhaps *benighted* and *sensitive* express appropriateness/inappropriateness rather than momentousness.

Looking at the right-hand co-text when *moment to* is immediately preceded by an article (indefinite, 11 instances; definite, 27 instances) and other grammatical words — *his* (2), *this* (5) — we find: *moment to/ savour* (4)/ *ask* (2)/ *bring* (2)/ *start* (2)/ *strike* (2)/ *add/ apply/ ban* (2)/ *be learning/ boast/ caress/ cash/ celebrate/ deepen/ ease/ embrace/ end/ force/ give/ impress/ introduce/ launch/ leave / lift/ make/ mention/ move/ name/ press/ realise/ remind / spill/ take stock/ tell/ *Thai die/ test/ waste/ wind*. The implication that the moment is important seems to be missing here, except perhaps in the cluster *a moment to savour*.

*Moment* is, of course, a polysemous word: as well as the “brief interval of time” meaning, we have also that found in Hamlet’s soliloquy “enterprises of great pitch and moment” (*Cambridge New Shakespeare*, Act 3 i, 86 (Wilson 1969)), which has the sense of “importance in influence or effect” (*New Penguin English Dictionary*). *Moment for* obviously descends from this second meaning and perhaps the word always has the possibility to slip between the two meanings.

In the *Guardian* article under consideration, the main head is *A moment for truth*. The importance of this moment is immediately indicated in the sub-head: *Blair and Britain face historic choices*. The expression is picked up again in the peroration in the final paragraph, where we find two overt and two elided instances of *moment for*: *This is a moment for cool reason and utmost seriousness* — the importance of the moment is underlined by the adjective *utmost*; there follows *It is a moment for citizens to choose, [it is a moment] for MPs to vote, and [it is a moment] for every minister, in public or in private, to weigh what Washington is asking in the balance of our domestic and international goals* — here, too, the moments are important ones and we may note the repetition of the key word *choose*; next the syntax changes somewhat, presumably for stylistic variation, while still preserving the word *moment*, with the introduction of a clause dependent on *moments when: This is one of those rare moments when threats of resignation, and even resignations, may be responsible options*. It seems clear that the author is using the expression to emphasize the momentousness of the moment.

That *moment for* is acting as a lexical cohesive device is fairly obvious: we can plainly see that the words *moment for* are repeated; but to appreciate the full force of the message — that this moment is momentous — we need the techniques of corpus analysis.

To return to the idea of lexical cohesion being important for the structuring of the text, the phraseology *moment for* is predicted by the headline and then picked up again in the final paragraph, which represents the peroration, or point of maximum persuasion, of the editorial. The ‘momentous moments’ are produced when the leader writer most wants to persuade.
5. Aunt Sallies and lexical cohesion

5.1 Definitions

The *New Penguin English Dictionary* defines an Aunt Sally in the following way.

*Aunt Sally* /'sali/ noun 1 an effigy of a woman at which objects are thrown at a fair.  
2 an easy target of criticism or attack.

The meaning which interests us in this paper is the second one: “an easy target of criticism or attack”. The concept is very similar to that conveyed by another equally quaint expression: the Straw man or Man of straw. The *New Penguin English Dictionary* defines the expression in this way:

*Man of straw* noun […] 3 dated a sham argument or adversary, set up only to be easily countered.

The expressions may be rather quaint and dated even, but they do represent a rhetorical device which is much used, especially in newspaper discourse. The previous sentence, in fact, is one kind of Aunt Sally. I have set up the idea that the words are ‘quaint’ and ‘dated’ — and therefore not very serious, but then I knock the argument down by pointing out how useful they are because they describe a common
phenomenon in newspaper discourse. Not all Aunt Sallies have this precise logical structure, but they are all in some way similar. Hunston (2001), for instance, discusses the functioning of may not be a. She gives the example: *There may not be a woman in the pulpit yet in [place name], but there is a preacher of sorts* and then comments: “In most cases, the noun group beginning with a indicates something that is highly valued, while the counter-assertion (the contrasting clause) indicates something that is less highly valued, though perhaps more realistically achievable.” (Hunston 2001:25) Schematically: **may not be a — Perfection, but, Good but not perfect.** My sentence has a different rhetorical structure, but is clearly the same kind of mechanism.

This links with what Thompson and Zhou (2000) say of the general phenomenon of concession:

[B]oth propositions are presented as valid, but the second is presented as in some way more valid than the first. This arises partly from the expectation in argumentation that, in cases of incompatibility, the second of the two propositions will normally be the one which expresses the writer’s own view. […] One way of capturing the impression given is by saying that the first proposition is conceded, whereas the second is asserted. (Thompson & Zhou 2000:126)

Aunt Sallies are a particularly interesting kind of concessive mechanism because they can operate over very large stretches of discourse and consequently may be used to give a backbone to the argument. They also tend to be found in thematic position and are frequently, though not uniquely, associated with prepositional phrases. The expressions which trigger the Aunt Sally phenomenon are to a greater or lesser extent conventionalised. The last three considerations make them relatively easy to identify and to teach to students. Students will find it useful to identify them as they help to answer the question of the writer’s commitment to her/his assertions, which so frequently seems to perplex students.

5.2 In itself

Let us now return to the Guardian (1 February 2003) article which we discussed in Section 3 and look at a further element, the item *in itself.*

<First Paragraph>*In itself, the fact that a majority of the country either opposes or is sceptical towards his policy on Iraq is not necessarily a reason for Tony Blair to change course. If a particular policy is right, then a minister is entitled to stick with it in the face of opposition, to trust to his own judgment and to take the consequences. That is how Mr Blair seems determined to see things on Iraq, fighting a lonely battle to persuade a doubting nation that George Bush has the British people’s best interests at heart. </First Paragraph>
This expression, as used in this example, creates the expectation that somewhere in the co-text at least a partial contradiction will appear. And, indeed, three sentences later we find the contradiction in the words:

> But Mr Blair’s view of the crisis — and of the role he is playing — remains dangerously double-edged. Mr Blair’s whole approach on Iraq may be put to its ultimate test very soon. [...]

The precise meanderings of the logic of the argument are a little difficult to follow; they can be summed up schematically in this way: in itself — Acceptable, but, More acceptable. The author has made the first part of the argument attractive only to dismiss it in favour of the more attractive second part. We have been presented with an Aunt Sally.

We are aware of this mechanism because of our intuition as expert readers of texts; those who read Guardian editorials regularly are even more expert in dealing with this sort of text. In Hoey’s (2004, 2005) terms, we have been primed to read the argument in this way. But most learners, even quite advanced ones, will have some difficulty in following the argument. Is the author saying that Blair should change his course or not?

In order to confirm our intuition we could use a dictionary, though in this case it would not be very helpful: not even the full-sized OED comments on the pragmatic force of this use of the phrase In itself. An alternative would be to use a corpus. We might turn to one of the general corpora, the BNC or the Bank of English, but as newspaper discourse, and especially the kind of argumentative strategies found in it, tend to be sui generis, I have chosen to use the corpus of newspaper texts called Papers93 described in Section 1.

There are 25 instances of the expression In itself, in the sentence initial or thematic position, in Papers93. Twenty-three of them clearly function in the way described above. In the two examples which function differently, an examination of the expanded context reveals a fundamental difference of meaning. The words in itself in the sequence In itself a grapevine can be a beautiful thing might be glossed: “a grapevine taken on its own and taking no account of other considerations”. Here In itself is roughly equivalent to the Latin eo ipso or the commonly used German philosophical tag an sich, “In itself, in the abstract, not in relation to anything else” (OED definition of an sich). In all the other 23 conforming examples the gloss would be significantly different: “as opposed to other considerations which are different from and in contrast to the one(s) cited, but perhaps more important and which I am going to mention”. This meaning accounts for 92% of the instances. In the preponderant meaning, In itself is part of an evaluative rhetorical structure, the second element of which it may be said to trigger off; the “grapevine meaning” does not have this function.
Speaking of lexical priming, Hoey (2005) makes clear throughout his book that lexical items only tend to behave in a certain way. We very rarely find an item which always behaves in the same way. This seems to be true also for Aunt Sallies. What is perhaps surprising is that In itself has the function of triggering off a counter argument more than nine times out of 10. So powerful is this priming that we need to read the whole of the paragraph in which the grapevine example occurs before we realize that there in itself is not a trigger.

In this paper I deal only with examples which occur in sentence initial or in thematic position. In itself and the other phraseologies which are discussed occur frequently in other syntactic positions but their interpretation in these positions, while not entirely different, is more problematic. An initial discussion of this can be found in Morley (2004b).

5.3 Prepositional phrases as Aunt Sallies

As I said in the last paragraph, other prepositional phrases function in the same way. Once again, I am considering only phraseologies found in the thematic position.

5.3.1 In theory

In 6,745,309 tokens from Guardian (Papers93) there are 39 instances of In theory in initial position of which 31 are followed by a counterstatement. An example:

Observing owls at close quarters is a rare event. In theory it should have been an occasion for undiluted pleasure. In practice it was a moment of confusion and we were unable to decide which way to look. (Guardian 4 January 1993)

In the eight other instances where it is not followed by a counterstatement, In theory is used literally to mean that the theory states that what follows is possible / natural / necessary. This is particularly evident in the following example, in which academic lawyers are quoted concerning the necessity of the Maastricht Treaty being ratified by parliament. Here In theory means literally that the theory of international law states what follows.

But British experts in community law said ratification was a matter for the Crown and Britain could ratify the treaty with or without the European Communities (Amendment) Bill. “The power to bind in international law is a Crown prerogative”, said Piers Gardiner, director of the British Institute of International and Comparative Law. In theory, we don’t need what’s going on in Parliament at all. (Guardian 12 February 1993)

However, it is sometimes difficult to escape the feeling that there is an unexpressed in practice lurking somewhere in the author’s mind.
GLAXO is trying to resolve a row with the powerful American Food & Drug Administration (FDA), which has accused it of illegally promoting its Zantac anti-ulcer drug in the US, in violation of US drug laws. Zantac is the world’s best-selling drug and the source of 44 per cent of Glaxo’s £4 billion annual turnover. In theory the FDA could get Zantac taken off the US market, which accounts for 40 per cent of sales, dealing a serious blow to the UK’s biggest company. (Guardian 8 March 1993)

We do not really think Zantac is going to be taken off the US market. This is not due simply to the traditional Guardian reader’s cynicism about the relationship between big business and the government in America, it is also because once a phraseology has acquired a semantic prosody, or in this case more correctly a rhetorical prosody, the prosody is felt even when it is not expressed: here, we mentally add but in practice it’s not very likely. As Sinclair (1996) observes once a prosody “is identified with a phrasing it will be part of the meaning even if it has no clear expression.” (Sinclair 1996:92)

5.3.2 In the past
If we leave aside expressions like in the past two years, there are 59 instances of In the past in thematic position in the same c. 6.7 million words from the Guardian (Papers93). In 43 cases (73%) it clearly triggers a following contradictory statement.

In the past, if companies overstepped the mark in advertising, they might have been slapped on the wrists. Now the FDA takes an adolescent, bullying attitude. (Guardian 8 March 1993)

There are three other instances where it might be said to trigger an opposition where there is no marker of the contrastive now element.

In the past they would have paid a full quarterly charge — even though they might have borrowed for a few days.
To sweeten the pill, the bank has abolished the pounds 20 charge for writing to customers and the £20 charge it currently levies for using a cash card when there are insufficient funds in an account. (Guardian 20 March 1993)

Here, although, the marker is absent, the context establishes that the words has abolished refer to the present situation. I have not counted this kind of example in the statistics.

5.3.3 Differences in Aunt Sally expressions
In theory and In the past have a slightly different rhetorical structure from In itself. We have described the In itself structure schematically as:

In itself — Acceptable, but, More acceptable.
The schematic structure of *In the past* would be:

**In the past — Was the case, now — Is the case**

For *In theory*

**In theory — Ought to be the case, in practice — Is the case.**

What the rhetorics have in common is that (1) they trigger later counter arguments, (2) their structure is fundamentally concessive, (3) they contain some kind of evaluation, in that one of the elements, usually the second, is considered ‘better’ than the other and consequently more likely to be the one the author embraces or avers.

Not all prepositional phrases dependent on *in* will act as Aunt Sallies. The expression *In the natural order of things* in sentence initial position might lead us to expect some comment on what really happens in the sublunary world in which we live but, in the corpora I have examined including the Web, it seems much more likely to be used to introduce a scientific argument about natural order.

Likewise, how often these phrases will behave as in this way cannot be predicted with absolute reliability: in the corpus examined *In itself* behaved as an Aunt Sally 92% of the time, while *In the past* so behaved only 73% of the time, or 78% of the time, if we count the three instances with no introductory marker. It would be well always to test our intuitions against a large corpus. Furthermore, I make no claims for the applicability of these statistics to other types of discourses than those found in newspapers. If they are accurate, they are true for the specialized corpus of a hundred million words of quality newspaper discourse produced in 1993. A newer, bigger and less discourse specific corpus might give different results. I have, however, checked out some of my findings with the *Bank of English* (500 million plus words) and on the Web without too contradictory findings.

Lastly, it is arguable that *in itself* would be better described as a grammatical item, perhaps an adverbial disjunct, whereas *in the past* has a more lexical feel, as it contains a lexical word which has not been desemanticized. Once again, as we observed in Section 2, we see how grammar and lexis work together to create cohesion. Hunston’s (2001:23) reference to Hoey’s plenary paper at the 1998 BAAL meeting seems particularly relevant here. In this paper Hoey suggests that collocation, or the patterning associated with lexical items, provides a bridge between lexis/grammar and text, as well as between lexis and grammar.

Expressions like *in itself, in the past, in theory*, by providing links within the discourse and thus creating cohesion, also bring together notions which in the past were considered totally distinct, like grammar, lexis and text / discourse.
5.4 Set phrases as Aunt Sallies

Moving on from prepositional phrases, we find that set phrases, too, if they are in the thematic position, can trigger a later counter argument. The metaphorical expression *To the naked eye* occurs five times in the hundred million words of *Papers*93 and on each occasion provokes a counterstatement. An example:

> Perhaps the most surprising disclosure is the freedom with which the diptych is painted. *To the naked eye* it looks undeviating in its details, but in fact the hair and faces of the angels are formed from myriad rapid, overlying strokes of various colours in an almost Impressionist technique. (Telegraph 28 August 1993)

Proverbs in particular lend themselves to being Aunt Sallies. They encapsulate given wisdom, which journalists can take for granted that their readers possess; they also give the journalists the opportunity to show their cleverness by overturning the wisdom.

> They say *familiarity breeds contempt*, but in the fourth meeting of these clubs in three weeks at The Valley last night, Charlton showed an exaggerated respect for the sanctity of their south London neighbours’ goal. (The Times 6 October 1993)

Over at the Mirror, page 12, "Our Man Hazza Gazza!", by Anthony Harwood, told an eerily familiar story. *Great minds think alike*, they say. *They also say fools never differ.* (Guardian 14 August 1993)

6. Conclusions

In conclusion, what I have suggested is that, first of all, lexical cohesion not only operates at the componential level, but can help to structure the discourse, in the present case in newspapers. As we attempted to show in Section 2 and Section 3, when analysing the expressions *choice* and *a moment for*, headlines are particularly powerful not just in indicating the lexical fields which the following article will contain but also, sometimes, in suggesting the kind of argumentation the body of the text will use. Coming to what I have called Aunt Sallies, we might say that certain lexical items, which are found in thematic position and are generally stereotypical in nature, trigger evaluation in the form of counter arguments and thus help to pilot us through the rhetoric of the text. They also represent one of the many linguistic devices that are beginning to be studied with increasing interest in the field of evaluation / stance / appraisal.

From the point of view of language learners, the recognition of the existence of all these signposts can be of great help in answering that very difficult question: But does the author really mean this? If he or she has used the kind of concessive
mechanism we have discussed in this paper, then the answer is that the author's commitment is limited, and s/he thinks that something else is better or more relevant now or more likely in the real world. More advanced students can test their intuitions against results from corpora. There is also a lot of room for research about these same phraseologies when they occur in non-thematic positions, for which the use of corpora is essential.

Notes

1. For a fuller description of this approach, which Partington calls Corpus-Assisted Discourse Analysis (CADS), see Partington (2004b).

2. The article is too long to quote in full. For an extensive discussion see Morley (1999).

3. To be more precise, as the headline is generally produced by a sub-editor and not the journalist who has written the article, the headline exploits the cohesion already present in the text. It remains true, obviously, that the reader encounters the headline first.

4. I am being very strict here in defining a semantic field. It could be argued that expressions like policy on Iraq, change course, stick with it, judgment, consequences, see things, view of the crisis, decides, good judgment and no doubt others belong to this semantic field at least in the sense that they are very likely collocates of choice.

5. This is not an attempted sleight of hand. I am well aware that the precise words in the phraseology are crucial. It does seem, however, clear that moment always has the potential to indicate a momentous moment and that the semantic prosody of moment for is carried over to the clause introductory moment when.

6. See Morley (2004a) for a discussion of the peroration in editorials. He argues that the linguistic elements that distinguish editorials, such as modal verbs and stance adverbials, increase in frequency towards the end of the article.

7. It is perhaps worth acknowledging the dialogic nature of these concessive mechanisms as pointed out by Winter (1982) and Hoey (1983). See also comments on White's Appraisal Forum (http://www.grammaticas.com/appraisal/).

8. The structure of Hunston's may be a is a mirror image of this. See Morley (2004b) for a more detailed discussion.

9. It may not be immediately apparent that seems determined to has a negative semantic prosody. There are 24 instances of the expression in c. 27 million words of the Guardian (Papers93), all of them are connected with negative following actions.

10. For rhetorical prosody see Morley (2004b).

11. The words familiarity breeds content are not sentence initial, but they are clause initial and in the thematic position, in Halliday's terms.
References


Lexical bundles and discourse signalling in academic lectures

Hilary Nesi and Helen Basturkmen
Coventry University / University of Auckland

This paper discusses some approaches to the categorisation of cohesive devices with reference to spoken academic discourse, multi-word units, and strings of frequently co-occurring words (lexical bundles). It goes on to investigate the cohesive role of lexical bundles in a corpus of 160 university lectures (120 from the BASE corpus and 40 from MICASE). Like the bundles from the T2K SWAL teaching subcorpus, investigated by Biber, Conrad and Cortes (2004), the bundles in the lecture corpus included both ‘oral’ and ‘literate’ elements. The majority of frequently occurring bundles were found to be used to signal discourse relations, although their cohesive function was not necessarily obvious when listed out of context.

Keywords: lexical bundles; linking ideas; lectures; cohesion; academic discourse; corpora; BASE; MICASE

1. Introduction

1.1 Cohesion and multiword units

Halliday and Hasan’s seminal work on cohesion in English (1976) identified four broad categories of textual relation: ‘addition’, ‘comparison’, ‘time sequencing’ and ‘result’. Subsequent researchers have developed this system, altering the terminology and categories to suit their individual purposes, but although there are many terminological differences between one system and another, the categories of additive, adversative, sequential and causal relations are common to most classifications (Louwerse & Mitchell 2003).

The Longman Grammar of Spoken and Written English (Biber et al. 1999) distinguishes between ‘coordinators’, ‘subordinators’ and ‘linking adverbials’ as grammatical devices which explicitly signal the connections between passages of
text. While coordinators and subordinators signal meaning relations within sentences, linking adverbials signal intrasentential relations, and can be compared to ‘cohesive conjunctions’, as described by Halliday and Hasan (1976). Biber et al. (1999) divide linking adverbials into six semantic groups: enumeration and addition (to list and add items to the ongoing discourse), summation (to conclude preceding discourse), apposition (to indicate equivalence or inclusion), result/inference (to signal consequence), contrast/concession (to indicate dissimilar comparison) and transition (to signal asides and lack of continuity with preceding discourse). Biber et al. (1999) report differing patterns of cohesion in conversation, academic prose, fiction and news reports, finding that linking adverbials occur more often in conversation and academic prose than in fiction and news. Adverbials signalling enumerative/summative and appositional relations are most common in academic prose.

Many of the items that have been identified as serving a signalling function in discourse are multi-word units rather than single words. Of course, some of the adverbs that typify cohesive conjunctions (Halliday & Hasan 1976) originally presented as two- or three-word units (furthermore, however, nevertheless etc), but came to be regarded as single words as they acquired a conventionalised pragmatic meaning. (Other such compounds seem to be in the process of transition — inasmuch as, insofar as, instead of etc.) In Halliday and Hasan’s scheme, conjunction is realised by either single word adverbs or multi-word prepositional phrases (such as in addition, as a result of that, in spite of that), but most recent research admits a broader range of syntactic structures into the cohesive system. Biber et al. (1999), for example, include within their category of ‘linking adverbials’ which is to say (as an adverbial of apposition), and to conclude (as an adverbial marking summation). Hyland (2004) categorises while it is true as a conjunction indicating concession.

Multi-word discourse devices are typically identified by combining text analysis with intuition about the way words are used. DeCarrico and Nattinger (1988) and Nattinger and DeCarrico (1992) examined the collocations within a span of five lexical items surrounding pre-selected node words in a corpus of lectures. They counted as lexical phrases those groups of words which appeared to be prefabricated, rather than generated by syntactic competence, and which served a pragmatic function in the discourse. A variety of multiword discourse devices were identified by this method, but the identification process required the researchers first of all to notice the existence of a prefabricated chunk and define its boundaries. For this reason Nattinger and DeCarrico’s lexical phrases tend to be perceptually salient, and structurally complete.
1.2 The identification and description of lexical bundles

An alternative method of identifying multi-word units is empirical rather than intuitive. Strings of frequently co-occurring words can be identified within a given corpus regardless of syntactic boundaries or their salience as meaningful units independent of context. Strings of any given length can be identified, and their meaning and function can be considered subsequently. Groups identified by this means are referred to by a variety of different terms, as Stubbs (2002:230) points out: they are known as ‘clusters’ (Scott 1997:41), ‘recurrent word-combinations’ (Altenberg 1998:101), ‘statistical phrases’ (Strzalkowski 1998:xiv), ‘lexical bundles’ (Biber et al. 1999:993) and ‘n-grams’ (Banerjee & Pedersen 2003).

Lexical bundles are discussed at length in the Longman Grammar of Spoken and Written English (Biber et al. 1999, Section 13.2). In this work a bundle is defined as ‘a recurring sequence of three or more words’ (1999:90), although most of Biber et al.’s analysis concerns longer sequences, particularly the four-word bundle. Biber et al. found almost ten times as many three-word bundles as four-word bundles in their corpus of conversation and academic prose, and almost ten times as many four-word bundles as five-word bundles. Four-word bundles included extended versions of the most common three-word bundles, and, as they were plentiful, they restricted their analysis to those occurring ten or more times per million words, in five or more different texts, in order ‘to exclude individual speaker/writer idiosyncracies’ (Biber et al. 1999:993).

One disadvantage of this method is that it does not permit the identification of discontinuous frames (for example not only... but also...). The pre-specification of string length does not preclude the identification of longer strings, however, as there is usually some overlap between the most frequently occurring strings of any given length. As the Longman Grammar of Spoken and Written English points out, ‘longer lexical bundles are usually formed through an extension or combination of one or more shorter bundles’ (Biber et al. 1999:993). By way of illustration, Biber et al. show the development of a string from three to six words:

\[
\begin{align*}
do\ you\ want;\ you\ want\ me;\ want\ me\ to;\ me\ to\ do & \rightarrow \\
do\ you\ want\ me;\ you\ want\ me\ to;\ want\ me\ to\ do & \rightarrow \\
do\ you\ want\ me\ to;\ you\ want\ me\ to\ do & \rightarrow \\
do\ you\ want\ me\ to. & 
\end{align*}
\]

Later studies have continued the practice of focussing on four-word bundles, often setting more rigorous cut-off points. Cortes (2004), for example, only considers bundles occurring 20 or more times per million words, and Biber, Conrad and Cortes (2004) only examine those occurring 40 times per million. Lexical bundles that occur with very high frequency across a range of texts are likely to be stored in
memory as unanalysed chunks, a particularly interesting consideration in view of
the fact that bundles tend to bridge syntactic boundaries and do not generally have
idiomatic meaning, and are therefore not very salient, either to the listener/reader
or to the language researcher. As Biber, Conrad and Cortes (2004:377) point out:

for the most part linguists have not noticed these high frequency multi-word se-
quen ces, probably because most previous research has focussed on grammatical
phrases and clauses, disregarding the possibility of lexical units that cut across
grammatical structures.

Biber, Conrad and Cortes (2004) compared lexical bundles in academic prose,
conversation and teaching sessions, using text samples taken from the Long-
man Spoken and Written English Corpus and the TOEFL 2000 Spoken and Writ-
ten Academic Language Corpus (T2K SW AL). Previous work on lexical bundles
had concentrated on written text and conversation, but Biber, Conrad and Cortes
(2004:382) found that classroom discourse made far more frequent use of bundles,
and contained the greatest variety of different bundle types. They concluded that
‘the extremely high density of lexical bundles in classroom teaching exists because
this register relies heavily on both ‘oral’ and ‘literate’ bundles. In their data ‘oral’
bundles, typical of conversation, are characterised by declarative and interrogative
clause fragments, while ‘literate’ bundles, typical of academic prose, contain noun
phrases and prepositional phrases. Their teaching subcorpus bundles contained
more nouns than the bundles in the conversation subcorpus, but more verbs than
the bundles in the subcorpora of textbooks and academic prose.

1.3 The cohesive role of lexical bundles

Biber, Conrad and Cortes (2004) do not explicitly discuss cohesion with reference
to lexical bundles, and do not provide many examples of bundles which perform a
cohesive role. Nevertheless some of the functions of lexical bundles they describe
seem comparable to those assigned to conjunctive relations in the model of cohe-
sion proposed by Halliday and Hasan (1976), and to functions identified in other
accounts of metadiscourse in academic text (for example Hyland 2004). Three pri-
mary functions are identified: I. stance (expressing attitudes or assessments of cer-
tainty), II. discourse organization (reflecting relationships between prior and com-
ing discourse), and III. reference (referring to physical or abstract entities, or to the
textual context). The bundles expressing function II and (to a lesser extent) function
III appear to have the greatest potential to perform a cohesive role. Bundles express-
ing function II are subdivided into two further categories: A, topic introduction/
focus, and B, topic elaboration/clarification. Category IIA includes such bundles
as if you look at; going to talk about; what I want to etc. These can be equated with
'frame markers' in Hyland's model of interactive metadiscourse in academic texts (Hyland 2004), and mark transition points in text. It is possible that they play a cohesive role as continuative items, fulfilling similar functions to *now* or *well* in the semantic system described by Halliday and Hasan (1976). Category IIB includes such bundles as *has to do with*; *you know I mean*; *on the other hand*; *as well as the* etc. These can be equated with 'transitions', the items in Hyland's model which are used to signal semantic relations between main clauses (Hyland 2004). These items likewise express some of the conjunctive relations identified by Halliday and Hasan (1976). Bundles expressing function III are subdivided into four further categories: A, identification/focus, B, imprecision, C, specification of attributes, and D, time/place/text reference. Most of the bundles assigned to categories IIIB (eg *and stuff like that*) and IIIC (eg *the size of the, the nature of the*) do not fit into Hyland's model of metadiscourse in academic texts (Hyland 2004) and do not appear to perform a cohesive function (with the exception of *as a result of*, which apparently marks a causal conjunctive relation). Many of the bundles in category IIIA, however, seem to signal exemplification (*that's one of the, one of the things, of the things that* etc) and might thus perform a cohesive role similar to that of 'additive apposition' in the system proposed by Halliday and Hasan (1976). Category IIID includes some items that might function as 'frame markers' in Hyland's model (*at the end of the, the beginning of the* etc) (Hyland 2004), and might perform the same role as temporal conjunctions (Halliday & Hasan 1976).

1.4 Hypotheses and research questions

Biber, Conrad and Cortes (2004) claim knowledge of only one study prior to theirs which examines the use of multi-word units in university lectures. This is the work of DeCarrico and Nattinger (1988) (also reported in Nattinger & DeCarrico 1992), which, as discussed earlier, identified units primarily on the basis of perceptual salience, and therefore tended to ignore those frequently occurring clusters that are syntactically incomplete, and composed of very common words which do not take on any special idiomatic meanings when they co-occur.

Biber, Conrad and Cortes (2004) examined lexical bundles that had been identified automatically, in the manner described in the *Longman Grammar of Spoken and Written English* (Biber et al. 1999). However most of the 'lectures' in their teaching subcorpus did not represent monologic, lecture-style discourse, because in T2K SWAL about three quarters of the recorded teaching sessions are more like lessons than lectures, with many short turns of less than 100 words (Biber, Conrad, Reppen, Byrd, Helt, Clark, Cortes, Csomay & Urzua 2004). This kind of classroom discourse is characterised by 'interactions among participants, and a focus on the
speakers’ personal concerns’ (Biber, Conrad & Cortes 2004:378), leading to more 'oral' lexical bundles. Lectures, on the other hand, are less interactive and more pre-planned, and in this respect may be closer to academic writing.

The circumstances that Biber, Conrad and Cortes (2004) consider responsible for the unusually high number of lexical bundles in their teaching subcorpus should also apply to monologic lectures, however. Like participants in conversation, both lecturers and teachers face real-time production constraints which encourage the use of prefabricated chunks. Moreover in both teaching contexts the information content is high and the pedagogic function will require that connections between propositions are made clear, although the fact that pre-planned monologic lectures are not co-constructed, and lack normal opportunities for negotiation of meaning, suggests that they may have a greater need for discourse structuring devices. Because of the differences between our lecture data and classroom discourse, and because of our particular focus on cohesion, we expect in this study to shed light on some devices that have not been discussed in previous accounts of the cohesive system.

As an exploration of the cohesive role of lexical bundles in a corpus of lectures, this study addresses two research questions:

1. How do the lexical bundles used in lectures compare to those in other registers, and particularly to those used in classroom teaching?
2. How do the lexical bundles used in lectures help to create cohesion?

2. Procedure

Our corpus was made up of 160 monologic lectures from the British Academic Spoken English (BASE) corpus and the Michigan Corpus of Academic Spoken English (MICASE). To create this corpus thirty lectures from BASE and ten from MICASE were selected from each of four broad disciplinary groupings: Arts and Humanities (325,873 words), Social Sciences (360,037 words), Life Sciences (316,762 words) and Physical Sciences (268,126 words). The total corpus size was 1,270,798 words, similar to that of the classroom teaching subcorpus used by Biber, Conrad and Cortes (2004) (1,248,800 words).

The BASE component was considerably larger than the MICASE component (882,980 words as opposed to 387,818 words) because although the two complete corpora are roughly equivalent in size, MICASE covers a wider range of speech event types and contains fewer examples of lectures dominated by a single speaker.

Lexical bundles in this corpus were identified using WordSmith Tools (Scott 1997). WordList cluster processing was activated by choosing the option Settings/
Min. & Max. Frequencies from the main WordList menu. A cluster size of four words was selected, with a minimum frequency of ten.

3. Results

3.1 The frequency of lexical bundles in lectures

Most of the commonest bundles occurred with very similar frequency in both the British and American lectures, although in the United States (relatively rare in BASE) was the 15th most frequent bundle in the MICASE component, and in the UK (relatively rare in MICASE) was the 10th most frequent in BASE. Slight differences in the way the two corpora were transcribed may have affected findings to a small extent: MICASE makes greater use of contracted forms such as gonna and wanna (the bundle we’re gonna talk about occurred 24 times in the MICASE sample) while the BASE transcribers were more likely to use going to and want to. Fillers and hesitation markers are marked as # in the BASE corpus and represented lexically in MICASE as hm, huh, mm, mhm, uh, um etc. However none of the most frequent bundles in either corpus contained markers of this kind, a further indication of their prefabricated nature.

There were 33,761 instances of four word bundles occurring at least ten times in the corpus as a whole. One thousand two hundred and sixty-six different four word bundles occurred ten times or more — about 996 per million words. Of these, 34 occurred at least 60 times — roughly 27 per million words. This finding is almost identical to that of Biber, Conrad and Cortes (2004); 28 four-word bundles occurred at least 60 times per million words in their teaching subcorpus, as opposed to 44 in conversation. Like classroom teaching, lectures use ‘a large set of different lexical bundles, while conversation relies on the extremely frequent use of a smaller set of bundles’ (Biber, Conrad & Cortes 2004:379).

3.2 The characteristics of lexical bundles in lectures

Table 1 compares the word classes of lecture bundle endings with those occurring in conversation and academic writing, as described in the Longman Grammar of Spoken and Written English (Biber et al. 1999:997 Table 13.3).

Like the bundles in the teaching subcorpus, the lecture bundles included both ‘oral’ and ‘literate’ elements. Table 2 reveals in more detail the mixed ‘oral’ and ‘literate’ nature of the lecture bundles. Their proportional distribution is compared with that of bundles in conversation and academic writing, as described in the Longman Grammar of Spoken and Written English (Biber et al. 1999:996 Table 13.1).
Table 1. The grammatical category of words ending bundles in lectures, conversation and academic prose (approximate proportional distribution)

<table>
<thead>
<tr>
<th>Word class</th>
<th>Example</th>
<th>Lectures</th>
<th>Conversation</th>
<th>Academic prose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td><em>I want to know</em></td>
<td>17%</td>
<td>40%</td>
<td>5%</td>
</tr>
<tr>
<td>Pronoun</td>
<td><em>well that’s what I</em></td>
<td>11%</td>
<td>15%</td>
<td>–</td>
</tr>
<tr>
<td>other function word</td>
<td><em>I went to the/in the case of</em></td>
<td>50%</td>
<td>40%</td>
<td>85%</td>
</tr>
<tr>
<td>Noun</td>
<td><em>at the same time</em></td>
<td>17%</td>
<td>–</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 2. Structural patterns in lectures, conversation and academic prose (approximate proportional distribution)

<table>
<thead>
<tr>
<th>Structure</th>
<th>Example</th>
<th>Lectures</th>
<th>Conversation</th>
<th>Academic prose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal pronoun + lexical verb phrase (+ complement clause)</td>
<td><em>I don't know what</em></td>
<td>15%</td>
<td>44%</td>
<td>–</td>
</tr>
<tr>
<td>Pronoun/NP (+auxiliary) + copula be (+)</td>
<td><em>it was in the</em></td>
<td>10%</td>
<td>8%</td>
<td>2%</td>
</tr>
<tr>
<td>(auxiliary +) active verb (+)</td>
<td><em>have a look at</em></td>
<td>14%</td>
<td>13%</td>
<td>–</td>
</tr>
<tr>
<td>Yes-no and wh-question fragment</td>
<td><em>can I have a</em></td>
<td>2%</td>
<td>12%</td>
<td>–</td>
</tr>
<tr>
<td>(verb +) wh-clause fragment</td>
<td><em>know what I mean</em></td>
<td>4%</td>
<td>4%</td>
<td>–</td>
</tr>
<tr>
<td>Noun phrase with post-modifier fragment</td>
<td><em>the nature of the</em></td>
<td>19%</td>
<td>4%</td>
<td>30%</td>
</tr>
<tr>
<td>Preposition + noun phrase fragment</td>
<td><em>as a result of</em></td>
<td>9%</td>
<td>3%</td>
<td>33%</td>
</tr>
<tr>
<td>Anticipatory it + vp/adjective P (+complement clause)</td>
<td><em>it is possible to</em></td>
<td>1%</td>
<td>–</td>
<td>9%</td>
</tr>
<tr>
<td>Passive verb + PP fragment</td>
<td><em>is based on the</em></td>
<td>–</td>
<td>–</td>
<td>6%</td>
</tr>
<tr>
<td>(verb +) that-clause fragment</td>
<td><em>should be noted that</em></td>
<td>3%</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>(verb/adjective +) to-clause fragment</td>
<td><em>are likely to be</em></td>
<td>3%</td>
<td>5%</td>
<td>9%</td>
</tr>
<tr>
<td>Other expressions</td>
<td></td>
<td>17%</td>
<td>6%</td>
<td>6%</td>
</tr>
</tbody>
</table>

Biber, Conrad and Cortes (2004) found that whereas conversation made greater use of VP-based and dependent clause bundles, and textbooks made greater use of NP/PP-based bundles, classroom teaching made roughly equal use of all three kinds of bundles. The first four structures in Table 2 are commoner in conversation than in academic writing, and the second four structures are commoner in academic writing than in conversation. They all occur with some frequency in the lecture corpus, although question fragments are not well represented. In our corpus questions sometimes serve as a means of structuring discourse and managing topic change, but there is very little use of the questioning strategies typical
of classroom interaction (as schematised by Sinclair & Coulthard (1975) in their Initiation Response Feedback text pattern). In contrast Biber, Conrad and Cortes (2004:382) report that classroom teaching ‘makes dense use of lexical bundles that represent declarative and interrogative clause fragments’.

Structures within the ‘other expressions’ category in Table 2 included four word numerical expressions, long nominal compounds and adverbial clause fragments such as if-clauses.

We examined in particular detail those four word lexical clusters that occurred at least ten times within each disciplinary grouping, and over 50 times in the corpus as a whole. By selecting according to range of occurrence as well as frequency we eliminated from our study numerical sequences and technical terms (e.g. the DNA and RNA Polymerase, both of which occurred 50 times in the corpus as a whole, but only in one disciplinary grouping). By this means we also eliminated any clusters that were idiosyncratic to a single speaker.

Table 3 shows the 20 most frequent lexical bundles in the corpus, all of which occurred at least ten times in each broad disciplinary grouping. Four pairs of bundles in this list can be combined to form some of the most frequently occurring five word bundles:

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>MICASE</th>
<th>BASE</th>
<th>Life</th>
<th>Arts</th>
<th>Social</th>
<th>Physical</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>THE END OF THE</td>
<td>171</td>
<td>44</td>
<td>127</td>
<td>29</td>
<td>76</td>
<td>36</td>
</tr>
<tr>
<td>2.</td>
<td>AT THE END OF</td>
<td>152</td>
<td>37</td>
<td>115</td>
<td>35</td>
<td>57</td>
<td>31</td>
</tr>
<tr>
<td>3.</td>
<td>IF YOU LOOK AT</td>
<td>132</td>
<td>38</td>
<td>94</td>
<td>30</td>
<td>32</td>
<td>43</td>
</tr>
<tr>
<td>4.</td>
<td>TO BE ABLE TO</td>
<td>128</td>
<td>43</td>
<td>85</td>
<td>43</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>5.</td>
<td>AT THE SAME TIME</td>
<td>118</td>
<td>36</td>
<td>82</td>
<td>24</td>
<td>42</td>
<td>35</td>
</tr>
<tr>
<td>6.</td>
<td>IF YOU WANT TO</td>
<td>113</td>
<td>13</td>
<td>100</td>
<td>31</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>7.</td>
<td>ONE OF THE THINGS</td>
<td>110</td>
<td>43</td>
<td>67</td>
<td>34</td>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>8.</td>
<td>IN TERMS OF THE</td>
<td>109</td>
<td>29</td>
<td>80</td>
<td>23</td>
<td>22</td>
<td>45</td>
</tr>
<tr>
<td>9.</td>
<td>IS GOING TO BE</td>
<td>104</td>
<td>14</td>
<td>90</td>
<td>18</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>10.</td>
<td>AND YOU CAN SEE</td>
<td>91</td>
<td>22</td>
<td>69</td>
<td>28</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>11.</td>
<td>A LITTLE BIT OF</td>
<td>83</td>
<td>32</td>
<td>51</td>
<td>37</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>12.</td>
<td>YOU CAN SEE THAT</td>
<td>82</td>
<td>13</td>
<td>69</td>
<td>31</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>13.</td>
<td>OF THE THINGS THAT</td>
<td>76</td>
<td>31</td>
<td>45</td>
<td>31</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>14.</td>
<td>THE REST OF THE</td>
<td>75</td>
<td>34</td>
<td>41</td>
<td>27</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td>15.</td>
<td>AND THIS IS THE</td>
<td>74</td>
<td>18</td>
<td>56</td>
<td>30</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>16.</td>
<td>I WANT YOU TO</td>
<td>72</td>
<td>30</td>
<td>42</td>
<td>16</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>17.</td>
<td>YOU LOOK AT THE</td>
<td>69</td>
<td>16</td>
<td>53</td>
<td>21</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>18.</td>
<td>I’M NOT GOING TO</td>
<td>68</td>
<td>6</td>
<td>62</td>
<td>20</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>19.</td>
<td>TO DO WITH THE</td>
<td>65</td>
<td>18</td>
<td>47</td>
<td>13</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>20.</td>
<td>A LOT OF THE</td>
<td>63</td>
<td>14</td>
<td>49</td>
<td>15</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>
Seventeen of these 20 most frequent bundles are listed by Biber, Conrad and Cortes (2004:384–8) as common bundles in classroom teaching. Three did not occur frequently in any of their subcorpora (*you look at the*; *and you can see; you can see that*).

3.3 The cohesive role of lexical bundles in lectures

The objective of the next stage of the research was to investigate whether the bundles played a discourse signalling role. The concordance function from the software package MP 2.2 (Barlow 2002) was used to retrieve and display the concordances surrounding the occurrences of these bundles. We used a method similar to that used by Flowerdew (2003) in an investigation of signalling nouns in biology textbooks and lectures. A qualitative analysis of the concordance lines displayed on the computer screen was conducted. We examined the concordances to identify whether in some instances the bundles appeared to function to link parts of the discourse, and when this appeared to be the case the software was used to retrieve a longer section of co-text. In general this involved tracking the text back to the point where the topic of the text segment in which the lexical bundles appeared was introduced into the discourse, and tracking the text forward to the point where the second idea unit ended. This long section was then examined with the aim of identifying the type of cohesive relation involved. We made use of the semantic categories of linking adverbials listed in the *Longman Grammar of Spoken and Written English* (Biber et al. 1999): enumeration/addition, summation/conclusion, apposition, result/inference, contrast/concession and transition.

To help determine the type of relationship involved we tried, on an intuitive basis, to replace the bundles with tokens widely recognised as signalling different types of relationship between parts of text, such as *in addition*, *so*, *in other words* and *by the way*. No attempt was made to identify all instances in the data when lexical bundles appeared to signal discourse relationships, or quantify the extent
to which bundles perform this role. Our aim was to establish whether bundles had the potential to contribute to the cohesive system in lectures, and to draw attention to some of the ways in which this might be achieved.

With the exception of *is going to be* and *the rest of the*, all the frequent lexical bundles listed in Table 3 were observed to function at times to signal discourse relations. In other contexts, all the bundles also served other functions, often as directives or as expressions of stance. The bundles *if you look at*/ *you look at the* → *if you look at the*, for example, functioned variously as directives, ‘topic introduction/focus’ discourse organizers, and referential expressions (signalling exemplification). Biber, Conrad and Cortes (2004:383) point out that bundles can have multiple functions even within a single occurrence.

In this article there is only space for a few examples to show the main cohesive relationships signalled by bundles in the corpus. An examination of the bundles in their surrounding text revealed that their discourse signalling functions were of two kinds. They appeared firstly to signal how one idea or piece of information was related to another idea or piece of information in the lecture. In the discussion below such signals will be referred to as ‘referential expressions’ (borrowing the term used for category III markers in Biber, Conrad & Cortes 2004). They appeared secondly to signal the relationship between topics and activities in the lecture, roughly equivalent in function to ‘frame markers’ (Hyland 2004:138), which are references to text boundaries, and devices used to sequence and label stages in the text and announce changes in topic and discourse goals. In the following discussion, markers with this function will be referred to as ‘discourse organizers’ (borrowing the term used for category II markers in Biber, Conrad & Cortes 2004).

3.3.1 Referential expressions

Bundles sometimes seemed to indicate logical relationships such as apposition, contrast/concession and result/inference (all semantic categories of linking adverbials in the *Longman Grammar of Spoken and Written English*, Biber et al. 1999).

The relationship of apposition is described in the *Longman Grammar of Spoken and Written English* as “showing that the second unit of text is to be treated either as equivalent to or included in the preceding unit… an appositive linking adverbial can be used to show that the second unit is to be taken as a restatement of the first, reformulating the information it expresses in some way or stating it in more explicit terms” (Biber et al. 1999:876). In our corpus we found appositive relations signalled by the bundles *and you can see*/ *you can see that*; *if you look at*/ *you look at the*; *one of the things*/ *of the things that*; *a lot of the*; and *and this is the*. Examples 1 to 3 illustrate propositions related by apposition. After each example, one of the linking adverbials listed in the *Longman Grammar of Spoken and
Written English under apposition is provided. It is suggested that these adverbials could replace the lexical bundles in the extracts with small changes to the sentence structure but almost no change in meaning.

Example 1
He takes the name victorious general that’s what imperator means so you can see now when you meet him on the street you could say oh how are you you are the victorious general son of a god Caesar how very nice for you and you can see how Roman names can be used to carry a message (in other words)

Example 2
They got a B on a paper they thought they were going to get an A on and then miss something like a thyroid diagnosis so it’s really important to get a grip on the physical one of the things that is very easy to forget to ask about especially when you are confronted with a person who is clean cut looks good very pulled together very articulate is to ask about drug and alcohol abuse (for example)

Example 3
There is one more polarity which you need to register it’s crucial because it lies at the very heart of Huck’s moral dilemma and this is the opposition between conscience and heart (namely).

In our corpus it was not uncommon for lecturers to use frequently occurring lexical bundles alongside other items conventionally recognised as cohesive conjunctions. The following two examples show this; apposition is signalled by both a lexical bundle (if you look at) and a linking adverbial (for example in Example 4, and for instance in Example 5).

Example 4
well there is lots of apparent evidence which has been interpreted as evidence of social learning for example if you look at different colonies or troops of chimps the same species but living in different parts of Africa you find that different ch-troops show different behaviours

Example 5
once you’ve got them and you’ve got some numbers on the diagram you can de-cide which are the sort of dominant issues and if something isn’t going to be im-portant in the overall heat transfer for instance in this problem if you look at the thermal resistance of the glass it’s very very small so in terms of changing the heat transfer it really wouldn’t matter if you made the glass a bit thicker.

The category of contrast/concession contains, according to the Longman Grammar of Spoken and Written English, ‘items that in some way mark incompatibility between information in different discourse units, or that signal concessive relationships’ (Biber et al. 1999:878). Interestingly, we found that the lexical bundle
at the same time often signalled contrast, although its residual meaning suggests a temporal relationship (the phrase is given as an example of a temporal conjunction in Halliday and Hasan’s summary table of conjunctive relations (1976:242–3)). Examples 6 and 7 illustrate the contrastive use of at the same time. In Example 6 the lecturer is discussing problems with Canada Geese. Contrast is signalled by but, and reinforced by the contrastive use of at the same time.

Example 6
They defecate in tremendous amounts, um, on sidewalks where runners are jogging around the lakes in the twin cities there’s always a crisis about too many geese, but at the same time, um, when they tried to do something about it, there were people who wanted them there, even in large numbers.

In Example 7 three parallel structures (‘If … he says…’; ‘If he says…’; ‘if he then goes on to say…’) introduce three examples of different modes of thinking. The bundle at the same time before the second example is used in the same way as but before the third example, to signal the contrast between simplex, complex and multiplex thinking.

Example 7
If you go to an historian an American historian and you say tell me about Abraham Lincoln and what happened in the theatre and he says the bullet the calibre of the bullet was so and so the assassin was called such and such this is what happened that’s simplex. If he says at the same time there were twelve other people trying to assassinate him and security arrangements had been bunked up and the theatre had this he did okay that’s pretty complex but if he then goes on to say but if Abraham Lincoln had not been Abraham Lincoln who had been elected but somebody else this would have happened if Abraham Lincoln had not been shot then this is what would I think would have happened to American history that’s a bit of multiplex thinking.

Examples 8 to 11 show lexical bundles that signal result/inference. In Example 8 the two most frequent bundles in the corpus are combined: the end of the and at the end of. Typically these bundles did not serve as reference markers, being used in their literal sense to refer to points in time and place external to the lecture (the end of the war, the end of the stage etc.) or as a means of structuring and predicting events within the lecture itself (for example I’ll come back to it at the end of the lecture). There were, however some examples of reference marking with the idiomatic expression at/by the end of the day. COBUILD defines this expression in the following way:

You say at the end of the day when you are talking about what happens after a long series of events or what appears to be the case after you have considered the relevant facts. (Collins COBUILD Advanced Learner’s Dictionary 2003)
Example 8 is interesting because the intention to communicate result/inference (what appears to be the case after you have considered the relevant facts) overrides the more residual meaning which might be paraphrased as ‘finally’ (what happens after a long series of events). The lecturer recommends that statistical advice should be taken ‘at the end of the day’, as a logical consequence of the arguments put forward earlier, although the taking of statistical advice is also in fact the first step in the process of making a proposal.

Example 8
So you’ve got to be sure that when you make your proposal not only is your methodology right but you’ve got to have access to the cere-, to the to the facilities to do the work and at the end of the day a bit like statistical power really you ought to have statistical advice in other words take statistical advice before you start not at the end.

Example 9 describes a causal chain of events. The first causal link is not explicitly signalled: knowledge moves on and (implicit — because of this) you realise a mistake has been made. The second causal link is signalled with the cohesive conjunction so: you realise a mistake has been made (explicit — so) you go back and try to re-think. A third causal link is signalled by the bundle to be able to: you need the chain (explicit — to be able to) get back. The relationship between the latter two propositions could be restated in the following way — You need the chain because you need to get back.

Example 9
Knowledge moves on you realise that a mistake has been made way down there somewhere and so you go back and try to re-think you need the chain you need to be able to get back so what you leave in place is not that you are responsible for the whole pyramid but you are responsible for your brick and stone.

As in Examples 4 and 5, the lecturers in Examples 10 and 11 make use of two methods of signalling, both linking adverbials and lexical bundles. In Example 10 hence reinforces the function of the bundle the end of the day, and in Example 11 so reinforces the function of in terms of. In both cases the linking adverbials therefore or consequently could be substituted for the multi-word expressions.

Example 10
What we’ve done is we’ve included within their utility function the idea of everything that gives them satisfaction and hence at the end of the day they must allocate all of their incomes to those things that give them satisfaction including major saving
3.3.2 Discourse organizers

Some of the frequent bundles in our data appeared to be used to signal how the topic and/or activity of the discourse in one part of a lecture related to that in another. ‘Transition’, a category of linking adverbials in the *Longman Grammar of Spoken and Written English*, contains items signalling the introduction of discourse that is loosely connected or not connected to previous discourse. These adverbials “mark a transition to another, usually tangential, topic” (Biber et al. 1999:879). Signals of transition were relatively rare in all the registers investigated (conversation, academic prose, fiction and news reports); this is probably because the written registers could make use of typographical devices to signal topic change and asides, whilst in conversation the topics are not usually pre-planned, but are negotiated by all the participants.

Biber, Conrad and Cortes (2004:386) identified a number of ‘topic introduction/focus’ discourse organizers which were common in their teaching subcorpus (occurring 40 to 99 times per million words) but were rare in other registers (occurring fewer that ten times per million words). These included *if you look at; take a look at; if you have a; if we look at; going to talk about; to look at the; to go ahead and; I want to do; what I want to; want to do is; want to talk about; you know if you; a little bit about*. The same or similar bundles also had this function in our lecture corpus. Example 12 shows the use of *if you look at* to signal transition. The end of discussion of the previous topic is signalled by the linking adverb *finally*, and the start of discussion of the next topic is signalled by both the linking adverb *so* and the lexical bundle.

Example 12

and *finally* in this sort of hierarchy of grafting we have a xenograft and xenogeneic graft comes from or xenogeneic graft comes from a member of a different species for example pig to man yeah *so let’s if you look at* the range of transplant medicine what is done and why i’m gonna talk briefly about currently successful grafts

As Biber, Conrad and Cortes (2004:382) point out, ‘topic introducing bundles often result in syntactic blends’. In Example 12, *if you look at* is inserted mid sentence immediately following ‘let’s’, and functions as if it were a non-finite verb form. As a result the sentence ‘finishes up in a way that is syntactically inconsistent with the way it began’ (Biber et al. 1999:1064):
Some of the other bundles Biber, Conrad and Cortes (2004) list as signalling
topic introduction/focus were not common enough to be listed in the top 20 in
our corpus, but occurred with a certain amount of frequency, for example I want
to do (47 instances), what I want to (73 instances) and want to do is (37 instances).
All of these seemed to function on occasion as signals of topic change. The more
frequent cluster I want you to is categorised as a stance expression with a directive
function by Biber, Conrad and Cortes (2004), rather than as a topic introduc-
tion/focus discourse organizer. In most of the examples in our corpus, however, it
functioned simultaneously as a directive and a signal of activity change. It was fre-
quently preceded by another marker such as so or now, as in Examples 13 and 14:

Example 13
So I want you to go through this diagram and make sure that you can do that for
each step

Example 14
So now I want you to spend a few minutes, um, thinking about the consequences

Example 15 provides further evidence of the prefabricated nature of bundles. The
speaker makes a false start, backtracks by inserting I want you to, and then re-
sumes the directive.

Example 15
I'm going to ask you questions and think I want you to think back to France and
Germany and to the Netherlands.

Example 16 illustrates the use of I want you to with the downtoner a little bit.

Example 16
okay operation mode which we'll look at in a minute and an effective address well
i want you to do a little bit now i've been talking for twenty minutes about time
you did something so write down all the possible forms all right for variants one
and two okay.

In this example the lecturer is preparing the students to perform a task, and the use
of a little bit may help to mitigate the threat to the audience's face by indicating that
the imposition on their time and concentration will not be great.

The bundle a little bit about, listed by Biber, Conrad and Cortes (2004) as a
common topic introduction/focus discourse organiser in classroom teaching, oc-
curred 56 times in our lecture corpus but was not amongst the 20 most frequent
bundle types. In almost every case it helped to signal topic change, as in Examples
17 and 18, or it functioned summatively to mark the end of a topic, as in Examples
19 and 20:
Example 17
Let's just think a little bit about this planar haem molecule

Example 18
I want to talk first before I stop a little bit about this issue of tissue matching

Example 19
I've talked today a little bit about the attributes of nationalism

Example 20
So that's a little bit about myself.

Biber, Conrad and Cortes (2004) list the variant a little bit of as a referential expression, linking propositions rather than topics or activities, but several of the instances in our corpus seemed to function to indicate topic change, often in conjunction with another signalling cluster. Examples 21 and 22 show the prospective use of a little bit of, proceeded by the frame markers now and what I want to do now:

Example 21
Now I'm going to dip into a little bit of my own research

Example 22
What I want to do now is to get back to a little bit of economics.

Example 23 illustrates retrospective use to signal the end of a digression, in conjunction with the frame markers all right and so:

Example 23
All right so that's a little bit of a sideway sideline.

Any sort of engineered change in topic or activity is essentially directive, and therefore potentially face threatening, and in all the above contexts the expression a little bit seems to help maintain face; either that of the audience, by downplaying the task imposition (Examples 17, 18, 21, 22), or that of the lecturer, by downplaying apparent digressions (Example 23, and possibly 20) and by giving the audience to understand that the information provided is only a small part of what there is to know and what the lecturer actually knows about the subject (Examples 18, 19 and 22). Similarly in Examples 20 and 21 the lecturers mitigate a potential threat to their audience's negative face by using a little bit to downplay talk about themselves and their own contribution.

The bundle in terms of the is also listed by Biber, Conrad and Cortes (2004) as a referential expression, like a little bit of. In our corpus in terms of the sometimes seemed to be used to establish logical relationships in the text (as in Example 11) but also seemed to be used to mark a transition between topics, as in Example 24.
In this example, following an aside by the lecturer on his or her own ways of coping with study, work and family as a student, *in terms of the* is used to signal a return to the main topic, the assignment. The bundle appears to signal retrospectively that the preceding discourse was tangential to the main discourse.

**Example 24**

I survived an MBA a family and a job purely by working with a group of mates here we sorted out the competencies of each of us the time of each of us and sort of divvied up the work. *In terms of the* presentation of the final paper, that’s an individual piece of work and must be handled as such.

Finally, the following examples show the use of *I’m not going to* prospectively, to narrow the focus of discussion (Examples 25 and 26), and summatively, to signal the end of a lecture topic (Examples 27 and 28).

**Example 25**

Yeah now this is what this says and *I’m not going to* talk about the second equation but just the first one

**Example 26**

I told you last time because I didn’t finish the lecture on Rawls and utilitarianism that I was going to finish it today but actually I’ve decided *I’m not going to* I know you’ll be very disappointed but I think you’ve had enough of that so I’m going straight into Nozick and his criticisms of Rawls

**Example 27**

You can get some idea of the w- the snow water content there and *I’m not going to* do any more and talk about this

**Example 28**

I’ve set out my defence there — you’ve got it *I’m not going to* say anything more.

4. Summary and conclusion

The analysis in the previous section indicates that four word lexical bundles can play a discourse signalling role in lectures, and we would argue that it is important for language learners to be aware of this. While native speakers of English can be expected to have implicit knowledge of the function of bundles, non native speakers are much less likely to have this understanding because they have consciously learned the language, rather than acquired it, and the role of lexical bundles as discourse signals is yet to be acknowledged in most language teaching materials.
In lectures students are required to process relatively long stretches of discourse featuring a complexity of ideas, discourse topics and activities. Discourse signals are intended to help the listener predict the nature of upcoming ideas and information, and a student who is unable to recognize these signals will be faced with additional cognitive processing demands, having to deduce both the intrinsic meaning of propositions, and make inferences about the relations between them.

Experimental studies have indicated that discourse signals in written text affect a reader’s ability to comprehend the text and recall information from it (Martínez 2002). Hoey (2001) argues that readers formulate hypotheses about how written text will develop, and this helps them interpret it correctly. According to Hoey (2001:32) “accurate recognition of the signals and their significance for the text’s development can greatly ease a reader’s processing burden in that it lessens the need for large-scale hypothesis forming at the same time as all that detailed micro-processing going on’. We would argue that this is also the case when listeners interact with lectures.

A considerable body of research has focused on the devices used in written text to make the links between sentences unambiguous for the reader (a recent example being Cortes 2004). The devices used in spoken text have been less extensively researched, however, and only limited information about them is available to teachers. Parrot (2000) makes some attempt to offer such a description, including, for example, items in ‘general use’ for signalling exemplification (for example, for instance and e.g.) and providing one item ‘used mainly in speaking’ (say). It is clear that more data-driven descriptions of the features of spoken language are needed to supplement the lists of discourse markers typically presented to learners of English as a foreign language.

Some writers (Flowerdew 2003; Thurstun & Candlin 1997, 1998; Weber 2001) suggest an inductive approach to teaching vocabulary use, in which students employ corpus-based research methods, referring to on-screen concordances (or print outs from them) to examine how words are used in context. This might also be a suitable way to teach students about discourse organising devices. If applied to spoken discourse, such an approach would entail providing transcriptions for students to examine, or requiring the students themselves to transcribe pre-selected excerpts from recordings (such as the examples given in Section 3 of this paper). They might then be required to identify the ideas and topics in the excerpts, and discuss the relationships between them and the means whereby these relationships are signalled.

As far as we are aware, the cohesive role of lexical bundles has not yet been examined in the classroom in this way. Now that lists of frequently occurring bundles are becoming available, however, we think that it is useful for learners
to investigate their use as signalling devices in authentic text, alongside the better known exponents that have come to feature prominently in English language teaching materials since the publication of the seminal work of Halliday and Hasan (1976).

Notes

1. The British Academic Spoken English (BASE) corpus was developed at the Universities of Warwick and Reading under the directorship of Hilary Nesi and Paul Thompson, Corpus development was assisted by funding from BALEAP, EURALEX, the British Academy (2000–2001: SG 30284), and the Arts and Humanities Research Council (2001–2005): RE/AN6806/APN13545).

2. MICASE is on-line, searchable collection of transcripts of academic speech events recorded at the University of Michigan. http://www.lsa.umich.edu/eli/micase/index.htm

References


Lexical bundles and discourse signalling in academic lectures


Cohesive chains and speakers’ choice of prominence

Martin Warren
The Hong Kong Polytechnic University

This paper describes lexical cohesion across participants in a discourse, and across discourse events, and the additional contribution made by speakers’ choices of prominence in the cohesive chains. The choice of prominence is made to communicate what is perceived to be situationally informative in the local context of interaction (Brazil 1985 and 1997). This study describes a connection between a speaker’s choice of a lexical or non-lexical word in a cohesive chain and the speaker’s choice of prominence and, in doing so, offers a partial answer to the question Hoey (1991:17) asks of Winter (1974 and 1979): ‘under what circumstances do we use one (cohesive device) rather than the other?’. It also describes how the selection of prominence across speakers in a discourse can signal divergence and convergence between them. Lastly, the paper discusses the pedagogical implications of the study with regard to English language learning and teaching.

Keywords: lexical cohesion, cohesive chain, discourse intonation, Hong Kong Corpus of Spoken English, prominence, job placement interview

1. The aim of the study

The aim of this ‘corpus-driven’ (Tognini Bonelli 2002:75; Sinclair 2004:12) study of spoken discourse is three-fold. First of all, it describes a connection between a speaker’s choice of a lexical or non-lexical word in a cohesive chain. Then, it explores the role played by discourse intonation in creating additional layers of meaning to lexical cohesive links in order to examine the communicative value of discourse intonation in lexical cohesion. Lastly, it discusses the pedagogical implications of the findings. The genre of discourse examined in this study is job placement interviews in the context of five-star hotels in Hong Kong. The study deals with the ‘interactional’ aspect of lexical cohesion in job placement interviews;
in other words, how speakers use words in a ‘systematic and patterned way for interactive purposes’ (McCarthy 1988:181–182). As Partington (1998:89) points out, there has been relatively little textual analysis based on corpora, compared to a relatively large number of studies of lexis and syntax. Nonetheless, Partington (1998:89–106) demonstrates how the methodological combination of wordlists, concordance lines, and viewing the lexical word under investigation in its original context, can enable the researcher to examine patterns of textual cohesion.

2. Grammatical and lexical cohesion

The contribution of lexical cohesion to the organizational properties of texts has long been acknowledged by discourse analysts. According to Hoey (1991:10), lexical cohesion, which forms multiple relationships within a text, is the main source of textual coherence, and lexical cohesion is the most common form of cohesive tie, accounting for 40% of the cohesive ties cited in Hasan (1984). While some studies tend to distinguish between grammatical and lexical cohesion or repetition (see for example Halliday & Hasan 1976), the notion of what may be considered to be part of the same cohesive chain in a discourse proposed by Winter (1974, 1979) is broader because he includes ellipsis, substitution, reference and lexical repetition. In the present study examining cohesion in spoken discourse, Winter’s (1974, 1979) approach to the ties between lexical and non-lexical cohesion is adopted as a more appropriate way of portraying the complexity of lexical cohesion, especially when the use of ellipsis and reference are generally far more prevalent in spoken discourse than in written discourse (see, for example, Cheng & Warren 1999). Hoey (1991) notes that Winter’s broad definition of cohesion basically describes all of the cohesive devices as functioning as repetition, and Hoey then asks an important question which he says needs further investigation: ‘under what circumstances do we use one rather than the other?’ (Hoey 1991:17). This study attempts to offer at least a partial answer to Hoey’s question.

In his research on vocabulary patterns in conversation, McCarthy (1988:185) examines the lexical chains in the form of what he terms ‘relexicalisations’ across turns and across speakers in conversational data. These relexicalisations cover non-identical repetitions that have been re-formulated or re-structured. McCarthy’s study is relevant to this study because he also examines the speakers’ choice of discourse intonation (Brazil 1985 and 1997). McCarthy (1988) makes the case that to fully understand reiteration (i.e. lexical chains), along with other forms of cohesion, the analyst working with spoken discourse needs to take into account the communicative role of intonation within the cohesive chain. McCarthy (1988:183)
argues that, in spoken discourse, intonation choices made by speakers also serve to signal ‘here-and-now lexical relationships projected by the speakers’. One means by which speakers do this is through one of the four systems in discourse intonation: prominence (the other three being tone, key and termination) and it is the communicative value of this intonation choice that is explored here.

This study is also in line with McCarthy’s (1988) attempt to include an analysis of discourse intonation in the study of lexical cohesion, but differs from his in that it studies a wider range of words within a cohesive chain. By taking into account the interdependent relationship of both lexical and non-lexical cohesion, and speakers’ intonation choices in the data, the study examines the contribution of cohesion to the overall organisation of the discourse and the interactional management of cohesion over speaker turns.

In spoken English, received wisdom is that it is generally lexical words that tend to be made prominent by speakers while non-lexical, or grammatical, words tend not to be made prominent unless for a reason (see, for example, Chun 2002:9; Clark & Yallop 1990:296). This study, following Brazil’s (1997) discourse intonation system, aims to investigate, with examples drawn from naturally occurring spoken discourse, the reasons for this phenomenon. It also investigates the exploitation by speakers of prominence on occasion to indicate convergence or divergence with the other participant(s).

3. The system of prominence in discourse intonation

Half of the Hong Kong Corpus of Spoken English (HKCSE), equivalent to one million words, including the job placement interviews, has been prosodically transcribed (Cheng, Greaves & Warren, 2005). The prosodic transcription, which is based on the discourse intonation system developed by Brazil (1985 and 1997) and others (see, for example, Coulthard & Brazil 1981; Coulthard & Montgomery 1981; Sinclair & Brazil 1982; Hewings 1990; Cauldwell 2002), is primarily concerned with the function of intonation in English and its communicative value. This system is of particular relevance to those wishing to study the communicative role of intonation. The choice of discourse intonation for the prosodic transcription of the Hong Kong Corpus of Spoken English (HKCSE) is also in line with those (see, for example, Couper-Kuhlen & Selting 1996:12–13) who call for the examination of the functions of intonation in naturally occurring discourse to better determine their pragmatic and situated meanings.

Discourse intonation consists of a set of choices available to speakers and these are not formulated with reference to grammar and do not have fixed attitudinal
meanings. A study by Chun (2002:15–45) of the various approaches to the phonological organisation of prosody points out that discourse intonation is a break with other traditions. It is in contrast to those (see, for example, Chomsky & Halle 1968; Liberman & Prince 1977; Pierrehumbert 1980) who see rule-driven generative phonology as a natural follow-up to work in generative grammar. Later work by the generative phonologists has tried to assign meaning to intonation, but the data used has been experimentally acquired for the most part (see, for example, Pierrehumbert & Hirschberg 1990). Discourse intonation is also in opposition to those, such as O’Connor and Arnold (1973) and Crystal (1975 and 1995), who seek to describe tones based on the fixed attitudes that they convey.

According to Brazil (1997:23–25), the choice of prominence on a syllable in a lexical word, which is realised through the speaker choosing to ‘stress’ the word, is not determined by grammar or word-accent/stress. Instead, speaker decisions within the prominence system are made on the basis of the speaker considering the status of individual words (Brazil 1997:39). For Brazil (1997), speakers have available to them two paradigms: existential and general. The existential paradigm is the set of possibilities that a speaker can choose from in a given situation. The general paradigm is the set of possibilities that is inherent in the language system. Brazil (1997:22–23) exemplifies the two paradigms with his well-known queen of hearts said in response to which card did you play. In this utterance, of is a product of the general paradigm because the speaker is limited in this context to this word by the language system. Conversely, queen and hearts are choices limited by the contents of the pack of cards rather than the language system and are thus part of an existential paradigm as opposed to a general paradigm. The choice of prominence in naturally-occurring spoken discourse is made when the speaker chooses from the existential paradigm that is available at that point in the discourse. In other words, making a word prominent denotes that the word is perceived by the speaker to be more situationally informative at that specific point in the context of the discourse, whereas non-prominence is for those words for which there is no alternative in the context. It needs to be added that not every syllable in a word has to be made prominent for the word to have the status of prominence in a tone unit. In continuous speech, speakers divide up their utterances into tone units. In discourse intonation, a tone unit contains at least one prominent syllable and the last prominent syllable (the tonic syllable) is the site of a pitch movement or tone.

Brazil (1997:18) states that there is a distinction between what is termed ‘word accent’ and the choice available to speakers to superimpose upon ‘accented’ syllables prominence or non-prominence. This is an important distinction and underlines the fact that it is the speaker who decides in what way her/his intonation choices ‘project a certain context of interaction’ (Brazil 1997:27). Thus the choices
with regard to prominent syllables are not reactive but rather proactive choices by speakers from known alternatives.

4. Data

The data in this study is a specialised corpus of job placement interviews. This corpus was extracted from the business sub-corpus in the HKCSE. The HKCSE is a two-million word corpus primarily made up of naturally occurring spoken discourse between Hong Kong Chinese and native speakers of English. In the HKCSE, there are four sub-corpora of equal size: academic discourses, business discourses, conversations and public discourses. The business sub-corpus is made up of a variety of discourse types such as meetings, presentations, interviews and service encounters, and in this study eleven hotel job placement interviews, amounting to 216 minutes of spoken data, were studied.

These interviews were held in five star hotels in Hong Kong and involved one or two hotel managers interviewing students applying to be placed in the hotel concerned for four months as a compulsory part of their BA (Hons) Hotel Management programme of the Hong Kong Polytechnic University. The purpose of the placement was for the students to acquire industry-based practical experience in relevant operational areas. The academic department provided assistance in soliciting suitable placements.

The job placement interview is a high stakes discourse event because the hotels are not obliged to accept applicants and the hotels have strict acceptance criteria and, in terms of the interviewees, this placement forms an important part of their CVs upon graduation and, if successful, can lead to a full-time offer of employment from the hotel. A critical part of the placement interview is the discussion of where the interviewee wishes to be placed in the hotel versus where the hotel would like to place the interviewee. This aspect of the interview is considered important from the point of view of the interviewer because the hotel has specific organisational needs to meet. It is also important for the interviewees because they often have specific career paths in mind and so wish to gain experience in a particular area of the hotel’s work.

5. Cohesive lexical chains

Lexical chains in the specialised corpus of hotel job placement interviews were identified by establishing the most frequent lexical words, by means of generating
a word frequency list and associated word lists by means of iConc®. Because the ‘basis of language is lexis’ (Lewis 1993:133), the lexical words that have high frequencies in the specialised corpus form a valid starting point. It needs to be acknowledged, however, that studies in phraseology (Sinclair 2005) show that it is often grammatical words rather than lexical words that provide the all important identification of meanings in phrase patterns. Table 1 lists the ten most frequently occurring lexical words in the corpus in rank order in relation to the overall word list. These words can be confidently predicted to be the source of lexical chains both within and across the job placement interviews in the specialised corpus.

Table 1. Ten most frequent lexical words

<table>
<thead>
<tr>
<th>Word</th>
<th>Rank order</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>hotel</td>
<td>25</td>
<td>281</td>
<td>0.62</td>
</tr>
<tr>
<td>front</td>
<td>45</td>
<td>170</td>
<td>0.37</td>
</tr>
<tr>
<td>department</td>
<td>47</td>
<td>159</td>
<td>0.35</td>
</tr>
<tr>
<td>office</td>
<td>48</td>
<td>154</td>
<td>0.34</td>
</tr>
<tr>
<td>work</td>
<td>62</td>
<td>123</td>
<td>0.27</td>
</tr>
<tr>
<td>guest</td>
<td>70</td>
<td>106</td>
<td>0.23</td>
</tr>
<tr>
<td>industry</td>
<td>77</td>
<td>99</td>
<td>0.22</td>
</tr>
<tr>
<td>learn</td>
<td>78</td>
<td>99</td>
<td>0.22</td>
</tr>
<tr>
<td>food</td>
<td>60</td>
<td>41</td>
<td>0.23</td>
</tr>
<tr>
<td>training</td>
<td>90</td>
<td>80</td>
<td>0.18</td>
</tr>
</tbody>
</table>

The top two words from the wordlist — hotel and front — were selected for this study. The word hotel is ranked 25 with a frequency of 281 (0.62%), and front is ranked 45 with a frequency of 170 (0.37%). A two-word association list was then generated. The combination of lexical words, front office, was selected because of the very high frequency at 142. In the instances of front office, 92.2% of the instances of office are in combination with front; and 83.5% of the instances of front co-occur with office. In addition, in the word frequency list, office is ranked 48 with a frequency of 154 (0.34%). The reason for front and office occurring so frequently in combination is that the term is viewed as a preferred area of work in hotels (i.e. as opposed to the back office). This is because working in the front office may include working at the front desk which brings hotel employees into direct contact with hotel guests, and this is often the most popular placement experience for the interviewees.

After key lexical words were established, the corpus was further examined through concordances of the above selected words to further contextualise and describe features of the interviews. Individual job placement interviews were then qualitatively examined to analyse the role played by discourse intonation in
creating additional layers of meaning to lexical cohesive links, and particularly the relation between lexical and non-lexical cohesion and prominence.

6. The communicative role of prominence in a cohesive chain

When discourse intonation is taken into account, speaker decisions within the prominence system are made on the basis of the speaker considering the status of individual words (Brazil 1997:39). The choice between making lexical or non-lexical words prominent is linked to what speakers determine to be situationally informative at that point in the discourse (Brazil 1997). The other three systems in discourse intonation, tone, key and termination, are not attributes of individual words but of the tonic segment (i.e. that section of the tone unit that falls between the first and the last prominent syllable) and are not considered here.

The communicative role played by prominence in cohesive chains is illustrated in the following examples (see transcription conventions at the end of the paper). Example 1 begins when the interviewer is talking to the interviewee about the long hours faced by many employees in the hotel industry.

Example 1

B: interviewer  a: interviewee
1. B: // and long hours in // OUR INDUSTRY // is very easy // for one simple
2. reason // is that the hotel never closes //
3. a: // mm //
4. B: // so // you can basically be here // twenty four hours a day //
5. a: // mm //
6. B: // it's not like another shop // or bank // it closes // so you can't // (.) // but
7. // basically we are open all the time //
8. a: // mm //
9. B: // so // I think // in OUR INDUSTRY // you // when you're committed //
10. a: // mm //
(HKCSE: B062)

In Example 1, the interviewer on lines 1–2 and 9 selects prominence on ‘our industry’ and ‘hotel’, and this has the effect of the speaker denoting these words to be situationally informative at that point in the unfolding discourse. This is not the case for the other words in this cohesive chain, here, it and we on lines 4, 6 and 7 respectively. These words, which in this context refer to either ‘hotel’ or ‘our industry’, are non-prominent which means that in the specific contexts in which they occur, the speaker deems them not to be situationally informative. What is interesting here is that it is both lexical and non-lexical words that have been chosen to be made prominent and non-lexical words that are chosen to be
non-prominent. While this is a common enough phenomenon, given that Brazil (1997) emphasizes that prominence is not a fixed property of the word itself, this suggests that there might be a connection between a speaker choice between a lexical or non-lexical word in a cohesive chain and her/his perception as to what is currently situationally informative. In other words, the decision as to whether to use a lexical or non-lexical word in a cohesive chain might be partly driven by a speaker’s decision as to what to project as being situationally informative at that particular point in the unfolding discourse.

In Example 2 below, we see this phenomenon repeated. The transcription is marked up in the same way and the extract begins with the interviewer asking why the interviewee wishes to work at that particular hotel.

Example 2

B: interviewer  a: interviewee
1. B: // very well // er // now // I’d like to ask you er // my second part of the
2. question which was // what you’re interested in HYATT // (.) // or is there
3. any // special reason // or you just dropped off in the MTR // and you saw
4. HYATT // oh // [let me walk in HERE // ((laugh))]
5. a:  
6. a: // um // (.) // actually // um // I’m being assigned // to work in the // (.) //
7. HYATT // (.) // HYATT REgency //
8. B: // oh you’ve been assigned // (.) // [it’s not your choice //
9. a: [ // yes // ((laugh))]
10. a: // yes // er // but er // when I know that // I was being assigned // to //
11. er // to apply for the // internship trainee // to work in the hyatt REgency //
12. I’m quite happy // cos um // ((laugh)) // it’s location // is very con- //
13. convenient // it’s just near the // MTR station // and um // I just er // come
14. here // go HERE // to um // by MTR //
15. B: // uhuh //
16. a: // and um // also // um // I know that// er from the brochure // of the hyatt
17. REgency // I know the // I know that // er // THIS hotel // always //
18. emphasize // on the staff training // and they treat their staff // not like
19. their // (.) // employee // just like their guest // (.) // and // they have a
20. good // and er well organised // training programmes // for the staff //
21. therefore I’m quite happy to be assigned to // to work HERE //</HKCSE: B062>

In Example 2, there is a cohesive chain made up of both lexical and non-lexical words. The lexical words consist of ‘Hyatt’, ‘Hyatt Regency’ and ‘hotel’ on lines 2, 4, 7, 11, 16–17 respectively. The speakers also consistently choose to make these words prominent, perceiving them to be situationally informative at those points in the discourse. The exception is ‘this hotel’ (line 17) uttered by the interviewee where it is the demonstrative pronoun ‘this’ and not ‘hotel’ which is perceived
to be more situationally informative. In this context of interaction, the speaker emphasises that this particular hotel has very good staff training and so implies that other hotels do not. In the case of the non-lexical words in the cohesive chain (eleven in all) — ‘here’ (4), ‘it’ (3), ‘they’ (2), ‘their’ (2) — only three are made prominent by the speaker, and in each case, it is ‘here’ (lines 4, 14 and 21) that is perceived to be situationally informative. A possible explanation is that at this point in the interview, the focus is on why the interviewee wishes to work here in this particular hotel and not some other hotel. The other non-lexical words are non-prominent, and so are not perceived by the speaker to be situationally informative in their local contexts. The general pattern for selecting prominence in Example 2 provides more evidence that, generally speaking, there is a connection between a speaker’s choice of lexical or non-lexical words in a cohesive chain and what she/he perceives to be situationally informative in the unfolding talk, but also that both lexical and non-lexical words can be made prominent or non-prominent by speakers.

Example 3 shows the interviewer questioning the interviewee about her choice of university programme (BA (Hons) in Hotel and Tourism Management).

Example 3

a2: interviewer  a1: interviewee
1 a2: // I see // actually since when // did you have this intention // of joining the
2 hotel // I am sure // I mean // () // er // you don't started this // well I
3 mean after you have finished at form seven right //
4 a1: // mh[mm //
5 a2: // [should be a bit earlier // or // or // did you just choose it // when you
6 finished form seven //
7 a1: // um // I think I just choose it // er when I // () // need to // er make a cho-
8 // [make my choice to // to study //
9 a2: // [I see //
10 a2: // I see // () // do you // have any other choice // apart from hotel // or do
11 you have a first second // () // choice // something like this //
12 a1: // um // THIS course // is the third CHOICE // actually //
13 a2: // THIRD choice // oh //
(HKCSE: B061)

In Example 3 the interviewer is asking the interviewee whether she had a choice of programmes when she applied to university in terms of listing first and second choices. The interviewee replies that Hotel and Tourism Management ‘is the third choice’ and selects prominence on ‘choice’, which indicates that she considers ‘choice’ to be situationally more informative at this point in the discourse. For the interviewer, however, it can be seen that it is the fact that the interviewee placed Hotel and Tourism Management third that is situationally informative when she
responds, saying ‘third choice oh’ with prominence on ‘third’. This example serves to illustrate the way in which a speaker’s choice of prominence can impact the communicative value of words in a lexical cohesive chain, and is not determined by the word itself, but is determined by the speaker’s real-time decisions to project what is informative in a particular context.

In Example 4 below, another phenomenon related to the selection of prominence in cohesive chains by speakers is described. This might be termed ‘convergence’ (see Chun 2002:39–42, for other forms of intonation behaviour signalling convergence and divergence) which describes the way in which speakers who initially diverge in terms of which syllable is made prominent, when a choice is possible, tend to converge and adopt the same syllable to be made prominent as the cohesive chain unfolds.

**Example 4**

B: interviewer  a: interviewee

1 B: // okay // (. ) // R__ has told me a little bit // so far // I er // sorry // I should
2 introduce myself // I work in the front Office // I’m [the front Office
3 a: [ // mhm //
4 coordinator // [so you have the front Office manager // (. ) // [then you
5 a [ // mhm //
6 have the assistant // and then you have me // [afterwards okay // (. ) // R__
7 a: [okay
8 has told me that you’re interested // in working in the front Office //
9 a: // yes //
10 B: // is that right //
11 a: // yes //
12 B: // why would you be interested in // (. ) // coming to front Office //
13 a: // mm // because I think er FRONT office // department // is a very
14 // um // important role // in the // hotel um // in a hotel // because // um //
15 B: // mm mm //
16 the main function of a hotel // is to // sell rooms // to provide
17 accommodations // apart from // food and beverage // other services //
18 and um // (. ) // and I think er // it’s also important // because // it provides
19 information // to guests //
20 B: // mm [mm //
21 a: [ // and er answer guests’ // enquiries // mm // and because // it’s // (. )
22 // because of the challenging nature // of the // (. ) // work // and // also
23 because of my personality // 1 // (. ) // because I’m quite er service //
24 oriented // [person // um // (. ) // I like to serve people // and I gain
25 B: [ // mhm //
26 satisfaction // when I // deal with them and er // (. ) // solve the problems //
27 [mm
25 B: [why would you choose // front Office // over // (. ) // food and beverage //
In Example 4, the interviewee wishes to work in the front office of the hotel based at the front desk (i.e. reception area) and the interviewer is not keen to commit himself until he has ‘tested’ the interviewee. This divergence is realised in a number of ways but, interestingly, it is also realised through the choice of prominence chosen for ‘front office’ by the respective speakers. The interviewer chooses prominence on ‘office’ while the interviewee consistently chooses to make ‘front’ prominent. This serves to indicate what each speaker considers to be situationally informative in the discourse. The interviewer might be said to be exploiting his choice of prominence by downplaying the importance of where the interviewee may be placed to work in the hotel, while the interviewee does the opposite. On
lines 40–77, the interviewer tests the interviewee's ability to handle difficult guests in the context of a front desk worker. Each time that the interviewee responds satisfactorily, the interviewer makes the scenario increasingly difficult by introducing an additional complication. Finally, on lines 78–80, the interviewer appears to be satisfied that the interviewee is up to the task when he says *that's good I was I was pushing you there ((laugh)) okay* on lines 80. It is at this point that he converges with the candidate's choice of prominence and, for the first time on line 82, chooses to make 'front' and not 'office' prominent when he says *when you're working in FRONT office where which department do you think you want to start with first.* The interviewer can be seen to then continue with this choice of prominence for *FRONT office* on lines 86 and 88. It is interesting to note that at the point in the discourse when the interviewer's reticence with regards to the interviewee's suitability for the front office ends, his selection of prominence converges with that of the interviewee. This alignment in prominence choice may be heard as an additional indication that their views on the matter of where in the hotel the interviewee should be placed are now also aligned. The issue is resolved; the interviewee is to be placed in the front office in line with her preference and it is now simply a matter of which department she wishes to start in.

To illustrate that the convergence described in Example 4 is not an isolated incident, the same phenomenon is to be found in Example 5. It is important to point out that the example is taken from a different placement interview with different participants. Example 5 begins at a point in the interview where the interviewee is explaining why she would like to be placed in the front office of the hotel.

**Example 5**

b: interviewer    a: interviewee

1 a: // and so // er // rooms division // er // rooms division // management // can
2   // er help me to understand the operation // of the FRONT office // er the
3 b: // [mm //
4   housekeeping department // and also // um // I // I can know // the // some
5   er // some // the experience from our lecturers //
6 b: // mm //
7 a: // and to see er their // er // how to deal with the guest // when // the guest
8   have er // when the guest is demanding // [and so I think it's very
9 b: [mm
10 interesting for me //
11 b: // oh // I see // ve- // very good // er // do you know any // difficulty or
12 problems // () // as a receptionist // in the front Office // yeah // because
13 quite a lot of student // er // always told me // I like front Office the most
14 // I want to be a receptionist // after graduate // but er // do you know // the
15 difficulty // of working in a front Office //
16 a: // um // I think // working in the FRONT Office // um // () // it must
Cohesive chains and speakers' choice of prominence

17 patient // [because if the guest // um complain about something // and we
18 b: // [mm //
19 can't argue // with them // [and we must keep calm // and // and // tackle
20 b: // [mm //
21 the problem properly //
22 b: // mm //
23 a: // and to see whether the managers // can help us [to solve the problem //
24 b: // [mm //
25 a: // [and I think the most important // problem is // (.) // how to deal // with
26 b: // [mm //
27 the complaints // from the guest //
28 b: // mm mm hmm hmm // very good // (.) // mm // (.) // mm // er // one thing
29 er // I would like to remind you // (.) // is that // (.) // er as a trainee in this
30 hotel // er // my own expectation on trainee //
31 a: // mm //
32 b: //yeah // is human relationship // all the colleagues // in this hotel // er
33 because er // if you have a harmonious relationship with // everybody // if
34 no matter // in the F and B // front Office or housekeeping // so // er //
35 they will have // (.) // more initiative // to teach you // teach you the
36 things //
37 a: // mm //
38 b: // mm // okay do you have any question to ask //
39 a: // um // I would like to ask // in my internship // um // um // we // er I
40 usually // will go to // er // what department // which department // that I
41 will // receive the training //
42 b: // mm // basically // for half a year training // er //we'll arrange one
43 department // (.) // [in a month // let's say the first month // er I will
44 a: // mm //
45 arrange you // to F and B first // because er // handover period will be at
46 the end of er // June // therefore I would prefer you stay in F and B // er
47 especially // in banquet // (.) // service department //
48 a: // mm //
49 b: // yeah to see handover function // [I think this only one chance // for
50 a: // [mm //
51 everybody //er // to have this experience // [(.) and then afterwards // er //
52 a: // [mm //
53 I'll be I'll plan for the // schedule // let's say house keeping // FRONT
54 office // and // even // if // you would like er // I can arrange // to stay in the
55 office // let's say account accounting office // or personnel office // er
56 which department // you prefer // er // er spend // more time //
57 a: // erm // I think // FRONT Office //
58 b: // FRONT Office //
59 a: // yes //
In this second example of convergence, in the early part of the extract the two speakers differ in their choice of prominence with regard to ‘front office’. For the interviewee, who is very keen to work as a front desk receptionist in the front office, it is ‘front’ which she chooses to make situationally more informative in this context of interaction on lines 2 and 16. The interviewer, on the other hand, chooses prominence on ‘office’ (lines 12, 13, 15 and 34) because, as it becomes clear later in the extract, his plans do not match with the interviewee’s aspirations (i.e. to spend the placement at the front desk of the hotel). This is the same behaviour as that observed in Example 4. When the topic of where the interviewee is to be placed is raised by her on lines 74–76, the interviewer finally makes his position clear on lines 78 and 80, saying I will arrange you to F and B first (F&B is Food and Beverage Department), but he then offers a glimmer of hope that placement in the front office is one of a number of possibilities after spending time in F&B. When he says FRONT office on lines 88–89, he now converges with the interviewee’s choice of prominence and makes ‘front’ prominent, and therefore more situationally informative. The interviewer asks which of the possible departments the interviewee would like to work in after she has completed her time in F&B on line 91, and she says FRONT Office on line 92 and this time selects prominence on both words. This choice of prominence is then selected by the interviewer on line 93, which again demonstrates convergence between the two speakers. On line 95, the interviewer asks how long the interviewee would prefer to work in the front office, and his choice of prominence is on ‘front’ not ‘office’. This choice of prominence is made again by the interviewee on line 98 when she replies to the question.

The form of convergence discussed above has not been described by others, but Chun (2002:39–42) discusses the ways a number of studies have shown speakers may use other forms of converging/diverging intonation behaviours to promote co-operation between themselves. For example, Yang (1995) describes speakers’ pitch movements mirroring each other across utterances. Similarly, Brazil (1997)
describes the prevalence of pitch concord, whereby the choice of pitch at the end of one speaker’s utterance is echoed by the next speaker at the start of his/her utterance. Also, Couper-Kuhlen and Selting (1996:46) describe the exploitation by speakers of pitch, timing and rhythm between speakers to either converge or diverge to indicate their orientation to the other participants.

7. Conclusions and pedagogical implications

In this study, a specific discourse type, job placement interviews, was examined, and since the corpus used has been prosodically transcribed, it proved possible to conduct an investigation into the relation between lexical and non-lexical cohesion and the communicative role of prominence, a system within discourse intonation (Brazil 1997). Several preliminary conclusions are drawn. First, lexical words are more likely to attract prominence than non-lexical words; the opposite was the case when speakers choose to use non-lexical words in a cohesive chain, and, importantly, this suggests that speakers do not randomly select lexical or non-lexical words in a lexical chain. Speakers may be guided in their choice between lexical and grammatical words, at least in part, by the decisions they make at the same time within the system of prominence in discourse intonation in terms of determining what is situationally informative at that point in the discourse. In addition to this observation, it was found that, while the lexical word being made prominent is a pattern widely found in the cohesive chains, in certain contexts of interaction the pattern may be reversed by speakers, and so lexical words are made non-prominent and non-lexical words made prominent in a chain. This further supports the view that prominence is not an inherent property of words, but subject to context-specific speaker choices. Another conclusion is that when two speakers choose prominence differently, in the sense of selecting different syllables in the same lexical word, or a different word in a fixed combination of words, as prominent, there is a tendency towards converging and adopting the same practice, although it is found that such convergence might be delayed, or withheld, when the speakers disagree in terms of the subject under discussion.

Another conclusion is concerned with our understanding about how prominence is allocated to a word by a speaker. Instead of only focusing on a speaker’s choice of prominence in distinct lexical or non-lexical words, the study analyses prominence choices in lexical or non-lexical words in a cohesive chain, and has found a connection between a speaker’s choice of a lexical word in a cohesive chain and his/her perception of situational informativeness at the moment of speaking, which can explain the reason for these patterns. This conclusion also constitutes
a partial answer to Hoey’s (1991:17) question: ‘under what circumstances do we use one (cohesive device) rather than the other?’ A speaker’s choice of prominence is not determined by the word class, i.e. lexical or grammatical words, but by the speaker’s decision to project what is more important or relevant in a particular context of interaction, which then impacts the choice of prominence and the choice of a lexical or a grammatical word.

Finally, this study has dealt with the interactional aspect of lexical cohesion by examining speakers taking turns and managing the interaction. It shows the way in which the speakers negotiate meanings through their choice of words and their decisions about making certain words prominent and others non-prominent ‘in a systematic and patterned way for interactive purposes’ (McCarthy 1988:181–182).

The findings in this study have potentially important implications for English language learning and teaching. Currently in the upper secondary schools in Hong Kong at least, discourse intonation has yet to be given attention in the teaching materials, despite studies that have advocated or investigated this area, including Brazil, Coulthard and Johns (1980), Hewings (1986), McCarthy (1988), Cauldwell (2002), Cheng (2004a, 2004b, 2004c), Cheng and Warren (2003), Warren (2004), Cheng, Greaves and Warren (2005), let alone the role played by speakers’ intonation choices as they navigate between lexical and non-lexical words in cohesive chains. The study of discourse intonation should become a staple part of English learning and teaching, especially to intermediate and advanced students. Corpora such as the one examined in this paper could well serve as the basis for learning and teaching materials and offer learners the opportunity for both the quantitative and qualitative study of discourse intonation in real world contexts.

Transcription conventions

Tone unit boundaries: //
A very slight pause: (.)
Prominence: upper case letters
Speaker identification: female indicated by A and a; male by B and b; Hong Kong Chinese indicated by lower case letters and native English speakers by upper case letters

Acknowledgment

The work described in this paper was substantially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. A-PF50). I am grateful for the helpful comments of the editors and two anonymous reviewers on earlier versions of this paper.
Notes

1. iConc® is designed and implemented by Chris Greaves, Senior Project Fellow, English Department, The Hong Kong Polytechnic University.

References


Describing the extended meanings of lexical cohesion in a corpus of SARS spoken discourse

Winnie Cheng
The Hong Kong Polytechnic University

The study reported in this paper applies Sinclair's (2004) descriptive model of lexical items, which consists of five categories of co-selection: two obligatory categories the core and semantic prosody, and three optional categories collocation, colligation and semantic preference. The study examines a selection of spoken discourse events collected in Hong Kong during and in the immediate aftermath of the SARS crisis in 2003. These discourse events form part of the Hong Kong Corpus of Spoken English (HKCSE). The findings show that once the overlapping patterns of co-selection of the most frequently occurring lexical words in the SARS corpus have been determined, it is possible to describe the cumulative effects of the habitual co-selection in the lexical items that contribute to textual meanings and coherence within and across the texts. It is argued that patterns of co-selection provide a fuller picture of textual and intertextual coherence than concentrating solely on lexical cohesion.

Keywords: colligation, collocation, core, extended units of meaning, Hong Kong Corpus of Spoken English, lexical cohesion, lexical item, semantic preference, semantic prosody

1. Introduction

Severe acute respiratory syndrome (SARS) is a viral respiratory illness that was recognized as a global threat in March 2003. SARS originated in southern China and was carried into Hong Kong Special Administrative Region (Hong Kong), and then on to Hanoi, Toronto, and Singapore in late February. SARS is primarily caused by a previously unrecognized corona virus, called SARS-associated corona virus (SARS-CoV). During November 2002 through July 2003, a total of 8,098
people worldwide became sick with severe acute respiratory syndrome that was accompanied by either pneumonia or respiratory distress syndrome (probable cases), according to the World Health Organization (WHO). Of these, altogether 774 died, 299 in Hong Kong. On 23 June 2003, the WHO removed Hong Kong from its list of areas with recent local transmission of SARS (http://www.cdc.gov/ncidod/sars/faq.htm). By late July 2003, no new cases were being reported, and WHO declared the global outbreak to be over.

The present study takes as its starting point that lexical cohesion contributes in part to textual and intertextual meanings and coherence in a shared environment. The study identifies forms of lexical cohesion and then proceeds to discuss patterns of co-selection among words in lexical items, based on Sinclair’s (2004:148) remark that ‘the word is not the best starting-point for a description of meaning, because meaning arises from words in particular combinations’. Sinclair’s notion of ‘lexical item’ is ‘characteristically phrasal although it can be realized in a single word’ (ibid.:122), and he observes that ‘all complete lexical items realize an element of meaning which is a function of the item in its context and context’ (ibid.:121). A similar discussion about ‘text and environment’ in the construction of extended units of meanings can be found in Tognini-Bonelli (2001:101–130). Meaning-making in text is a product of multiple occurrences of the same or semantically related items within and across texts; in other words, the lexical items that recur within and across different texts cumulatively impact textual and intertextual meanings.

The present study uses the theoretical concept ‘lexical item’ (Sinclair 2004:148) as a way of analysing text. It aims to show that it is the overlapping patterns of co-occurrence captured by a discussion and analysis using the ‘lexical item’ as a category that are contributing to the coherence of the text (Stubbs 2001:109), i.e. the overlapping of co-occurrence patterns is cohesive. The study examines a collection of extended naturally occurring SARS-related spoken public texts in Hong Kong in order to illuminate topical- and context-specific meanings related to SARS through the recurrent patterns of co-selection within and across the public texts due to the speakers’ choice of lexical items in their speech. The study also investigates the SARS-related public discourse as a possible site of extended units of meanings which are short-lived.

This paper describes an initial quantitative corpus-driven investigation of the discourse of SARS health-related speech events which then leads to the exploration of the data in more qualitative ways. The paper also shows how such a combination of quantitative and qualitative research approaches can offer more detailed language descriptions which can have useful pedagogical implications.
2. A model for describing the lexical item

Sinclair (1996, 1998, 2004) postulates an abstract model of lexical description that accounts for the internal structure of a lexical item (Sinclair 2004:141–8). The model comprises ‘five categories of co-selection’, two obligatory and three optional, which are put forward as components of a lexical item. The lexical item ‘reconciles the paradigmatic and syntagmatic dimensions of choice at each choice point’ (ibid.:141), using the five descriptive categories to describe both dimensions (ibid.:148). The first obligatory category is the ‘core’, which is ‘invariable, and constitutes the evidence of the occurrence of the item as a whole’ (ibid.:141). The other obligatory category is the ‘semantic prosody’. It is ‘the determiner of the meaning of the whole’ lexical item, expresses the ‘function’ of the lexical item, and shows ‘how the rest of the item is to be interpreted functionally’ (ibid.:141). A word may be said to have a particular semantic prosody if it can be shown to co-occur typically with other words that belong to a particular semantic set and display ‘a subtle element of attitudinal, often pragmatic meaning’ (ibid.:145). The three optional categories are collocation (Firth 1935, 1957), colligation (Firth 1935, 1957) and semantic preference. Collocation refers to ‘the co-occurrence of words with no more than four intervening words’ (Sinclair 2004:141). Colligation is ‘co-occurrence of grammatical choices’, and semantic preference is ‘the restriction of regular co-occurrence to items which share a semantic feature, e.g. about sport or suffering’ (ibid.:141). The three optional categories ‘realize co-ordinated secondary choices within the item, fine-tuning the meaning and giving semantic cohesion to the text as a whole’ (ibid.:141).

The five categories of co-selection represent a hierarchy of units of different sizes sharing the realization of meaning. A lexical item, for Sinclair, is the realization of the five categories of co-selection, with an identifiable and invariant core. From the point of view of the speaker/writer, the textual process of the five categories of co-selection (Sinclair 2004:34) is that first the speaker/writer selects a prosody of \( x \) applied to a semantic preference \( y \). The semantic preference in turn controls the collocational and colligational patterns, and the final component of the lexical item is the invariable core. Thus Sinclair (2004:34) argues that ‘the initial choice of semantic prosody is the functional choice which links meaning to purpose; all subsequent choices within the lexical item relate back to the prosody’.

As an illustration, Sinclair (2004:35–36) examines the collocation *true feelings* found in the Bank of English. He concludes that the collocation *true feelings* is the core of a compound lexical item which has the following inherent components: a semantic prosody of ‘reluctance/inability’, a semantic preference of ‘expression’ (and a strong colligation of a verb with the semantic preference), and a colligating possessive adjective.
This approach to characterising extended units of meaning is shared by Stubbs, who states that linguistics lacks ‘a descriptive theory of meaning’ (Stubbs 2001:62–63) and cites Sinclair’s (1996) call for ‘extended units of meaning’, based on Sinclair’s hypothesis that most units of meaning are phrasal. Stubbs (2001:64–66) cites Sinclair’s (1996) four types of co-occurrence, namely collocation, colliagination, semantic preference and discourse prosody. These are four of the ‘five categories of co-selection’ (Sinclair 1996, 2004), the fifth being the identification of the ‘core’. Stubbs (2001:65–66) prefers the term ‘discourse prosody’ to semantic prosody (citing Tognini-Bonelli (1996:193 and 209) as the source of this term) as it emphasises the creation of discourse coherence. Unlike Sinclair’s (1996, 2004) view of semantic prosody as obligatory, Stubbs (2001) views discourse prosody as optional. With regard to the borderline between discourse prosody and semantic preference, Stubbs suggests that it is not always clear cut. For example, the lemma ‘cause’ has a ‘strong negative prosody and there are relations of semantic preference between the verb cause and sets of abstract nouns such as illnesses and personal feelings’ (Stubbs 2001:65–66). Sinclair (2004:32) also notes that on occasion the semantic preference and semantic prosody may be ‘fused’ as a result of the speaker’s choice of co-selected words, for instance, the adjective ‘invisible’ in the lexical item ‘invisible to the naked eye’ is an example of a semantic preference of visibility and a semantic prosody of difficulty that are ‘fused’ in the same word.

Since semantic prosody is the critical, functional choice, the use of it in this study needs to be clarified. The difference between connotation and semantic prosody has been discussed. Partington (1998:66) describes ‘expressive connotations’ which involve the speaker or writer selecting a lexical item, and thus implying a ‘favourable or unfavourable evaluation’ of what is under discussion. He argues that one ‘particularly subtle and interesting aspect of expressive connotation which can be highlighted by corpus data’ is termed ‘semantic prosody’ which refers to the spreading of connotational meaning beyond the boundaries of single words. Thus a positive or negative connotation is not to be found in a single item, but in its association with collocates. Stubbs (2001:197–219) examines words, phrases and connotations, and he uses ‘connotation’ to describe the speaker’s point of view, attitude and evaluation. Sinclair (2004:121) notes that semantics distinguishes between two kinds of word meaning: denotative and connotative. He offers the term ‘semantic prosody’ rather than ‘connotation’ as connotation is word-based, whereas semantic prosody ranges over the whole lexical item and encompasses the effect of employing a particular lexical item. Given the differing views relating to semantic prosody (see, for example, Whitsitt 2005), it is important to make clear that this study adopts Sinclair’s five categories of co-selection as the framework for analysing extended meanings of words, which further supports the importance of
lexical cohesion in understanding textual coherence. Thus the term semantic prosody is employed here as it better captures the sense of meaning being determined by patterns of co-selection among the same or semantically related words, rather than simply the individual words. The semantic prosody functions to integrate the item with its surroundings (Sinclair 2004:34).

3. The SARS specialised corpus

Spoken corpora have been and remain far rarer than written, largely due to the difficulties in obtaining and transcribing them, but corpora of native speaker varieties of English (most noticeably, British English and American English) have been analyzed (see, for example, Aijmer 1996; Biber, Conrad & Reppen 1999; Carter & McCarthy 1997; Sinclair 2001). A corpus allows a combination of quantitative and qualitative analysis, and new patterning is much easier to detect through quantitative analysis than by qualitative analysis alone.

The data examined in this study is obtained from the Hong Kong Corpus of Spoken English (HKCSE) (Cheng, Greaves & Warren 2005). It contains nine public speech events recorded from April to June 2003 in Hong Kong. The length of the speech events ranges from 26 minutes to 30 seconds, totalling 116 minutes and 17,898 words of recorded data. The SARS data set comprises a range of genres including public speeches, Q&A sessions, radio announcements and forum discussions, and the speakers are medical professionals, businessmen and politicians. For instance, the ‘Business After SARS Conference’ was organised by ‘Operation Unite’ which was an action-oriented campaign to draw the Hong Kong community together to fight against the challenges posed by SARS in 2003. The luncheon speech followed by a Q&A on ‘Health Care Cooperation’ was organised by the Hong Kong General Chamber of Commerce to discuss what role the private sector can play in the development of Hong Kong’s health care industries across the border.

4. Analysis of multiple patterns of co-selection

To identify the key lexical words associated with the SARS corpus, ConcGram© 1 was used to generate a word frequency list. Table 1 lists the ten most frequently occurring lexical words, which are the likely source of lexical cohesion both within and across the texts in the corpus, and may also be predicted to be wholly, or part of, the core of the lexical item.
Table 1. The top ten most frequently occurring lexical words in the SARS corpus

<table>
<thead>
<tr>
<th>Word</th>
<th>Rank order</th>
<th>Frequency</th>
<th>Percentage of lemmas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong Kong</td>
<td>16</td>
<td>140</td>
<td>0.77</td>
</tr>
<tr>
<td>health</td>
<td>23</td>
<td>97</td>
<td>0.53</td>
</tr>
<tr>
<td>care</td>
<td>34</td>
<td>67</td>
<td>0.37</td>
</tr>
<tr>
<td>hospital</td>
<td>35</td>
<td>65</td>
<td>0.36</td>
</tr>
<tr>
<td>private</td>
<td>41</td>
<td>59</td>
<td>0.33</td>
</tr>
<tr>
<td>SARS</td>
<td>46</td>
<td>51</td>
<td>0.28</td>
</tr>
<tr>
<td>people</td>
<td>53</td>
<td>44</td>
<td>0.24</td>
</tr>
<tr>
<td>sector</td>
<td>59</td>
<td>41</td>
<td>0.23</td>
</tr>
<tr>
<td>virus</td>
<td>60</td>
<td>41</td>
<td>0.23</td>
</tr>
<tr>
<td>China</td>
<td>67</td>
<td>36</td>
<td>0.20</td>
</tr>
</tbody>
</table>

After identifying the potential sources of lexical cohesion in the SARS corpus through generating the word frequency list, the corpus is further examined through concordances of the most frequent lexical words as the starting point to describe the lexical items in the context of the Hong Kong SARS discourse. Sinclair’s five categories of co-selection form the systematic basis for the description of the lexical items. Identifying the various interrelated forms and patterns of co-selection, along both syntagmatic and paradigmatic dimensions, in the most frequently occurring lexical words in the SARS corpus would help to describe lexical themes within and across the discourse events and add to our understanding of the extent of lexical cohesion present. In other words, while this study focuses on analysis of particular ‘lexical items’ (Sinclair 2004:148), it does so with the aim of revealing the overall textual relationships, meanings and coherence in the corpus.

The study begins by examining the concordance lines containing the six most frequent lexical words, namely Hong Kong, health, care, hospital, private and SARS. After that, virus is also studied in conjunction with SARS.

4.1 Hong Kong

With 140 instances each of Hong and Kong in the SARS corpus, the proper noun Hong Kong is identified as the most frequent. Given that all of the discourse events took place in Hong Kong and the focus of the talk was to discuss the medical, health-related administrative and economic issues in the immediate aftermath of SARS being brought under control, the prevalence of Hong Kong is not surprising. Figure 1 shows sample concordance lines for Hong Kong.
for most of us this is an opportunity of a lifetime. any challenges and overcome any adversities and strengths the strategic position and the unique role of each task often appeared to be daunting but we knew that cosmopolitan enterprise and connected so is pro-business environment unmatched in Asia and Hong Kong continues to maintain a clear advantage. SARS has caused we are in fact telling the world that Hong Kong will emerge stronger. Not only just strength the strategic position and the unique role of Hong Kong cannot be replaced over the years our competitive advantage the premier international business cent Hong Kong cannot be replaced, it has unique competitive advantage and Hong Kong has had many unique competitive advantages. This is the success of Hong Kong, this is the Hong Kong spirit that we all cherish. Thank you very much ladies and gentlemen.

In the last three months Hong Kong has taken on enormous challenge of fighting SARS and we are in fact telling the world that Hong Kong is ready to take off again. I believe that Hong Kong now must harness this energy and move forward the other positive outcomes from the epidemic are beginning to pick up once again. We have recovered from the trade mark. Ours is a resilient economy. I would tell everyone around the world we are made in Hong Kong and we are proud of it. Thank you very much.

**Figure 1.** Sample concordance lines for *Hong Kong*

Most of the uses of *Hong Kong* (65.6%, 118 out of 140 times) are speakers talking of *Hong Kong* as if it is a living entity (e.g., *Hong Kong has faced many crises*), and it is this usage which is analysed here to describe the five categories of co-selection. The other 22 instances of *Hong Kong* are speakers talking of *Hong Kong* purely as a geographical location (e.g., *Standard Chartered Bank has been in Hong Kong for one hundred and fifty years*), or as part of the name of an organisation (e.g., *Hong Kong University*), and they are not analysed here.

*Hong Kong* in the SARS corpus has no strong patterns of collocation or colligation. Collocation and colligation are viewed as optional in a lexical item (Sinclair 2004:161) and this is observed here. At N-1 position, *Hong Kong* is preceded by a preposition (52 times), the paradigmatic choices of which include *in* (26 times), *of* (12 times), *to* (7 times) and so on. Syntagmatically, *Hong Kong* is often spoken in association with phrases such as *cost of doing business, pro-business environment, unique competitive advantages, absorb the volatility, new and exciting opportunities, a Closer Economic Partnership Agreement, boost to tourism, and resilient economy*. This shows that *Hong Kong* in this corpus has the semantic preference of ‘business and economy’ in 72.3% (86 out of 118) of the instances. The other semantic preference is ‘medical field’ in 20.0% (24 out of 118) of the instances, and the associated phrases include *employment of medical nursing and other help, private hospitals, the next virus* and *public health authorities*.

*Hong Kong* in this corpus also has an interesting semantic prosody in the hands of the politicians, business and medical professionals. As exemplified partly in Figure 1, examples of words surrounding *Hong Kong* in the concordance lines are *emerge stronger, cannot be replaced, continues to maintain, bounced back, ready to take off again, exciting opportunities, the positive spirit, thrive again* and so on. The cumulative effect of such words, across 94.9% of the instances (112 out of 118), is to underscore the sense that *Hong Kong* is synonymous with an entrepreneurial ‘can do’ spirit and that *Hong Kong* is the eternal ‘comeback kid’. Whatever problems
Hong Kong goes through, deeply ingrained in the local culture is an abiding sense of ‘optimism’ that it can overcome any obstacle, and it is this optimism that is the chosen semantic prosody associated with Hong Kong. The function of Hong Kong, therefore, finds its place in a lexical item which is ‘optimism about recovery in business, economy and the medical field’. With respect to the core, because Hong Kong has no strong collocates, Hong Kong is the core of this lexical item.

It is worth noting at this point that while Hong Kong is rarely, if ever, criticised, or found wanting, in these discourse events, this is not the case with specific areas of activity in Hong Kong as is exemplified below.

4.2 health

The most frequently occurring lexical word after Hong Kong is health which is spoken 97 times in the SARS corpus, and some sample concordance lines are shown in Figure 2.

Figure 2. Sample concordance lines for health

Health has no strong collocates to the left, except for a tendency to collocate to the right at N+1 with care 33 times (34%). Health is colligationally followed by a noun at N+1 72 times (74%). The paradigmatic choices include care 33 times and other nouns related to an area of work or activity 39 times: service(s) (5 times), management (5 times), authorities (3 times), organisation (3 times), departments, workers, services, administration and so on.

The semantic preference associated with health in the SARS corpus is ‘institutionalised health provision’, and this strong preference is found in 97.0% (94 out of 97) of the instances. Regarding the semantic prosody, 69.0% (67 out of 97) of instances carry the meaning of ‘a desired state of existence that has yet to be achieved’; in other words, ‘lacking’. This is because [health + noun] is often preceded or followed by phrases such as a good opportunity to develop, a good opportunity to provide, encourage development of, to improve and so on. It needs to be remembered when analysing the semantic prosody associated with health, and of other items later in the study, that these SARS discourse events took place during
and in the immediate aftermath of the terrible SARS crisis that really shook the people of Hong Kong and the institutions responsible for dealing with it. We are able to see in these events, genuine critical self-reflection on how to learn from the tragedy and to better prepare for such crisis in the future. Indeed reflections such as these led to the resignations of key officials and the drafting of proposals for institutional reforms. With respect to the ‘core’, which is what ‘constitutes the evidence of the occurrence of the item as a whole’ (Sinclair 2004:141), the analysis shows that health is the core of the lexical item as it has no typical collocates.

To conclude, analysis of health shows strong evidence for a lexical item which has a semantic prosody of ‘lacking’, in a semantic preference of ‘institutionalised health provision’ which is realized by colligation of a noun of efforts, work, activity, people and governmental and non-governmental agencies, with no typical collocations. There is a strong colligation with a noun (care, services, management, authorities, organizations, and so on) that follows the single word core, health.

4.3 care

The lexical word care occurs 67 times in the SARS corpus, but two instances are removed (e.g. would you care to comment). Among the 65 instances of care examined, 51.0% (33 times) of them collocate with health at N-1. In the 32 instances when not preceded by health, there are no other strong collocates for care, although there are primary care (6 times) and intensive care (3 times). Figure 3 shows some concordance lines for care.

Figure 3. Sample concordance lines for care

Care forms strong grammatical patterns with other words because it is typically preceded by a modifier (93.8%, 61 times) and ‘on occasion’ followed by a noun (38.5%, 25 times). There is a strong semantic preference of ‘institutionalised support of health provision’ in 98.5% of instances (64 out of 65). Examples are health care workers, health care services, primary care, elderly care, outpatient care and so on. This semantic preference has also been observed by Stubbs (2001:151–54) in his analysis of care in British English, and he states that care has changed over
the last one hundred years from predominantly personal uses to institutional uses with the growth of the welfare state, one of which is related to health services.

76.9% (50 out of 65) of the instances of care occur with words and phrases such as expanding, need, improve, modify, deficiency, shortage, provide more services, what are the challenges, also experience problems, opportunities and challenges, and so on, all of which serve to create the semantic prosody of ‘insufficiency’. It can be said that the semantic prosody associated with these instances in this SARS corpus is similar to that of the lexical item with health as its core, and provision of ‘care’ is perceived by speakers as something desirable but not yet fully realised, that is ‘insufficiency’. The similarity of the semantic prosodies associated with the lexical item with care as the core and that with health as the core demonstrates that in these texts, both institutionalised health services and institutionalised care provision in the public sector are portrayed by speakers as ‘insufficient’. Later in the paper, this will be shown to be interconnected with another lexical item with private as its core.

Since the lexical word care has no typical collocates, it alone is the core of the lexical item. It has been shown that the lexical word care is the core of a lexical item with the following inherent components: a semantic prosody of insufficiency, a semantic preference of institutionalised support of (health) care provision and a colligating modifier.

4.4 hospital(s)

The lexical word hospital is ranked 35th in terms of frequency and occurs 65 times in the SARS corpus. The frequency becomes 75 after adding ten instances of the plural form hospitals, but is reduced to 66 instances after excluding the use of hospital when it is part of a proper noun (e.g. Prince of Wales Hospital). See Figure 4 for some concordance lines for hospital(s).

Figure 4. Sample concordance lines for hospital(s)

The general noun hospital(s) contributes to lexical cohesion within and across the nine texts and also displays a number of patterns in the syntagmatic and
paradigmatic dimensions. First of all, there are no strong collocates to either the left or the right, with private hospital(s) occurring only 8 times and hospital services only 5 times. Nevertheless, unlike health examined earlier, hospital(s) has stronger colligational patterns to the left than to the right. It is found that hospital(s) is often (66.7%, 44 times) immediately preceded by a determiner, paradigmatically realised by the (30 times), a (7 times), this (6 times) and one instance of that. The determiner typically is found at N-1 (33 times) but is also found at N-2, where in all cases it appears before a modifier (11 times). To the right, colligational patterns are less strong and the tendency to be followed by a noun (16 times) is much less pronounced than is found in the case of health.

A semantic preference of ‘the future re-organisation and funding of hospitals’ is found in 86.3% (57 times) of instances rather than ‘the services provided by hospitals or the work that the hospitals carried out during the SARS outbreak’ (27.3%, 18 times). With respect to the semantic prosody, as illustrated partly in Figure 4, throughout the SARS corpus the various speakers overwhelmingly speak of hospitals within the context of increased privatisation of hospital provision, resultant profit making, and so on. In addition, in 86.3% (57 out of 66 instances) hospital(s) are to be encouraged, developed, provided and promoted, all of which have a semantic prosody of being ‘advantageous’.

The analysis has shown that the function of hospital(s) finds its place in a lexical item which is ‘hospital privatisation representing something mutually advantageous for both the business community, in terms of profit, and the wider community, in terms of greater choice of services’. Another component of the lexical item with hospital(s) as the core is a colligating determiner but only in 66.7%. Finally, since hospital has no fixed collocates, it is found to be the core of the lexical item, meaning that hospital ‘constitutes the evidence of the occurrence of the item as a whole’ (Sinclair 2004:141). Since hospital and hospitals constitute the core, there is an example of a core that includes grammatical inflection (plural) which is described by Sinclair (2003:173–4).

The lexical item with hospital as the core can be seen to be related to the lexical items with health and care as their cores, and also related to the following analysis.

4.5 private

The lexical word private is found among the top 100 words in a SARS corpus. Private is number 43 on the word frequency list with 59 instances. One of these instances Hong Kong Private Elderly Home Association is excluded because it is part of a proper noun while the rest have private as an adjective. It is interesting to see how a corpus-driven analysis is able to reveal why this lexical word is prominent
in this specialized corpus. *Private* has a strong tendency to collocate to the right at N+1 with *sector* (58.7%, 34 times), and to a lesser extent with *hospitals* (13.8%, 8 times) and *health care* (6.9%, 4 times) (see Figure 5).

In what way can we work in partnership with the private sector in terms of hospital care in terms of out
erader trend in health care is to encourage more private hospital development and in the light of er nt
to encourage more private resources to put in different services in private sector not against even outsourcing to the private sector even in during the yesterday’s meeti
en it could be er a share of investment or from the private sector as well as the government or it could be
d have read sending people and buying beds from private sector hospitals not just in Britain but in or
er ing time then also to promote the development of private sector when I was in UK or last year I heard
Kong three areas public sector academic sector and private sector first I talk about public sector er from e
evrybody every sector which start with government private sector professionals academic institutions
encourage private sector to promote development of ambulatory services reducing waiting time promote private provision of hospital services now this is a

**Figure 5.** Sample concordance lines *private*

*Private* has a clear pattern of collocation to the right as a modifier before a noun in all instances. The semantic preference here is the actual or potential contribution of the private sector to the provision of health-related services in Hong Kong in 94.8% of instances (55 out of 58). We find, for example, *private health facilities, private investments, more private hospital development, private resources* and so on. This semantic preference is clearly specific to this corpus. A look at the Bank of English Corpus shows that the private sector is typically juxtaposed with the ‘public sector/state/government’ but is not restricted to the provision of health-related services. In the SARS corpus, the frequency of occurrence of *public* is 28 times, and of these 12 instances explicitly link *public* and *private* and four instances are the name of a government body in mainland China.

In the previous analyses, when the semantic prosodies of lexical items with *health* and *care* as their respective cores are examined, a sense of ‘inadequacy’ is conveyed. Nevertheless, when the semantic prosody of *private sector* or *private* + noun is considered, the prosody does not convey the generally negative semantic prosodies of these other lexical items, but it is nonetheless related to the semantic prosody associated with the lexical item with *hospital* as its core. In the company of *private sector* or *private* + noun can be found *proven successful, live a better life, good opportunities for, the importance of, good to encourage, opportunities for collaboration, encourage more, promote the development of, and obligation to help*. In 87.9% of instances (51 out of 58) of *private*, the semantic prosody is ’preferred’ as the speakers in the discourse events involving politicians and business people present private interests as a panacea for the deficiencies in the present, health-related services in Hong Kong which are predominantly in the public sector. This semantic prosody is not surprising in the context of Hong Kong in which a strong commitment to capitalism and its free market economy is central to the extent of even
being written into Hong Kong’s mini-constitution. This point also relates back to the earlier discussion of the semantic prosody of *Hong Kong* and so provides further evidence of the way in which speakers build up intertextual coherence of their meanings and so reinforce their ideological positions. Here a semantic prosody of ‘preferredness’ is made in relation to the involvement of private interests in the provision of heath services in Hong Kong with *private* as the core of a compound lexical item always acting as a modifier.

4.6 SARS

SARS occurs 51 times. Figure 6 presents a sample of concordance lines for SARS. SARS does not display any strong collocates; it collocates with *epidemic* 8 times (15.7%) and *outbreak* 4 times (7.8%). SARS has colligational patterns of being preceded by the definite article (23.5%, 12 times) or a preposition (29.4%, 15 times). The semantic preference is typically (76.5%, 39 times) to do with the impact of SARS on Hong Kong and those directly affected by it, with phrases such as *turmoil, damage, disease, crisis, infected area, patient, epidemic, outbreak,* and *morbidity.*

SARS has provided an interesting example of the dynamics of semantic prosody, and is discussed below. The texts in the corpus are a mixture of texts produced during and after the SARS crisis in Hong Kong and this has led to two distinct semantic prosodies being selected by speakers. During the SARS crisis there are instances of *challenge of fighting, continue to battle, heavy pressure* and *severely affected by,* and hence a semantic prosody of ‘embattled/besieged’ in 66.7% (34 out of 51) of instances. Interestingly, the semantic preference and semantic prosody associated with SARS provide a good example of a phenomenon described by Sinclair (2004:35). Sinclair states that sometimes ‘the semantic preference and the semantic prosody are fused’ (for example, *invisible* has the semantic preference of visibility and the semantic prosody of difficulty), and this phenomenon is observed here in the cases of co-selections such as *epidemic, crisis* and *outbreak.* In certain co-selections, the semantic preference of ‘the impact of SARS’ fuses with the semantic prosody of a negative sense of ‘embattled/besieged.’ However, in the post-SARS texts a different semantic prosody is observed through the speakers’ choice of such phrases as *successfully brought (SARS epidemic) under control, positive outcome from the (SARS situation) and in spite of the (turmoil that SARS has caused)* which can be characterised as ‘positive assessment’ in 21.6% (11 out of 51) of instances. This semantic prosody also supports the semantic prosody of the lexical item with *Hong Kong* as its core, and demonstrates, again, the interconnectivity of lexical items across a corpus. It can be seen that meaning making and ideological positions are
reinforced across lexical items within and across texts in the corpus, creating cohesive links and textual and intertextual coherence. It can be concluded that during the SARS crisis, speakers chose the semantic prosody of 'embattled/besieged' in relation to the ongoing impact of SARS with SARS as the core, whereas in the post-SARS period, speakers chose a semantic prosody of 'positive assessment' in relation to the post-assessment of impact of SARS with SARS as the core.

4.7 virus

The lexical word virus is used 41 times and, interestingly, never in combination with SARS. It is used typically in the form of the (modifier) virus that refers to SARS. The strongest collocate is corona at N-1 (46.3%, 19 times) (see Figure 7 for examples). At N+1 less strong collocates are itself (12.2%, 5 times) and infection (7.3%, 3 times). The main patterns of colligation are the use of determiners at N-1 (31.7%, 13 times) and the + modifier + virus (41.5%, 17 times).

During the SARS outbreak, medical professionals were debating whether or not the SARS virus was a type of corona virus, and this explains the semantic preference found here which is ‘investigating virus typologies’ in 63.4% (26 out of 41) of instances. The word virus co-occurs with detected, found, learned, document, looking for, testing and so on. However, since the texts span both the time of the SARS crisis and the immediate post-SARS period, again two semantic prosodies are found as the medical experts' investigations led to greater certainty as to the identity of the virus. The semantic prosody associated with virus is sometimes the
sense of ‘unknown/uncertainty’ in 53.6% (22 out of 41) of instances as the speakers describe their investigations as inconclusive at that stage with expressions such as we don’t know, nothing was found, we still try to find and so on. It can be seen that in later uses, the medical researchers were sure that the SARS virus was a form of corona virus, and there is therefore a semantic prosody of ‘known/certainty’ in expressions such as I have no doubt and it’s very clear to me accounting for 31.7% (13 out of 41) of instances. It is also of interest to note that 97.6% (40 out of 41) of instances of virus are spoken by medical professionals, which confirms its semantic preference.

Semantic prosody, which is a dynamic property of lexical items, can be seen to be evolving here — today all medical experts agree that the SARS virus was indeed a type of corona virus. The analysis has shown that initially a semantic prosody of ‘unknown/uncertainty’ was chosen in relation to medical investigation of virus typologies with SARS as the core of a lexical item. In the post-SARS period the semantic prosody changed to ‘known/certain’ with the rest of the lexical item remaining intact.

5. Pedagogical implications

Despite the fact that lexical cohesion and overlapping patterns of co-selection are such prevalent phenomena in written and spoken discourse, they are rarely formally taught to learners of English. In Hong Kong, for instance, a survey of about 20 English course books used in the upper secondary forms has produced very little evidence of these aspects of language use being taught. Very often, in reading comprehension activities, advice is given that the meaning of vocabulary items needs to be understood in context. In one book, such advice is followed by the teaching of ‘-nyms’ words, with antonyms, synonyms and hyponyms presented together with homonyms, homophones and acronyms (Wong & Etherton 2000:70–71). In oral summaries and presentations, the course books advise the use of synonyms of phrasal verbs, verbs, nouns, adjectives (Leech 2002:50–54); general nouns instead of quoting ‘the exact names and numbers’ (Mau 2003:18); ‘alternatives which have the same meaning’ to replace words or phrases (Esser 1999:14); and ‘reword the content of the (reading) passage’ (Leech 2002:13). Other books suggest ways to make the writing and oral presentation ‘coherent and logical’ (Esser 2003:4; Li & Leech 2003:16) by means of ‘appropriate connectives’ (Leech 1998:13), ‘linking phrases’, ‘signal expressions’, or ‘transition words’ (Sutton 1999:42; Potter 2003a:17–18; Potter 2003b:24; Lee & Holzer 2003:45), as well as ‘the definite article, comparison, conjunctions, pronouns, substitution, subordination
and synonyms’ (Li & Leetch 2003:16). However, the emphasis tends to be on the surface non-lexical devices, and no definition of ‘coherent’ or any explanation about how the different devices contribute to coherence is provided. When discussing the importance of coherence of ideas, some books do include linking phrases, synonyms, ‘vocabulary families and key words’ with the warning that ‘just because two sentences contain the same words doesn't mean they are a pair’ (Li & Leetch 2003:22). The explanation is, however, very limited. This paper has shown that all of the above advice, while true, is insufficient preparation for learners. Corpus-driven studies of language emphasise the phraseological character of natural language, that is, the more-or-less fixed co-occurrence of linguistic elements (see, for example, Hunston 1995). The analysis of the semantic prosodies associated with the lexical items in a corpus is a way to acquire context knowledge which is of paramount importance for writers trying to master tasks within a specific genre (see, for example, Tribble 2000). It is, therefore, necessary for learners to understand the patterns of co-selections in context, in addition to the study of lexical cohesion. It would be useful if they understand that the same words, when occurring in different discourse events in the same genre or in different genres, are likely to create very different patterns of co-selection.

The pedagogic implications are that given the ease of access to commercially available corpora, or a self-compiled corpus, such as a learner corpus, it should not be an immense task for teachers to design and write materials to exemplify and teach these very useful concepts of co-selection.

6. Conclusions

The present study demonstrates a corpus-driven approach to describing lexical items (Sinclair 2004:148) across texts in a specialised corpus, by means of Sinclair’s (2004:141–148) five categories of co-selection. The description of the patterns of co-selection using the ‘lexical item’ as a category has been shown to be more productive by going beyond describing simple lexical cohesion. The study has demonstrated that it is possible to describe the semantic prosody associated with all of the lexical items which confirms Sinclair’s claim that prosody is always identifiable, that is, obligatory. This is contrary to others, for example, Stubbs (2001:106) who argues that semantic prosody is not always identifiable. This study provides evidence that semantic prosody is both dynamic and text or genre specific, as it has been shown that this is the case for a number of lexical items studied.

The advantage of working with a collection of specialised discourse events, such as the one studied here, is that the researcher is able to establish, by means of
the word frequency list, initial examples of lexical cohesion which run within and across the discourse events. It is then possible to describe the cumulative effects of any paradigmatic and syntagmatic patterns of co-selection on the co-construction of situated meanings and the establishment and maintenance of ideological positions. The study has shown that in this SARS corpus, the lexical items described become cumulatively and interconnectedly associated, for example, with an optimistic view of Hong Kong’s post-SARS future, while others become associated with a need for reform or further development, resulting from the identification of deficiencies, by the preferred means of involving private business interests. Sinclair (2005:7) outlines the possibility of an automated process to compare the lexical profiles or ‘aboutness’ of texts and to establish these as ‘partial rephrasings of each other’. This study has attempted a partially automated study of what Sinclair (2005) terms ‘document relativity’ and it has proved to be a significant addition to more traditional studies of lexical cohesion.

Public discourse such as this SARS corpus is possibly, in some cases, the site of extended units of meanings which are more short-lived. This is because, as Cheng (2004:48) points out, discourse in the public domain is more likely ‘to transmit short-term priorities or agendas, albeit reflecting underlying values, as opposed to longer term values and beliefs’. There is also the fact that the corpus examined is a specialised corpus which can have extended meanings which might not hold for a general corpus, and this is to be expected.

More research needs to be done with both specialised and reference corpora to better determine whether co-selection and its associated meanings require a combined approach, as this study suggests that within a larger corpus there may be sub-corpora with particular forms of lexical cohesion and patterns of co-selection that are not necessarily replicated across the main corpus.

As the analysis widens, we find lexical items, combining patterns of cohesion and co-selection, which in turn relate to other lexical items. The analysis can be seen to be revealing complex textual and intertextual relationships which together are part of the mechanism that creates surface lexical cohesion, but more fundamentally textual and intertextual meanings and coherence.

Acknowledgements

I am grateful for the helpful comments of the editors and two anonymous reviewers on earlier versions of this paper. Any fault with this paper is my own.

The work described in this paper was substantially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. B-Q714).
Note

1. ConcGram® is a search engine designed and implemented by Chris Greaves, Senior Project Fellow, Department of English, The Hong Kong Polytechnic University.

References


Use of signalling nouns in a learner corpus*

John Flowerdew
University of Leeds

Signalling nouns are nouns which have cohesive properties across and within clauses. A signalling noun is potentially any abstract noun the full meaning of which can only be made specific by reference to its context. Examples of nouns which can function as signalling nouns are attitude, assistance, difficulty, endurance, process, reason, result etc. Signalling nouns in discourse are closely associated with nominalisation and are problematic for learners. Based on a corpus of argumentative essays written by Cantonese L1 learners of English, this paper presents a taxonomy of error types and frequency data of the different error types in the use of signalling nouns. The paper then compares the average number of signalling nouns used per essay with grades awarded to the essays, on the one hand, and the numbers of signalling noun errors according to grades, on the other. In both cases there is a significant correlation. The findings confirm the intuitive idea that the use of signalling nouns adds to the overall coherence of a text.

Keywords: signalling noun, cohesion, lexical cohesion, coherence, learner corpus, learner error, error analysis

1. Introduction

The use of lexical cohesion, and within that, the use of what are referred to in this paper as signalling nouns (Flowerdew 2002, 2003a, 2003b), is a major challenge for writers, whether it be in their first or additional language(s). In the teaching of English as a second language, lexical cohesion has been neglected (for exceptions see Francis 1988; Thurstun & Candlin 1998), perhaps because there have been no good descriptions of the forms and functions of this phenomenon.

In their seminal book, Cohesion in English, Halliday and Hasan (1976) devote only a very short penultimate chapter to lexical cohesion, assigning the first part to what they refer to as general nouns, a superordinate term, which can be viewed
as incorporating signalling nouns. Since the publication of *Cohesion in English* a range of studies has focused on various aspects of signalling nouns (Winter 1977, 1992; Tadros 1985; Francis 1986; Ivanič 1991), but this research has only relatively recently been brought together in a systematic fashion (Schmid 2000; Flowerdew 2002, 2003a, 2003b).

Now that we have these more comprehensive accounts, as provided by Schmid and Flowerdew, ways can be developed for diagnosing learner problems in the use of signalling nouns. In the tradition of error analysis in general (Odlin 1989; Richards 1992) and learner corpora studies in particular (Granger 1998), this diagnosis can be applied to teaching and learning. The purpose of this paper is thus to analyze a corpus of learner writing, to develop a taxonomy of errors in the use of signalling nouns, and to provide some frequency data on the amount of use by learners of signalling nouns and the number of errors made in their use. The reason for the focus on signalling nouns is that they are particularly frequent in academic language, the register that has been most studied in learner corpus studies to date. In a study of a large corpus of academic English, consisting of research articles, textbook chapters and lectures, Flowerdew (in preparation) found that there was a significant difference in the frequency of use of signalling nouns across these different genres, with research articles making the most use of signalling nouns, followed by textbooks and then lectures. A comparison with spoken conversational data showed a much lower usage than in these academic genres. The appropriate use of signalling nouns can thus be seen as an important dimension in the development of academic literacy.

2. **Corpus-based error analysis**

Richards, Platt and Platt (1992:127–8) describe error analysis as developing in the 1960s with the goal of demonstrating that many learner errors were not due to the learners’ mother tongue, but to universals of second language acquisition. They also state that by the 1970s error analysis had largely been superseded by studies of interlanguage and second language acquisition. As Altenberg and Granger (2002:14) point out, error analysis has often been viewed negatively, “as retrograde, a return to the old days when errors were considered to be an entirely negative aspect of learner language”. However, as they argue (Altenberg & Granger 2002:14), “it [error analysis] is a key aspect of the process which takes us towards understanding interlanguage development and one which must be considered essential within a pedagogical framework.” In knowing what learners can be expected to have acquired at a certain stage in their learning, teachers and materials designers are put in a position where they can optimise their input. Analysing learner errors is a useful way of finding out the stage of learning learners have arrived at.
As Altenberg and Granger (2002:14) also point out, error analysis with the use of computers is quite different from its earlier manifestation, where the focus was on decontextualized examples. With the use of corpora and corpus tools, analysts are able to consider both the context of use and the context of individual errors. At the same time, it might be added, with the use of tagging, analysts are able to retrieve large numbers of errors of the same type, thus determining the prevalence of a particular error within a given group of learners. In addition, where learner corpora are compared with corresponding native speaker corpora by contrasting the number of uses of a given item, consideration can be given to the question as to whether learners are avoiding an item, a sensible strategy if learners are not confident in the use of the item, but one which needs to be attended to, if learners are to make their language more native-like. In addition, in some cases, because of pedagogical practice, students may overuse a given item; this again needs to be taken into consideration.

3. Error analysis and lexis

Another difference between old-style error analysis and computer-based approaches is in the focus of analysis. In the 1960s and 1970s linguistics and language teaching were under the influence of structuralism. This meant a focus on syntax and phonology, at the expense of lexis. While this situation was most predominant in the United States, in Great Britain, led by linguists such as Firth, Halliday and Sinclair, at the same time, a much different approach was adopted in linguistics (if not language teaching), with lexis viewed as an important linguistic level in its own right, although at the same time interdependent with grammar (Halliday’s “lexico-grammatical” level [Halliday 2004]). Lexis, according to this view, could be analysed syntagmatically, as well as paradigmatically. “You shall know a word by the company it keeps”, in the words of Firth (1957:179). This emphasis on the syntagmatic dimension of lexis, as mentioned at the beginning of this paper, was discussed by Halliday and Hasan (1976) in Cohesion in English, where the idea of collocation, or how words tend to co-occur, was developed. Collocation is concerned with purely lexical relations (Stubbs 2001:65). However, there are also relations between lexical and grammatical words, referred to as colligation (e.g. Stubbs 2001:64).

In this paper, in accordance with this syntagmatic view of lexis, the analysis of signalling noun errors will emphasise how they are misused on both the paradigmatic and the syntagmatic axes — how an inappropriate word (paradigmatic axis) or inappropriate collocation/colligation (syntagmatic axis) may be chosen.
4. Signalling nouns

A signalling noun is potentially any abstract noun which is unspecific out of context, but specific in context (Flowerdew 2002, 2003a, 2003b). To put it another way, the full meaning of the word is not made clear unless it is put in a specific context. Specifics may be realized in the clause, as in examples 1 and 2, where the full meaning of the abstract (signalling) nouns “threat” (example 1) and “question” (example 2) are realised in the underlined portion of text (all examples are drawn from the academic corpus used in Flowerdew [in preparation]).

(1) While it provides opportunities for interested consumers to participate in research studies, the threat that professional participants will skew results raises ethical issues for researchers who want impartial, naive study participants.

(2) And really addressing the question of whether, for all sorts of evidential and process reasons, we should be greeting silence as proof, and whether we should be putting accused persons under pressure to speak either in interview or in court.

Specifics may also be realised across clauses, as in examples 3 and 4. In example 3 “findings” is realised in the preceding underlined clauses, while in example 4 “risk” is realised in the clauses which follow. In their inter-clausal function signalling nouns may thus be realised either anaphorically (example 3) or cataphorically (example 4).

(3) However, recent laboratory experiments have demonstrated that they are not only strongly dependent on the carbonate chemistry of the culture medium but that the so-called ‘vital-effects’ are probably mediated via perturbations of the local carbonate system. These findings have an important impact on the interpretation of isotope data.

(4) Despite their benefits to medicine, vaccines are not without risk. If the vaccine maker does not completely inactivate a pathogen, the recipient can actually develop polio or another serious disease.

Signalling nouns may also rely on background knowledge in order to be made “specific”. In example 5, the variable meaning of “impact” is not supplied, an analysis of its context reveals. Readers have to use their background knowledge to provide a suitable possible interpretation of what the (favorable) “impact” might be.

(5) The ensuing sharp increases in grain prices in 1988 had a favorable impact on farmers.
5. The corpus

The corpus used in this study consists of 110,000 words, drawn from a larger corpus of 390,587 words, made up of argumentative essays written by Cantonese L1 first year students at a university in Hong Kong. The essays were written under examination conditions and the students did not have access to dictionaries. Students were given one hour 45 minutes to complete the task. They were given a word limit of 450–500 words. There were altogether 210 essays in the smaller corpus, on nine topics, as follows: Recycling, Smoking Ban, Credit Cards, Hong Kong Country Parks, Immigrants from China, Lowu Railway, Peer Assessment and Cyber Cafes. Students were given one of these topics and provided with six sources, each consisting of six short paragraphs. The sources took the form of extracts from books, journals and the internet. Students were told they were only allowed to use one direct quotation, but, where quotations were used, they were deleted before the scripts were put into the corpus. The essays from the larger corpus to be included in the smaller one were selected to include equal numbers of those receiving B+, B, C+, C, D+, and D grades. Essays receiving A and F grades were not included, as there were not sufficient numbers of these grades to balance those awarded the middle range grades.

The larger corpus was created with a view to being incorporated into the International Corpus of Learner English (ICLE) created by Sylviane Granger of the University of Louvain (Belgium). Although a version of the corpus has now been included in ICLE, at the time of the study this was not the case. Access to the data was kindly provided by Linda Lin, who created the corpus.

6. Methodology

The methodology for identifying and classifying signalling noun errors in the corpus was as follows. A trained research assistant, Delian Gaskell, who had previously worked on a funded research project on signalling nouns in academic discourse, systematically sampled the larger corpus referred to above (i.e. essays were taken equally for each letter grade), identifying signalling noun errors until a theoretical saturation point (Glaser & Strauss 1967) was reached, that is to say a stage was arrived at where no new categories were being discovered (124,000 words). As errors were found in the corpus, they were copied from the text and pasted into a separate file for later analysis. Together with the author, hypotheses were developed as to possible distinctive types of error as new examples were pulled from the corpus. Once broad categories of error types became apparent, the research assistant
continued to search through the corpus until a sizeable number of examples of each error type was found. As this process took place, where examples did not fit into the hypothesised categories, then new categories were created, sometimes replacing existing categories, if they had not been expanded greatly. The process was thus an iterative one, with the number of categories being expanded and collapsed until the collected errors could be accounted for in the most parsimonious way. The resulting taxonomy of errors was thus a joint production on the part of the research assistant and the author.

7. Findings

Figures 1 and 2 show how the errors identified in the corpus were classified. As these figures show, errors were classified into the following categories: colligation, collocation, incorrect signalling noun, and missing signalling noun. This section will describe and exemplify each category with corpus examples. The types are discussed in order of frequency. However, more detailed frequency data will be presented in the next section.

7.1 Colligation errors with signalling nouns

The most frequent category of error is colligation (see Section 3 for definition). Learner errors in this category were primarily with following prepositions. Signalling nouns can be followed by many different prepositions and there does not seem to be any apparent systematicity in their use (although this would be an area worthy of study). This would account for the difficulty learners have here. In his book, Schmid (2000) provides a wealth of examples of signalling nouns and many examples of the different prepositions which typically follow certain nouns. Example 6 provides a group of such problematic colligations (signalling noun in bold, problematic preposition underlined, appropriate alternative in brackets at the end of each phrase). The appropriacy of each colligation was checked against the British National Corpus, using the Phrases in English software (http://pie.usna.edu/ accessed 7 December 2005); also for collocation, see Section 7.3:

(6) a. the major argument in supporting the development of country parks (for)
   b. the chance to suffer lung cancer, heart diseases and respiratory diseases (of [suffering])
   c. discrimination to smokers (against)
   d. a great effort on banning smoking in restaurants (to [ban])
   e. the argument on developing country parks (for)
7.2 Incorrect signalling nouns

The second category is that of incorrect signalling noun. Errors here can be grouped according to meaning and form (see Figure 2).

7.2.1 Error in meaning i.e. wrong lexical noun

Examples 7 and 8 are examples of this category. In example 7 the realisation (underlined portion) refers to the effect of low wages on spending power. Clearly this cannot be labelled as “pollution”, as the writer has done. A more appropriate signalling noun might be something like “situation”, perhaps pre-modified with
a negative epithet (the student’s use of “pollution” would suggest that a negative connotation is perhaps being attempted). In example 8 “way” might perhaps be better replaced by a signalling noun such as “procedure”. On the other hand a post-modifier might make the use of “way” more acceptable, e.g. “stable way of dealing with the matter”

(7) It may further worsen our economic pollution also after the economic downturn. If their wages are low, they must not prefer to spend more in daily life

(8) The main purpose of establish this scheme that the government hopes all of groups will be able to support themselves. It seems to provide the stable way to them.

Although the signalling nouns in these examples are clearly inappropriate, they do show that the learners have grasped the concept of using signalling nouns to encapsulate a concept which is elucidated in the following (example 7) or preceding (example 8) context. Errors such as these might be considered developmental in terms of second language acquisition. The learners understand the functioning of signalling nouns; their problem is that they lack knowledge of the specific abstract noun in the given context (i.e. their problem is on the paradigmatic, rather than the syntagmatic dimension).

7.2.2 Errors of signalling noun form
Signalling noun errors concerning the form of the noun fall into three categories: noun form confusion; adjective vs. noun confusion; and gerund vs. noun confusion.

7.2.2.1 Noun form confusion. Examples 9 and 10 show noun form confusion.

(9) Nevertheless, modern recycling is a capital-intensive and relatively high-tech industry. The process of recycling demands large amount of capital. This expenditure may be greater than the cost of products which produce by virgin raw materials.

(10) People against this points out that this would hurt unempolyment and the prospection of local university graduates. Mainland supports large in numbers of these workers relatively to Hong Kong. So local intellectuals prospects may be affect because companies would employ Mainland specialists rather than local workers.

In example 9 “expenditure” seems to be an attempt at the signalling noun “expenditure”, while in example 10 “prospection” is close to “prospects”, which would be
appropriate. It might be argued that problems such as these are not really problems with the cohesive function of signalling nouns, but are more appropriately considered as problems of spelling or word formation. On the other hand, it can also be argued that such errors should be considered as signalling noun errors, due to the fact that a large number of signalling nouns are deverbal nouns and they follow the affixation rules associated with nominalization. This can be seen in Table 1, where out of the top 100 nouns from a corpus of academic English (Flowerdew in preparation referred to above), the following affixes can be noted (some of them more than once): -ation, -ure, -ition, -ity, -ion, -ship, -ism, -ance, -ence, -age, -istic and -ulty. It is ironic that the learners who produced examples 9 and 10, while understanding that signalling nouns often take affixes, have chosen the wrong ones. On the positive side, this can be considered as a stage in the learners’ interlanguage, where they at least have a partial understanding of signalling nouns and how they work.

Table 1. Top 100 signalling nouns from an academic corpus

| 1. example | 21. process | 41. argument | 61. period | 81. role |
| 2. case     | 22. factor  | 42. possibility | 62. stage | 82. objective |
| 3. model    | 23. type    | 43. ability    | 63. purpose | 83. coefficient |
| 4. result   | 24. fact    | 44. difference | 64. discussion | 84. decision |
| 5. way      | 25. principle | 45. law       | 65. failure   | 85. behaviour |
| 6. problem  | 26. reaction | 46. range     | 66. attempt   | 86. intention |
| 7. equation | 27. issue   | 47. area      | 67. property  | 87. prediction |
| 8. theory   | 28. approach | 48. concept   | 68. chapter   | 88. hypothesis |
| 9. term     | 29. experiment | 49. analysis | 69. feature   | 89. number    |
| 10. section | 30. procedure | 50. conclusion | 70. expression | 90. implication |
| 11. idea    | 31. form    | 51. situation | 71. potential | 91. advantage |
| 12. point   | 32. condition | 52. paper     | 72. technique | 92. definition |
| 13. figure  | 33. rate    | 53. policy    | 73. topic     | 93. observation |
| 14. thing   | 34. right   | 54. view      | 74. parameter | 94. notion    |
| 15. system  | 35. kind    | 55. response  | 75. mechanism | 95. characteristic |
| 16. table   | 36. solution | 56. relationship | 76. instance | 96. phenomenon |
| 17. question | 37. function | 57. strategy  | 77. evidence  | 97. target    |
| 18. reason  | 38. change  | 58. consequence | 78. part     | 98. word      |
| 19. effect  | 39. value   | 59. assumption | 79. introduction | 99. difficulty |
| 20. method  | 40. study   | 60. step      | 80. test      | 100. subject  |
7.2.2.2 Adjective-noun confusion. Examples 11 and 12 indicate adjective-noun confusion.

(11) **Students can purchase their reference books and materials by credit card through internet system in local and overseas. They just need to la in their account number and the expiry date. After all the transactions are completed. The goods will be received in a few days.** It allows students more **compatible** with the advance technology of the society.

(12) **For example, travels can gain any types information though internet, at the cafes in most major airports due to their requirements or interesting.** <Reference> states that the travelers and tourist can check information of most major airports and any kinds of information which they interest though internet inside cyber cafes. For example, the travelers who want to get some guides for their travel.

In example 11 the writer could have used a noun “compatibility” instead of the adjective “compatible”. In example 12, interestingly, the writer refers to “requirements or interesting”, thus selecting a noun (“requirements”) appropriately, but making an error with the selection of an adjective (“interesting”), where the noun “interests” was needed.

These errors seem to be due to a lack of knowledge of the nominal form of an adjective which is already known. The use of the adjective rather than the noun at least performs the function of conveying the semantics, if not the morphology, of the message the learners want to convey. Again, this could be claimed to be a problem of general lack of knowledge of word formation. Given that adjective/noun confusion is a general problem, then it should not be treated as a specific signalling noun problem, but perhaps tackled with frequent signalling nouns. On the other hand, as noted earlier, in academic writing, these signalling nouns are very frequent and therefore deserve a separate treatment and adjective/noun confusion could be part of that treatment.

7.2.2.3 Gerund-noun confusion. In examples 13 and 14 writers have used gerunds inappropriately.

(13) **According to<R>, breathing secondhand smoke increases the risk of lung cancer and heart disease by about 25%, the reason behind this increasing is that there is no smoke-banning in public places, especially restaurants.**

(14) **The decreasing of economy in Hong Kong is seriously.** Many people lose their job or earn a lower salary than before.
In example 13 “increasing” is used where “increase” is required. Similarly, in example 14 “decreasing” is used instead of “decrease”. These errors are understandable, given that many signalling nouns do take the -ing suffix (e.g. “easing”, “spending”, “writing”). We seem to have here a case of over-generalization; learners are relying on a limited repertoire of nominal suffixes, specifically the suffix associated with the gerund.

7.3 Collocation errors with signalling nouns

The third major category of error associated with inter-clausal uses of signalling nouns is to do with collocation (Figure 1). Such errors may concern either verbs or pre-modification. Looking first at verbs, examples 15 and 16 are instances of these.

(15) According to the Hong Kong SAR’s new policy, smoking will be banned in restaurant. After this law has been opened to the public, there have been a lot of argument on this law.

(16) <Reference> mentioned that the burning of plastic waste generates toxic by-products and plastic waste is nonbiodegradable, cannot decompose by micro — organism, when landfilled it is truly that plastic is made of some toxic element such as chloride. It will release lots of toxic gases during the process of burning. Also, after landfilling, the plastic waste still take up are third of landfill space. In conclusion, recycling seems to be the only method of waste management in Hong Kong as it can store the problems arising from the tranitional method and reduce pollution.

In example 15 the signalling noun “law” is collocated with the verb “opened”, where something like “introduced” would have been more appropriate. In example 16 the verb “store” is collocated with the signalling noun “problems”, where a verb such as “resolve” or “overcome” would have been more suitable.

Finally, looking now at pre-modification we have instances of this type of error in examples 17 and 18.

(17) On the other hand, however, a number of experts have presented that using credit card would case a great deal of burden on students. Lau pointed out that the main problem caused by permitting students to hold credit cards is that they may get into debt.

(18) Some people, however, think that it would draw a serious impact to the business, causing billions of dollars’ loss and unemployment.
In both of these examples the signalling nouns seem to be constricted in the collocations they may accept. Thus in example 19, where we have “a great deal of burden”, the prefabricated pattern “great burden” is acceptable, but not what the writer had produced. Similarly in example 20 the prefabricated pattern would be with “have” or “result in” or “create”, but the writer has used “draw”.

### 7.4 Omission of signalling noun

Turning now to the next category of error (Figure 1), a minor one, examples 19 and 20 show the omission of signalling nouns.

(19) Actually, most of the disposal in Hong Kong is plastic, if policy of recycling is adopted, serious pollution problem can be reduced. Also, recycling can conserve energy in managing waste disposal. In addition to these, the method of recycling can also reduce management cost in waste and thus increase competitiveness of manufacturers.

(20) Although, playing games is a way to gain relaxation, it may also contains some adverse effects on young people’s physically and mentally. As time goes by, it may harm their eyesight, moreover, majority of the online games contain real-life violence which may mislead the youth to use violence, to solve problems.

In example 19, although we have an anaphoric pronoun, “these”; a signalling noun such as “measures” would make the meaning clearer. It seems that the writer either does not know that a signalling noun is needed here or, if they do, do not know one which would be appropriate. In example 20 we have two adverbs, “physically” and “mentally”, where more appropriate would be adjectives followed by a signalling noun such as “well-being” or “health” (i.e. “physical and mental well-being/health”). (Alternatively, this could be corrected by deleting the possessive in “people’s” to give “some adverse effects on young people physically and mentally”).

### 7.5 Frequency data

Figure 3 provides frequency data for the different types of error. Figure 3 shows that the most common problem was with incorrect signalling nouns, followed by incorrect use of signalling noun, errors in collocation and no signalling noun. Out of an overall total of 1451 errors, there were 989 colligation errors, 283 incorrect signalling noun uses, 153 errors of collocation and just 26 examples of no signalling noun when there should have been one.

Figure 4 shows that the highest grade students, B+, use on average 3.95 signalling nouns per hundred words, while the lowest grade, D, use only 3.05, clearly
indicating that there is a general decrease in the use of signalling nouns by learners of lower grades, despite a slight upward number between C+ and C students (of course, we need to bear in mind here to what extent the grades awarded were reliable). An ANOVA test carried out on these figures shows that the difference in signalling noun use among the different grades of essays was statistically significant (p<0.05).

Figure 5 shows that lower grade students make a higher number of errors when they use signalling nouns. When B+ students use signalling nouns, only
27% have errors associated with them, while the lowest grade of student in the study, D, have 44% with associated errors, a difference of 17%. An ANOVA test confirmed that these differences in number of uses and number of errors were significant (p<0.05). Pearson’s Correlation Test was also used to find out the correlation between the grade of essays and percentage of errors made. This test revealed a positive, although weak-to-moderate, correlation between the grades of essays and the percentage of errors made (p<.05, \( r = -.232 \)), suggesting a confirmation of the trend that the lower the grade, the more errors are made.

8. Discussion

Based on the learner corpus of Hong Kong writers, this study has identified a range of errors in the use of signalling nouns. Specifically, problems it has identified are: 1. colligation errors; 2. incorrect signalling noun choice; 3. missing signalling noun and 4. collocation errors with signalling noun. Within the colligation category, problems are mainly to do with following prepositions. In the incorrect category, problems arise from incorrect choice of noun and incorrect word form, the latter category including noun form confusion, adjective vs. noun confusion and gerund vs. noun confusion. Within the collocation category, problems are to do with collocation with verbs and pre-modification. Some of these problems are likely to be related to transfer from Cantonese, notably adjective/noun confusion, gerund vs. noun form confusion, and collocation, although without a comparison with a corpus of learner data from other first language groups it is difficult to make strong claims here. Other problems may be to do with overgeneralisation, notably incorrect word form and gerund vs. noun confusion.

In the earlier part of this paper it was anticipated that signalling noun errors could possibly occur on either the paradigmatic axis (how an inappropriate word may be chosen) or the syntagmatic axis (how the word is related to other words).
As it turns out, both types of error have been identified, although the syntagmatic axis (at least in terms of frequency) seems to be more problematic. Inappropriate signalling nouns function on the paradigmatic dimension (e.g. examples 7 and 8) while colligation and collocation problems operate on the syntagmatic dimension (e.g. examples 6 and 15–18). Missing signalling nouns, on the other hand, could be functioning on either level (e.g. examples 19–20); the learner might not be aware of the appropriate noun to choose and so leaves it out (paradigmatic) or they might not realise that a noun is needed (syntagmatic).

9. Limitations of the Study

Before discussing possible implications of the study, some limitations and ways in which the research will be taken forward will be pointed out. First, no comparison with native English-speaker data are provided. This would identify if students are under- or over-using signalling nouns (intuition would suggest under-use) and to what extent their range of use (i.e. different types) might vary with that of native speakers (again intuition would suggest a narrower range on the part of the learners). This limitation will be rectified in a follow-up study which will compare the learner corpus used in this study with a native speaker corpus (such a corpus — the LOCNESS corpus — is available with the ICLE set of learner corpora). In addition, contrasts will be made with comparable learner corpora from other languages, with a view to seeing to what extent the problems identified with the Hong Kong Chinese learners are “universal” and to what extent they are due to the specific typology of the L1 of the Chinese writers. However, as Ellis and Barkhuizen (2005:66) warn, one needs to be very careful in assigning the cause of a given error, not least because many errors are likely to be due to multiple rather than single sources.

10. Implications of the study

In spite of the above limitations, the results of this study are not without value from the pedagogic point of view. They also have something to tell us about the importance of signalling nouns as a part of lexical cohesion in general. Taking the pedagogic aspect first, as the focus of this paper is primarily on pedagogy, to paraphrase Corder (1967, cited in Barlow 2005:51), as with error analysis in general, the results of this study are significant in three ways: 1. they serve a pedagogic purpose by showing learners what they have not yet mastered; 2. they serve a research
purpose by providing evidence about how languages are learned; and 3. they serve
a learning purpose by acting as devices by which learners can discover the rules
of the target language (i.e. by obtaining feedback on their errors). The taxonomy
and examples and the syntagmatic/paradigmatic distinction are clearly useful in
categorising errors with signalling nouns. The classification of signalling noun er-
rors in this way can provide a check-list which can be used for error diagnosis and
correction and for the design of teaching materials to improve the use of signalling
nouns by Chinese learners, if not groups from other language backgrounds as well.
The frequency data demonstrate that effective use of signalling nouns is a develop-
mental phenomenon which correlates with overall writing ability. Related to this
issue, consideration should be given to the question as to what degree the “errors”
identified in this paper represent different degrees in terms of developing interlan-
guage. In example 9, where the learner writes expenstion, they have clearly under-
stood quite a lot about the functioning of signalling nouns. They have realised that
a signalling noun is appropriate where they have placed it; they have identified the
right lexeme; they have realised that signalling nouns often end with a nominal
suffix; the only problem remaining is in the wrong choice of suffix (-tion instead
of -iture). A similar degree of competence can be noted in example 10, where the
writer has used prospection instead of “prospects”, the only problem again being
in the choice of suffix. These errors show that a high degree of acquisition would
seem to have taken place and that these errors might be viewed in a different light
to those, for example, where the wrong signalling nouns have been selected or, in-
deed, where a signalling noun might be seen to be missing, as in some of the other
examples presented in this paper.

Finally, turning now to the importance of signalling nouns as an aspect of
lexical cohesion in general, the data presented in this study have shown that those
learners who achieved higher marks tended to use more signalling nouns and
with greater accuracy. This confirms the intuitive idea that signalling nouns are
an essential dimension of lexical cohesion, which contributes to the coherence of
a text.

Notes

* I should like to acknowledge the help of my research assistants, Delian Gaskell and Jennifer
Eagleton in the preparation of this paper, along with Michaela Mahlberg and the anonymous
reviewers for comments on an earlier draft. I should especially like to acknowledge the generos-
ity of Linda Lin in making available her corpus for this study. This paper was funded as part of
a grant from the Research Grants Council of the Hong Kong Special Administrative Region,
China (Project No. CityU 1312\03H).
References


Lexical cohesion

Corpus linguistic theory and its application in English language teaching*

Michaela Mahlberg
University of Liverpool

Cohesion is generally described with regard to two broad categories: ‘grammatical cohesion’ and ‘lexical cohesion.’ These categories reflect a view on language that treats grammar and lexis along separate lines. Language teaching textbooks on cohesion often follow this division. In contrast, a corpus theoretical approach to the description of English prioritises lexis and does not assume that lexical and grammatical phenomena can be clearly distinguished. Consequently, cohesion can be seen in a new light: cohesion is created by interlocking lexico-grammatical patterns and overlapping lexical items. A corpus theoretical approach to cohesion has important implications for English language teaching. The article looks at difficulties of teaching cohesion, shows links between communicative approaches to ELT and corpus linguistics, and suggests practical applications of corpus theoretical concepts.

Keywords: lexical cohesion, lexical item, collocation, semantic prosody, communicative approaches

1. Introduction

The term ‘cohesion’ is used to refer to the property of connectedness that characterises a text in contrast to a mere sequence of words. Sometimes ‘cohesion’ is contrasted with ‘coherence,’ where the former focuses on features on the textual surface and the latter describes underlying meaning relationships which can, but need not, be reflected by features on the surface text (see for instance de Beaugrande & Dressler 1981). Cohesion should play an important role in English language teaching (ELT), as readers and writers need to be aware of the links that hold chunks of text together and that contribute to the creation of a text as a unit of
meaning. Cohesion can contribute to the readability of a text and have an impact on the comprehensibility and clarity of the argument. Additionally, the way in which links between textual chunks are signalled reflects genre-specific properties of texts. Thus, an appropriate use of cohesive devices is essential for language learners to develop a native-like competence of text production and reception. However, the teaching of cohesion does not seem to be a straightforward issue. When Cook (1989:127) complains that cohesion does not receive enough attention in traditional language teaching, he observes that “[c]ohesion between sentences is too easily seen as an aspect of language use to be developed after the ability to handle grammar and words within sentences”. Such an approach to cohesion can result from viewing words as fairly independent linguistic units and syntactic rules as the main principles that guide the combination of words into sequences. When the grammar and the vocabulary of a language are thus separated it is easy to talk about ‘grammatical’ and ‘lexical’ cohesion, a division that is not uncommon in language teaching textbooks.

It is clear that issues that we encounter in pedagogic approaches to cohesion are not merely a consequence of the requirements of the classroom, they also reflect general linguistic beliefs. The present article argues in favour of a corpus theoretical approach to cohesion, which views cohesion as a fundamentally lexical phenomenon, and suggestions will be made for the application of this approach in ELT. The article starts with a look at different types of cohesion (Section 2). Section 3 summarises the main difficulties of dealing with cohesion in the context of ELT. Section 4 looks at communicative approaches in ELT and how they link in with corpus linguistic ideas. Section 5 introduces a corpus theoretical approach to cohesion and Section 6 presents some textual examples. Section 7 then looks at implications of the corpus theoretical approach for ELT and Section 8 concludes the article.

2. Grammatical, lexical and other kinds of cohesion

A standard book on cohesion is Halliday and Hasan’s (1976) *Cohesion in English*. According to Halliday and Hasan (1976:299) “[c]ohesion expresses the continuity that exists between one part of the text and another”. The authors devote the biggest part of their book to grammatical cohesion (reference, substitution, ellipsis, and conjunction), but they also discuss lexical cohesion. Lexical cohesion can occur in the form of ‘reiteration’ or ‘collocation’. Reiteration is the repetition of a lexical item or the use of a synonym of some kind in the context of reference (Halliday & Hasan 1976:318), and collocation covers all types of lexical relations
that do not need referential identity and cannot be described as a type of reiteration (Halliday & Hasan 1976:287). In contrast to Halliday and Hasan (1976), Hoey (1991) gives lexical cohesion a more central role. He observes that “[l]exical cohesion is the only type of cohesion that regularly forms multiple relationships” (Hoey 1991:10) between elements in the text. According to Hoey (1991:10) “the study of the greater part of cohesion is the study of lexis, and the study of cohesion in text is to a considerable degree the study of patterns of lexis in text”. To describe patterns of lexis Hoey (1991) discusses various categories of lexical repetition. One of his categories is, for instance, ‘complex repetition’. The words *argue* and *argument* are illustration of complex repetition as they share a lexical morpheme, but are not formally identical. Another example is *meeting* as a verb and *meeting* as a noun, which are formally identical but have different grammatical functions. With the help of such categories, Hoey (1991) can achieve a detailed account of cohesive lexical relations in text. Hoey’s (1991) main interest is in the textual organisation to which cohesive links contribute. Cohesive links can be interpreted, for instance, as an indication of the centrality or marginality of sentences in texts.

Restricting a discussion of cohesion to the work of Halliday and Hasan (1976) and Hoey (1991) would neglect the many publications that have contributed to research in the field. For instance, the fairly brief account of lexical cohesion in Halliday and Hasan (1976) has been later revised by Hasan (1984) and Martin (1992). In the context of language teaching, a very useful account of vocabulary in text is given by McCarthy (1991) in his chapter on ‘Discourse analysis and vocabulary’. In addition to a section on lexical cohesion, McCarthy (1991) deals with issues such as discourse-organising words, which overlap with Winter’s (1977) vocabulary 3 and Francis’ (1986) anaphoric nouns. Words such as *problem*, *fact*, *situation*, etc. can package text and indicate larger text-patterns. Discourse-organising words can be seen as somewhere in between lexical and grammatical cohesion and thus are similar to Halliday and Hasan’s (1976) general nouns (a related concept is that of ‘signalling nouns’, see Flowerdew in the present issue). However, words that organise discourse are not restricted to the word class noun and words that organise discourse need not necessarily be as general in meaning as *problem* or *fact*. McCarthy (1991) illustrates, for instance, how vocabulary in text can be interpreted as a reflection of a problem-solution pattern. Words such as *concern*, *drawback*, *hamper*, *obstacle* can indicate a ‘problem’. Potential signals of a ‘response’ to the problem are *change*, *combat*, *come up with*. The words *answer*, *effect*, *outcome*, can be indicative of ‘solution/result’ and items such as *effective*, *overcome*, *unsuccessful* could be interpreted as indication of an ‘evaluation’ of the result (cf. McCarthy 1991:79). A more detailed discussion of such culturally popular patterns, although not with the main focus on language teaching, can be found, for instance, in Hoey (2001); for a recent corpus-based study of the problem-solution pattern see Flowerdew (2008).
Connectedness in text is not only reflected by the choice of vocabulary words or grammatical linking words; the choice of tense and aspect also contributes to textual relations (see for instance Quirk et al. 1985:1454ff.); we can include parallelisms and adjacency pairs in lists of cohesive devices (see for instance Morley 1999:51ff.); the flow of information, that often progresses from given to new, plays a role in the transition from single sentences to connected text (see for instance Biber et al. 1999:896f.); and eventually genre conventions have an impact on the links between parts of a text. Although I could do no more here than give a brief overview, it should have become clear that cohesion is a complex phenomenon to describe and that we can find cross-relations to a variety of fields.

3. Difficulties of describing cohesion in textbooks

The brief outline in the previous section has shown that a description of cohesive devices can be approached from different angles. However, in spite of the various modifications, revisions and additions to cohesive categories that have been suggested, accounts of cohesion tend to share some underlying assumptions about grammatical and lexical distinctions between cohesive devices, as we find them in Halliday and Hasan’s (1976) work. In particular, in publications relevant to language teaching the outline suggested in Cohesion in English often serves as a framework or a convenient starting-point; see, for instance, Salkie’s (1995) workbook or the textbook by Hatch (1992).

A distinction of cohesive devices into grammatical and lexical is indicative of an approach to text that builds on the grammatical categories used to characterise words in sentences. However, beyond sentence boundaries, it becomes increasingly difficult to present a systematic account of linguistic categories. This situation does not only hold for a theoretical discussion of cohesion but also for the role that cohesion plays in language teaching. Additionally, in language teaching the learners’ needs, teaching aims and questions of teaching and learning methodologies have to be taken into account. Thus, in teaching materials that aim at developing writing skills, for instance, complex issues of cohesion may be condensed into a list of words and phrases like on the one hand, on the other hand, additionally, in contrast, etc. that are presented as a help for learners to connect their arguments. Sometimes even longer phrases are suggested as text-building devices that could help learners to produce continuous text. Werlich (1988), for instance, suggests phrases of the type from the point of view of the setting/the problem/the thesis or it is interesting to note that in order to help students with the production of text.
Textbooks have the difficult task of choosing the right categories that are useful to learners. In the following, I will summarise some of the points that make cohesion a challenging topic for textbooks. It has to be noted that the present article focuses on written language; spoken discourse would add further points to the discussion.

- **Appropriate exemplification**: textbooks can often be criticised for using only few, or simply two, constructed sentences to illustrate cohesive links. Short examples provide a simplified picture neglecting features that span larger contexts and that create complex networks of meaning relationships. Taking real examples of cohesion would require more space and make an overview of cohesive devices fairly clumsy. On the other hand, finding a single natural text that could serve as illustration of most of the cohesive features that are listed in an overview is very difficult. Whereas most texts will provide several examples of pronouns functioning as ‘reference items’, it is more difficult to find a useful text for the illustration of the ‘substitute’ one, which Halliday and Hasan (1976:89) introduce with the help of the following example: *My axe is too blunt. I must get a sharper one.*

- **A detailed textual analysis is time-consuming**: it is not only space that is needed for real examples but also time. Time is a crucial factor in teaching and learning. The analysis of full texts requires more attention than the discussion of just a few sentences. Connected text can have many multi-directional links and complex clues to indicate these relations. Real text may also be more difficult than what the learner is able to handle at a certain stage, so more time is needed to fill in background knowledge. Thus, the simplification in the presentation of examples is also a matter of time-efficiency.

- **Cohesive links are genre-specific**: one of the reasons why a single text to illustrate a variety of cohesive devices is hard to find is that cohesive links are genre-specific. Narrative texts that deal with a central character, for instance, can provide many examples of reference and chains of reference items. In contrast, newspaper articles seem to be more likely candidates to illustrate lexical relationships where sentences share three or more lexical links of the type that Hoey (1991) discusses.

- **Generalisation**: textbooks have to generalise. They have to select and parcel facts about language for learners. Even if textbooks aim to give learners the opportunity to discover linguistic facts for themselves, they have to provide a general overview. They have to give some indication of how to organise linguistic knowledge and how to find systematic relationships between linguistic phenomena. Listing categories of grammatical cohesion seems to be an efficient way of introducing cohesive devices. The number of grammatical word classes
is limited, so a list of cohesive devices will be fairly manageable. In contrast, lexical items can be less clearly grouped into categories. Repetition and the use of synonyms or antonyms are some of the lexical possibilities to create cohesive links, and the number of words that could illustrate such cohesive links is endless. Therefore, an overview of lexical cohesion will have to be fairly abstract and restricted to selected examples.

4. Corpus linguistics, the communicative approach and the teaching of cohesion

So far we have focused on problems of describing cohesion without looking at a more fundamental issue: the linguistic framework that is taken as the basis for the description. The separation between lexis and grammar reflected in the teaching of cohesion is still often upheld in linguistics, but it is not taken for granted anymore. In particular, corpus linguistic research has been accumulating evidence that the boundaries between lexical and grammatical categories are less clear-cut than traditional approaches seem to suggest. With the observation of recurrent patterns of words, corpus linguistics draws attention to the importance of lexical patterns. The focus on the description of lexical phenomena that is advocated by corpus-driven linguistics in particular, also plays a role for language teaching. Early ideas on lexically oriented and corpus informed approaches in language teaching are put forward by Sinclair and Renouf (1988), who sketch a 'lexical syllabus', a concept which is further developed by Willis (1990). The central argument is that language teaching should focus on the common patterns of common words. Lexis also takes centre stage in Lewis’ (1993) suggestion of a ‘lexical approach’ to ELT. One of the key principles of the lexical approach is that “[l]anguage consists of grammaticalised lexis not lexicalised grammar” (Lewis 1993:vi). Whereas Sinclair and Renouf (1988) and Willis (1990) take a corpus linguistic approach to ELT, the work of Lewis is in the first place set in an ELT context. The lexical approach is less concerned with issues of corpus linguistics but more with the needs of the classroom, building on useful approaches to language teaching and experience of practitioners (see for instance Lewis 2000). Lewis (1993:vi) points out: “[t]he Lexical Approach develops many of the fundamental principles advanced by proponents of Communicative Approaches”. However, corpus linguistic research also lends support to the lexical approach. In its early stages, the lexical approach mainly concentrated on fixed phrases, whereas corpus linguistic approaches stress the variability of phrases (as we will see in the following section). More recent applications of the lexical approach have profited from the support of corpus linguistics.
Corpus linguistic theory and its application in ELT

Corpus linguistics has also developed further approaches that are directly related to language teaching. As McCarthy (2001:128) puts it “[t]he language of the corpus is, above all, real, and what is it that all language learners want, other than ‘real’ contact with the target language.” The contact with corpus data can be exploited to raise the learners’ awareness of textual patterns and make them discover linguistic facts for themselves (see for instance Johns 1991). Numerous suggestions have been made for how to use corpora in language teaching and publications that address language teaching with specific attention to national curricula have started to emerge (see Mukherjee (2002) on the situation in German classrooms). Furthermore, theoretical issues of corpus linguistic work have started to develop and have an impact on ELT. An important contribution has come from Hunston and Francis (2000), who suggest a Pattern Grammar to provide a systematic account of the patterns of words and relationships between words with similar patterns. Willis (2003), for instance, shows how a Pattern Grammar can become part of language teaching. A question that still needs more attention is how we can move on from lexical patterns to connected texts.

The centrality of lexis that becomes evident through corpus linguistic observations leads Müller-Hartmann and Schocker-von Ditfurth (2004:98) to conclude that “lexical patterns become the major category in learning and teaching discourse”. Müller-Hartmann and Schocker-von Ditfurth (2004) are not primarily concerned with a corpus linguistic approach to language teaching, but focus on the development of intercultural communicative competence. Still, their view shares ideas with the corpus theoretical approach that underlies the present article and that will be discussed in Section 5. Müller-Hartmann and Schocker-von Ditfurth (2004:93) criticise the more traditional approach of dealing with vocabulary and grammar in separate ways and point out that “when we view language as a tool that you use to create meaning, then it is more appropriate to look at the different sub-systems, such as words, grammar, and sounds as a coherent whole under the notion of discourse”. Words and their patterns are seen in their communicative and cultural contexts and meaning is viewed as a result of intertextual relations. However, Müller-Hartmann and Schocker-von Ditfurth (2004) seem to move on too fast from lexis to larger units of discourse without paying enough attention to the role of lexis in creating cohesion.

A focus on discourse and communicative skills does not necessarily answer all questions of cohesion in ELT, but it seems that communicative approaches play an important role in bringing ELT and corpus linguistics closer together. Communicative approaches and corpus linguistics share the view that language is used in context. A central notion of the communicative movement in language pedagogy after 1970 was “the conviction that language teaching should take greater
account of the way that language worked in the real world” (Howatt & Widdowson 2004:326). Interpreted in linguistic terms, a key concept of the communicative movement is the fact that “in communicative contexts language is viewed as a unified event” (Howatt & Widdowson 2004:332). Thus cohesion attracted considerable attention in ELT and the work of Halliday and Hasan (1976) has had its effects on teaching materials. Now that corpus linguistics can provide new access to language in the real world and approaches to language teaching have been supported by corpus linguistic research, the teaching of cohesion can also be seen in a new light. There are some suggestions that seem to point in the right direction. However, both corpus linguistics and ELT can profit from a discussion of the place of cohesion in a ‘corpus theoretical’ framework. The following section will make some tentative suggestions.

5. Cohesion in a corpus theoretical framework

There is still disagreement on whether corpus linguistics is mainly a methodology or needs its own theoretical framework. Advocates of corpus-driven approaches to the description of English claim that new descriptive tools are needed to account for the situation of real text, and ideas of theoretical frameworks to accommodate such tools have started to emerge. Mahlberg (2005) discusses corpus linguistic work that has theoretical implications or makes theoretical claims, such as Teubert (1999), Hunston and Francis (2000), Sinclair (2004) and Hoey (2005). On the basis of this discussion, key features of a ‘corpus theoretical approach’ are outlined. The main theoretical claims can be summarised here briefly as follows: language is a social phenomenon and meaning can thus be viewed as use; patterns of language use become visible in corpora and corpus evidence illustrates that meaning and form are associated. A corpus linguistic, bottom-up description of language prioritises lexis. In such a framework there is also room for the description of cohesion: cohesion is described from a lexical point of view as part of the local textual functions of lexical items (Mahlberg 2005). Local textual functions describe patterns of words in texts. They aim to characterise how units of meaning fit in with each other in the creation of text. Thus the continuity between parts of texts that is created by interrelated patterns of words can be regarded as a result of local textual functions.

An important concept in this corpus theoretical framework and the description of local textual functions is the ‘lexical item’ that is suggested by Sinclair (e.g. 1996, 1998).³ The lexical item characterises extended units of meaning. It is made up of five categories of co-selection: the ‘core’ and the ‘semantic prosody’ are obligatory
The kind described by Whitehouse. Her true feelings about the case can perhaps be Adams was unable to hide his true feelings. Not for the Leicester City manager gives them status." Benitez kept his true feelings to himself but said of Mourinho but are reluctant to express their true feelings. "Here everything is conciliatory," says Mr to really let rip with their true feelings, only time will tell. http://channel9\) of being able to hide her true feelings. In public life this would later they are goin through and their true feelings. And anyway Don, you say "sexuality of our inability to mask our true feelings was unearthed by Jon Hess, of leaves little doubt as to his true feelings. And as for Li Xian, the appear calm and composed, masking her true feelings. So her demeanour when giving evidence all accounts, has now conveyed his true feelings. We await the outcome with interest as a cloak to disguise their true feelings. Deborah Bruce's production combines imagination I've watched her. No doubt her true feelings were emerging from her body, like he may prefer to let his true feelings out in print. Today the governors that – even if those were his true feelings. Instead he did something far cleverer and moist-eyed, and makes known her true feelings, what does Parker do? He politely get a hint of Do Thi's true feelings but Caine and Fraser are on of being able to hide her true feelings. In public life this would later plans to invade Iraq. Whatever his true feelings about the war, the Labour lawyer was perhaps an indicator of his true feelings. FROM REED ORGAN TO CHEQUERED FLAG Only Mr Chapman's words revealed his true feelings towards his child's murderer: "I hope suffering, but unable to express his true feelings. Offstage in all these books, but Kent last week let slip her true feelings about royal life fortune she marries him, but her true feelings are even displayed at the wedding secretive wife, Snow (who records her true feelings in her precious diary), and, finally upon as evidence of "Old Europe's" true feelings towards Britain when relations have been speak the truth or reveal your true feelings because that's NOT a habit we run. If he was disguising his true feelings he did it masterfully. Instead he people being able to express their true feelings of rawness and to pretend they other Chilean cities to express their true feelings about this fateful day in Chile's

Among the prominent collocations are, for instance, the possessives his, her, and their. We can describe these collocations in a more general way with the help of the category 'colligation'. Colligation is “the co-occurrence of grammatical choices” (Sinclair 2004:32) and we can say that the lexical item with the core true feelings has a colligating possessive determiner, or colligates with other possessive forms. 'Semantic preference' is “the restriction of regular co-occurrence to items which share a semantic feature, for example that they are all about, say, sport or suffering” (Sinclair 2004:142). As Sinclair (2004:35) points out, and as we can see in the concordance sample above, true feelings has a semantic preference for 'expression', which is illustrated by the verbs express, reveal, convey or by expressions such as
let his true feelings out in print. The final category we need to look at is the ‘semantic prosody’. The semantic prosody of an item is a “subtle element of attitudinal, often pragmatic meaning” (Sinclair 2004:145). Sinclair (2004:35) chooses the label ‘reluctance’ or ‘inability’ for the semantic prosody of true feelings. This semantic prosody is illustrated by examples such as reluctant to express their true feelings or unable to express his true feelings. The semantic prosody and the semantic preference can also be fused as in hide, disguise, or mask, or in examples such as get a hint of Do Thi’s true feelings, or evidence of “Old Europe’s” true feelings. The short overview of the main components that make up the lexical item with the core true feelings shows that lexical patterns are not the same as fixed phrases. The description of a lexical item leaves room for the variability of lexical choices in text. It is this variability that is central to a corpus theoretical approach to cohesion.

The concept of the lexical item shows that a unit of meaning is not the same as a single word. Meaning is distributed over several words that are chosen together. The way in which words are chosen together can be observed in collocational patterns: the co-occurrences of words in texts. Such patterns create cohesion. As Stubbs (2001a) points out, the cohesive potential of collocation has not been widely recognised yet: “[s]uch syntagmatic patterning is much more detailed than is generally shown in grammars: it stretches well beyond words and short phrases, and provides a relatively unexplored mechanism of text cohesion” (Stubbs 2001a:309). Stubbs (2001b:108ff.) looks for instance at Sinclair’s (2004) naked eye example to illustrate how phrases and collocations can combine in texts. In Halliday and Hasan’s (1976) leading work on cohesion, collocation is only assigned very little space. The authors describe collocation as “the most problematic part of lexical cohesion” (Halliday & Hasan 1976:284). It is important to note that Halliday and Hasan’s (1976) use of the term collocation differs from how the term is understood by most corpus linguists. In corpus linguistics, collocation describes actual occurrences of words in text. Although there may be different ideas on issues of frequencies and the number of words that can intervene between collocates, in corpus linguistics the actual occurrence in text is crucial to provide evidence of collocations. For Halliday and Hasan (1976) the textual evidence is less central. More important are the meaning associations between words. Examples of collocations from Halliday and Hasan (1976:285) are antonyms such as like … hate, wet … dry, or pairs that are drawn from the same lexical set, such as basement … roof, car … brake, mouth … chin.

Halliday and Hasan (1976:285) claim that “[t]here is always the possibility of cohesion between any pair of lexical items which are in some way associated with each other in the language”. That such collocation is not automatically realised in text is illustrated by Stubbs (2001a). Stubbs (2001a) looks at the example of kick and foot
Corpus linguistic theory and its application in ELT

and points out that “the very fact that KICK implies FOOT means that the words tend not to collocate in real text, since they have no need to” (Stubbs 2001a:311): in almost 200 occurrences of KICK, Stubbs (2001a) found only half-a-dozen occurrences each of foot and feet. The span he investigated was ten words to the left and ten words to the right. This span is slightly larger than the distance between collocates that is normally taken into account. In corpus linguistics, collocations tend to be described with regard to short distances. Halliday and Hasan (1976) make it clear that they are primarily interested in collocations that can occur over larger distances and across sentence boundaries (Halliday & Hasan 1976:286). After all, their main interest is in cohesive ties across sentences (Halliday & Hasan 1976:9). Although not all examples of collocation understood in the sense of Halliday and Hasan (1976) seem to work equally well in natural texts, as Stubbs (2001a) shows, Halliday and Hasan (1976) raise an interesting issue, for which support comes from Hoey (2005).4

In the present article, however, cohesion that results from collocational patterns in texts will be interpreted as cohesion created by interlinking lexical items. Thus, when we describe cohesion, we can look at a text from the point of view of a specific lexical item. Near the core of the lexical item we will find more prominent collocations than closer to the boundary of the lexical item. But we have to move closer to the boundary of an item to account for larger passages of text. Thus the semantic prosody plays an important role in the creation of cohesion. As the brief summary of Sinclair’s approach has shown, the categories collocation, colligation, semantic preference, and semantic prosody become increasingly abstract. To talk about ‘colligation’ means to interpret collocations in grammatical terms. The semantic preference is a way of interpreting co-occurring words in terms of lexical sets. The most abstract category is that of the semantic prosody. The items that occur around the core and give an indication of its prosody need not necessarily be strong collocations. As Sinclair (2004) points out, the semantic prosody is subtle and “not subject to any conventions of linguistic realization, and so is subject to enormous variation, making it difficult for a human or a computer to find it reliably” (Sinclair 2004:144f.). However, the semantic prosody is one of the obligatory components of a lexical item. The semantic prosody is the reason why the item is chosen, even if it is not realised explicitly (Sinclair 2004:144f.). Without the semantic prosody “the string of words just ‘means’ — it is not put to use in a viable communication” (Sinclair 2004:34).

Thus the semantic prosody adds further detail to a corpus theoretical approach to cohesion: lexical items in texts are made of cores that are surrounded by some realisation of a more or less fixed pattern. The closer we come to the boundary of an item the greater the variability of its patterns. The variable patterns open up possibilities of combinations with other items. At the peripheral
end of lexical items we may find semantic prosodies that overlap with or shade into prosodies of other lexical items. As semantic prosodies are attitudinal they do not only add to the connectedness of text but they also play a part in what is sometimes called the ‘tone’ of a text and it is clear that semantic prosodies are not independent of the genre of a text. In the present article, we cannot go into further detail of features related to the semantic prosody which is a concept that is not uncontroversial in corpus linguistics (see, for example Whitsitt 2005). To give a clearer picture of the points that are crucial to the present approach, it is time to look at some examples.

6. True feelings in text

In the above analysis of concordance lines the semantic prosody of true feelings was labelled ‘reluctance/inability’. In his analysis of the lexical item, Sinclair (2004:36) concludes: ‘not only are our true feelings our genuine emotions, but we use this particular collocation when talking about our reluctance to express them, even to ourselves’. This is a very broad characterisation. Although the semantic prosody indicates how an item may integrate with other items in a text, we need to have a closer look at individual examples to find more detailed information. The point that people cannot or do not want to express their true feelings in a straightforward way has many facets. When true feelings are genuine emotions they may be contrasted with what is displayed on the surface. Thus we can find examples such as the following (line 28 in the concordance sample in Section 5).

The face he showed to the world was not that of a man on the run. If he was disguising his true feelings he did it masterfully. Instead he delivered one of the most informed explanations yet of his twisted ideology.


The example comes from an article on an interview that a Pakistani journalist managed to conduct with Osama bin Laden. As in the concordance sample discussed in Section 5, we have an indication of the semantic preference of ‘expression’ in the verb disguise. The semantic preference is fused with the semantic prosody as the example does not deal with the expression of true feelings, but with the possibility of them being hidden. The meaning of ‘expression’ is also present in the first sentence of this little paragraph, in the verb show. And the contrast between show and disguise is carried on by the word instead in the third sentence. Thus we see how the lexical item of true feelings fits in with its context and contributes to the cohesion of the text.
Another aspect of true feelings is the type of emotions concerned. One possibility is that the emotions are negative and refer to feelings of dislike or arrogance that people have towards others. Such feelings are preferred to be hidden because it would not be politically correct or socially acceptable to talk about others in a negative way. But humans are not saints, so sometimes such true feelings get expression and situations may arise as in the article with the headline: “I thought the mike was switched off…” Here is the beginning of this article from the Guardian:

‘I thought the mike was switched off…’

Matthew Tempest, political correspondent
Wednesday November 27, 2002

The Canadian prime minister’s press secretary, Françoise Ducros, was forced to resign yesterday after she referred to the US president, George Bush, as a “moron”, during a private conversation with a journalist at last week’s Nato summit in Prague.

Our political correspondent, Matthew Tempest, looks at other accidentally overheard political insults.

1. President Bush
One of the highlights of the 2000 presidential campaign came when the then governor of Texas, George Bush, turned to his vice-presidential running mate Dick Cheney at a rally in Chicago, pointed out a reporter in the pack, presumed the microphone was switched off, and murmured: “There’s Adam Clymer, a major league asshole from the New York Times.”

“Oh yeah, he is, big time,” Mr Cheney replied.

(Copyright Guardian Newspapers Limited 2002)

Throughout the article we find expressions illustrating that things are said which are not meant for a large audience. People do not want to be open about what they think or feel. In the first paragraph there is private conversation, then in the following sentence accidentally overheard. When Bush is described we hear that he presumed the microphone was switched off and he did not tell or say something but he murmured. These examples, and similar instances throughout the rest of the article, illustrate a network of meanings into which true feelings will fit when mentioned in the final section of the article (quoted below). The article illustrates how true feelings, when expressed, can become political insults. Seven examples of “accidentally overheard political insults” (see above quote) are listed. The final example in this list is about Dr Richard Simpson who is said to have expressed his true feelings in the following way:

7. Dr Richard Simpson
Few may have heard of the junior Labour minister at Holyrood with responsibility for the fire service, if he had not told dinner party guests his true feelings of
the striking firefighters. He said of them: “These people aren’t socialists, they’re protectionists, they’re fascists — the kind of people who supported Mussolini. We must not give in to these bastards.”

Despite insisting he was merely repeating other people’s views on the dispute, Dr Simpson fell on his sword yesterday.

(Copyright Guardian Newspapers Limited 2002)

Thus we have seen how words across the text can be interpreted as being linked by a semantic prosody characterising a reluctance to publicly reveal negative feelings. We have looked at these links from the point of view of the lexical item with the core true feelings. To view lexical cohesion as the results of networks of interlinking lexical items is not necessarily in contrast with previous approaches to cohesion, but can be seen as complementing the picture. If we take a more conventional approach to the cohesive devices illustrated by the above example, we can describe the cohesive function of true feelings in mainly two ways. On the one hand, true feelings packages a stretch of text, namely the words of Simpson that are presented in the form of a quotation. In this sense true feelings functions as text-organising vocabulary (see McCarthy 1991 in the above discussion) or as a ‘label’ as Francis (1994) would call vocabulary items that encapsulate stretches of text. On the other hand, feelings forms a link with views in the final sentence of the example, when we see feelings as a paraphrase not only of emotions, but also of views, or opinions. However, the link between feelings and views can be described in even greater detail by drawing on the semantic prosody of true feelings: the ‘reluctance’ to express true feelings is in the case of Simpson his attempt to distance himself from the words he cannot deny he has said: he was merely repeating other people’s views. This interpretation could not be justified by merely looking at the nouns feelings and views, which would typically be seen as the central elements in an account of lexical cohesion. A lexical relation between true and other people’s is difficult to see. In particular, since other is not a lexical, but a grammatical item in the traditional sense. Still, other is an important element for the link to the prosody of ‘reluctance’ that is part of true feelings.6

With true feelings as the core of a lexical item, the semantic preference can also be interpreted in terms of cohesive links. The concordance analysis showed that the semantic preference of ‘expression’ tends to be realised by verbs and in our example tell in if he had not told dinner party guests his true feelings is a realisation of this preference. In the short section on Richard Simpson, say and repeat are lexical repetitions of tell. Thus cohesive links between the three verbs get a further dimension through the link with true feelings.

To sum up the analysis of this example, we can say that the cohesion created by lexical items works on two dimensions. On the one hand, we have linear links
between words that are part of the realisation of a lexical item. Linear links characterise patterns of co-occurrence around the core of a lexical item. They may be described in terms of collocational patterns. On the other hand, we have non-linear links that spread across larger passages of text. Non-linear links illustrate how different lexical items merge beyond the occurrence in a sequence. These links may be interpreted as various types of lexical repetition links between the words in the narrow context around the core and words that spread across the text. In the example above, the links of *tell* with *say* and *repeat* can be viewed as non-linear and are describable as cases of simple paraphrase. They also form links with *conversation* at the beginning of the article, which could be viewed as an instance of complex paraphrase (slightly stretching Hoey’s (1991) terminology). Additionally, non-linear links may only become describable as realisations of semantic prosodies, as in the example of *other people’s views*. When it comes to semantic prosodies we enter a realm that is often seen as the more interpretative end of text analysis where we deal with elusive meanings that are to a great extent subjective. However, our interpretation was backed up by a concordance analysis that supports the observations made in a single text by generalisations found for a larger number of instances of the lexical item under investigation.

Still, the textual analysis that can be presented in the limited space of this article is only partial. The variability of lexical items is linked to a number of factors, and the lexical patterns into which *true feelings* enters will not be the same in every text. The realisation of the components of a lexical item are shaped, for instance, by the topic of the text and by the genre of the text. Moreover, we have only looked at links in the text from the point of view of a single lexical item. The picture will only be complete if we do the same for every item in the text. Then we can truly describe how the ‘interlocking’ and ‘overlapping’ of prosodies works in this text.

The important point with regard to the corpus theoretical approach is the focus on the link between lexical and textual properties. Starting with a concordance analysis we can identify detailed features of collocational patterns on the basis of huge numbers of texts. The analysis of individual texts can then reveal further detail on links in texts. These links are also part of the properties of lexical items. Within the space of this article only a few examples could be given, but it is clear that we need descriptive tools to capture patterns that go beyond what can be found in concordances. The concept of local textual functions is one suggestion for such tools. In the present section, the description of the lexical network of which *true feelings* is a part is a description of local textual functions of *true feelings*: the relationships into which the item enters in the creation of text. These relationships cover cohesive links, but also other textual relations. For instance, relations expressing contrasts, as in the Bin Laden example, or relations
expressing close links to the overall topic, as in the text dealing with accidentally overhead political insults. Such textual functions of lexical items are necessarily ‘local’, as we cannot claim that all items behave in the same way, or even that the same item behaves in the same way in different texts. We may, however, find generalisations when we look at items in texts with a similar purpose or genre, and we may find similarities when we look at items with similar frequencies (see for instance Mahlberg 2005). Against this theoretical background, we can now look at implications for language teaching.

7. Implications for the concept of cohesion in ELT

In the list in Section 3 we saw that the choice of examples, the time that is needed to analyse a text and the variety of textual features that have an impact on words in text, all make it difficult to give an account of cohesion that is useful to language learners. At first sight, the variability of lexical patterns, which has been illustrated by the example of true feelings, does not make the task any easier. However, if we are prepared to accept some fundamental changes to the theoretical framework within which we deal with cohesion, cohesion can become more manageable in ELT.

The examples in the present article suggest that cohesion is a fundamentally lexical phenomenon. For some words such as pronouns more typically ‘grammatical’ arguments may be helpful, but textual links are also to a large extent lexical and thus have to be described in lexical terms. The basic idea is easy to see, also for learners of a language: meaning is not mapped one-to-one to words but spans across words that tend to co-occur. In a text we can view words as cores of lexical items and their surroundings as realisations of lexical patterns. When explaining this idea of cohesion to learners we do not need all the theoretical background that concerns us as linguists. It is often useful to use a picture illustrating the situation of words in text in a very simple way:

-------- • ----- • --------------
----- ------------------------
--- • ------------------------
• --------------------------

The points are words that attract the attention of the reader as useful starting-points for a more detailed analysis. The lines are words whose functions in textual patterns are difficult to describe with only a first look at the text. The words represented by lines could turn out to be strong collocates or only words at the fuzzy edges of a lexical item. Information on how the lines are part of lexical items can be
gathered by doing concordance analyses. The students take the points as starting-points for a concordance analysis. The results of the concordance analysis will help them to shed light on the lines. Students will have to learn that not every item in every text will yield the same amount of useful information for a textual interpretation. The basic idea of cohesion will become clear as students gain experience in the analysis of lexical items in texts. It will be the task of the teacher to find suitable texts and provide help on which words to choose for a closer analysis. An exhaustive analysis of a single text is neither possible nor necessary to make the underlying principles clear. Such an approach to textual patterns may not be as systematic as working through neat lists of cohesive devices but it will help learners develop an awareness of properties of natural texts.

Thus a corpus theoretical approach to cohesion can help to cope with the difficulties listed in Section 3. Instead of teaching cohesion on top of vocabulary and grammar, cohesion will become a natural component of lexical items in text: general categories to describe cohesive links are part of the description of lexical items. Such a corpus theoretical approach to cohesion seems to work best in the context of a discourse approach to the teaching of vocabulary. Similar to the suggestions that McCarthy (1991) makes for the building of a ‘textually-based lexicon’ as an alternative to the random vocabulary list (McCarthy 1991:81), students could gather lexical and textual information by moving from words in texts to concordances and back to texts again. As the students gain more and more experience in this type of textual analysis, they will also be able to see similarities between lexical items and identify local textual functions that are shared by different lexical items.

With the continuing development of corpus linguistic theories, there will also be more and more reference materials that can help teachers to find the right texts and focus on the most useful words. We have seen that a lexical syllabus (see Section 4) stresses the importance of frequent words and frequent patterns, and Mahlberg (2005), for instance, provides a description of textual patterns of high-frequency nouns, which may be one of the sources to provide useful background information for the teaching of cohesion. And even if teachers cannot spend much time on corpus theoretical issues, with a tool such as WebCorp (2006) that easily provides concordance samples (cf. Section 5) there are enough possibilities to bring natural examples of cohesion into the classroom.
8. Conclusions

The present article has proposed a corpus theoretical approach to cohesion. A corpus linguistic theory provides a framework for a description of the English language that does not need a strict separation of lexis and grammar. As a result, cohesion is mainly seen from a lexical point of view. Cohesion describes the way in which the flexible boundaries of lexical items link in with other lexical items. Central to the corpus theoretical approach is the importance of language as a means of communication: language is action and meaning is use. With the focus on language in context, the corpus theoretical approach links in with communicative approaches to language teaching. In contrast to more traditional approaches, the corpus linguistic focus on communication does not stop at fixed phrases or cohesive devices described by clear-cut categories. Because of the variability that is allowed in a corpus theoretical approach, it cannot yet produce as systematic or as comprehensive textbooks as traditional approaches, but it suggests some simple ideas as a starting-point. With the help of easily accessible tools, this corpus theoretical approach can help learners to develop their experience and awareness of properties of natural texts.

Notes

* I would like to thank John Flowerdew and two anonymous reviewers for useful comments and advice.

1. Strictly speaking, Halliday and Hasan (1976:6) view conjunction as “mainly grammatical, but with a lexical component in it”.


3. Sinclair (1996) and Sinclair (1998) were later included in Sinclair (2004), so in the following I will only refer to the more recent publication.

4. Hoey (2005:116f.) introduces the notion of ‘textual collocations’ to stress the occurrence of collocations in larger textual environments.

5. The text is available at http://politics.guardian.co.uk/backbench/story/0,,1138512,00.html (last accessed March 2006).

6. When we look at a number of lexical items to observe their cohesive behaviour we will also see that there are differences between lexical items. These differences are not only due to the variability of individual lexical items, but also to the types of links that lexical items can create. In the example above, we saw how true feelings is linked to other people's views. We have seen that other people's links in with the prosody of 'reluctance'. However, the noun people is less clearly involved in the meaning relationships that form the cohesive network around true feelings. The
important element to express Simpson's desire to distance himself from his words is other. The function of people is more that of a 'support function' (cf. Mahlberg 2005): its meaning is in the background but it enables a construction that carries meaning relationships that are more central to the text.

References


Index

A
abstract noun 68, 88, 92
academic prose 26, 30
adjacency pairs 106
analysis of discourse 5
apposition 24, 32, 33, 34
argument structuring 5
Aunt Sally 6

B
Bank of English 6, 18, 67, 76
British Academic Spoken
English (BASE) corpus 28
British National Corpus 6, 90

C
classroom discourse 26, 28,
31, 32, 37
cluster 11, 25
corpus-driven 45, 66, 75,
80, 110
counter argument 19
counterstatement 16
course coherence 46, 65, 66, 68, 69,
77, 80, 87, 100, 101, 103
cohesive device 1, 12, 46, 60,
104, 106–108, 116, 119, 120
cohesive tie 6, 45, 78, 105, 107
colligation 18, 67, 68, 71–73,
75, 76, 78, 87, 90, 96, 98,
99, 111, 113
collocate 69, 72–75, 77, 78,
96, 112, 113
collocation 1, 24, 67, 87, 90,
95, 96, 98, 99, 104, 111, 113
communicative approach 108–110
complex paraphrase 117
complex repetition 105
componential device 6
componential level 19
comprehension and recall 41
concession 14
concessive mechanism 14
ConcGram© 69
conjunction 24, 27, 34, 79, 104
connotation 11, 68, 91
connotative meaning 68
contrast/concession 24, 32,
34, 35
corpus-driven 45, 66, 75,
80, 110
counter argument 19
counterstatement 16
classroom discourse 26, 28,
31, 32, 37
complex paraphrase 117
complex repetition 105
componential device 6
componential level 19
comprehension and recall 41
concession 14
concessive mechanism 14
ConcGram© 69
conjunction 24, 27, 34, 79, 104
connotation 11, 68, 91
connotative meaning 68
contrast/concession 24, 32,
34, 35
corpus-driven 45, 66, 75,
80, 110
counter argument 19
counterstatement 16
classroom discourse 26, 28,
31, 32, 37
complex paraphrase 117
complex repetition 105
componential device 6
componential level 19
comprehension and recall 41
concession 14
concessive mechanism 14
ConcGram© 69
conjunction 24, 27, 34, 79, 104
connotation 11, 68, 91
connotative meaning 68
contrast/concession 24, 32,
34, 35
corpus-driven 45, 66, 75,
80, 110
counter argument 19
counterstatement 16
classroom discourse 26, 28,
31, 32, 37
complex paraphrase 117
complex repetition 105
componential device 6
componential level 19
comprehension and recall 41
concession 14
concessive mechanism 14
ConcGram© 69
conjunction 24, 27, 34, 79, 104
connotation 11, 68, 91
connotative meaning 68
contrast/concession 24, 32,
34, 35
corpus-driven 45, 66, 75,
80, 110
counter argument 19
counterstatement 16
classroom discourse 26, 28,
31, 32, 37
complex paraphrase 117
complex repetition 105
componential device 6
componential level 19
comprehension and recall 41
concession 14
concessive mechanism 14
ConcGram© 69
conjunction 24, 27, 34, 79, 104
connotation 11, 68, 91
connotative meaning 68
contrast/concession 24, 32,
34, 35
corpus-driven 45, 66, 75,
80, 110
counter argument 19
counterstatement 16
classroom discourse 26, 28,
31, 32, 37
complex paraphrase 117
complex repetition 105
componential device 6
componential level 19
comprehension and recall 41
concession 14
concessive mechanism 14
ConcGram© 69
conjunction 24, 27, 34, 79, 104
connotation 11, 68, 91
connotative meaning 68
contrast/concession 24, 32,
34, 35
corpus-driven 45, 66, 75,
80, 110
counter argument 19
counterstatement 16
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>logical structure</td>
<td>14</td>
</tr>
<tr>
<td>Longman Spoken and Written English Corpus</td>
<td>26</td>
</tr>
<tr>
<td>M</td>
<td></td>
</tr>
<tr>
<td>metaphorical expression</td>
<td>19</td>
</tr>
<tr>
<td>Michigan Corpus of Academic Spoken English (MICASE)</td>
<td>28</td>
</tr>
<tr>
<td>monologue</td>
<td>27, 28</td>
</tr>
<tr>
<td>N</td>
<td></td>
</tr>
<tr>
<td>newspaper headline</td>
<td>5, 8</td>
</tr>
<tr>
<td>non-lexical cohesion</td>
<td>46, 47, 51, 59</td>
</tr>
<tr>
<td>nucleus</td>
<td>8</td>
</tr>
<tr>
<td>O</td>
<td></td>
</tr>
<tr>
<td>'oral' lexical bundle</td>
<td>26, 28, 29</td>
</tr>
<tr>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Papers93</td>
<td>6</td>
</tr>
<tr>
<td>peroration</td>
<td>12</td>
</tr>
<tr>
<td>polysemous word</td>
<td>12</td>
</tr>
<tr>
<td>prepositional phrase</td>
<td>14, 16, 24, 26</td>
</tr>
<tr>
<td>priming</td>
<td>11</td>
</tr>
<tr>
<td>problem-solution pattern</td>
<td>105</td>
</tr>
<tr>
<td>prominence</td>
<td>47–49, 51–56, 58–60</td>
</tr>
<tr>
<td>public discourse</td>
<td>66</td>
</tr>
<tr>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>qualitative approach</td>
<td>5, 66</td>
</tr>
<tr>
<td>quantitative approach</td>
<td>66</td>
</tr>
<tr>
<td>R</td>
<td></td>
</tr>
<tr>
<td>reference</td>
<td>1, 26, 33–37, 46, 104</td>
</tr>
<tr>
<td>reiteration</td>
<td>1, 46, 104</td>
</tr>
<tr>
<td>relexicalisation</td>
<td>46</td>
</tr>
<tr>
<td>repetition</td>
<td>1, 10, 46, 104, 105, 108, 117</td>
</tr>
<tr>
<td>result/inference</td>
<td>24, 32, 35, 36</td>
</tr>
<tr>
<td>rhetorical device</td>
<td>13</td>
</tr>
<tr>
<td>rhetorical effect</td>
<td>6</td>
</tr>
<tr>
<td>rhetorical function</td>
<td>8</td>
</tr>
<tr>
<td>rhetorical movement</td>
<td>5</td>
</tr>
<tr>
<td>rhetorical prosody</td>
<td>17</td>
</tr>
<tr>
<td>S</td>
<td></td>
</tr>
<tr>
<td>satellite</td>
<td>8</td>
</tr>
<tr>
<td>semantic field</td>
<td>5, 6</td>
</tr>
<tr>
<td>semantic preference</td>
<td>67, 68, 71, 73, 74–79, 111–114, 116</td>
</tr>
<tr>
<td>string</td>
<td>11, 25</td>
</tr>
<tr>
<td>string length</td>
<td>25</td>
</tr>
<tr>
<td>structuring</td>
<td>12</td>
</tr>
<tr>
<td>stylistic variation</td>
<td>12</td>
</tr>
<tr>
<td>sub-head</td>
<td>9</td>
</tr>
<tr>
<td>summation/conclusion</td>
<td>24, 32, 38–40</td>
</tr>
<tr>
<td>T</td>
<td></td>
</tr>
<tr>
<td>thematic position</td>
<td>14</td>
</tr>
<tr>
<td>TOEFL 2000 Spoken and Written Academic Language Corpus (T2K SWAL)</td>
<td>26, 27</td>
</tr>
<tr>
<td>topic introduction/focus</td>
<td>37–39</td>
</tr>
<tr>
<td>transition</td>
<td>24, 32, 37–40</td>
</tr>
<tr>
<td>V</td>
<td></td>
</tr>
<tr>
<td>vocabulary</td>
<td>3, 105</td>
</tr>
<tr>
<td>W</td>
<td></td>
</tr>
<tr>
<td>WebCorp</td>
<td>111</td>
</tr>
<tr>
<td>writer's commitment</td>
<td>14</td>
</tr>
<tr>
<td>specialised corpus</td>
<td>49, 50, 69, 80, 81</td>
</tr>
<tr>
<td>stance</td>
<td>19, 26</td>
</tr>
<tr>
<td>Straw man, Man of straw</td>
<td>13</td>
</tr>
<tr>
<td>string length</td>
<td>11, 25</td>
</tr>
<tr>
<td>structuring</td>
<td>12</td>
</tr>
<tr>
<td>stylistic variation</td>
<td>12</td>
</tr>
<tr>
<td>sub-head</td>
<td>9</td>
</tr>
<tr>
<td>summation/conclusion</td>
<td>24, 32, 38–40</td>
</tr>
<tr>
<td>thematic position</td>
<td>14</td>
</tr>
<tr>
<td>TOEFL 2000 Spoken and Written Academic Language Corpus (T2K SWAL)</td>
<td>26, 27</td>
</tr>
<tr>
<td>topic introduction/focus</td>
<td>37–39</td>
</tr>
<tr>
<td>transition</td>
<td>24, 32, 37–40</td>
</tr>
<tr>
<td>V</td>
<td></td>
</tr>
<tr>
<td>vocabulary</td>
<td>3, 105</td>
</tr>
<tr>
<td>W</td>
<td></td>
</tr>
<tr>
<td>WebCorp</td>
<td>111</td>
</tr>
<tr>
<td>writer's commitment</td>
<td>14</td>
</tr>
</tbody>
</table>
In the series *Benjamins Current Topics (BCT)* the following titles have been published thus far or are scheduled for publication:

14  **BAKER, Anne and Bencc WOLLE** (eds.): Sign Language Acquisition. 2008. xi, 163 pp. + index.
13  **ABRY, Christian, Anne VILAIN and Jean-Luc SCHWARTZ** (eds.): Vocalize to Localize. ca. 320 pp. *Expected Forthcoming*
12  **DROR, Itiel E.** (ed.): Cognitive Technologies and the Pragmatics of Cognition. 2007. xii, 186 pp.
10  **LIEBAL, Katja, Cornelia MÜLLER and Simone PIKA** (eds.): Gestural Communication in Nonhuman and Human Primates. 2007. xiv, 284 pp.
8   **TEUBERT, Wolfgang** (ed.): Text Corpora and Multilingual Lexicography. 2007. x, 162 pp.
7   **PENKE, Martina and Anette ROSENBACH** (eds.): What Counts as Evidence in Linguistics. The case of innateness. 2007. x, 297 pp.
5   **ANTHONISSEN, Christine and Jan BLOMMAERT** (eds.): Discourse and Human Rights Violations. 2007. x, 142 pp.