

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 conjunction¹), but they also discuss lexical cohesion. Lexical cohesion can occur in the form of ‘reiteration’ or ‘collocation’. Reitera-

tion 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).

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 textbuilding 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 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-driven² 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', 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, build-

ing 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 linguistics has also developed approaches that are more 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 categories and 'collocation', 'colligation', and 'semantic preference' are optional. One of the examples that Sinclair (2004) discusses is the lexical item with the core *true feelings*. The core of a lexical item is invariable, but moving further away from the core, patterns become more variable. One of the categories to describe the patterns around the core is 'collocation'. According to Sinclair (2004:141) collocation is "the co-occurrence of words with no more than four intervening words". The core *true feelings* is itself a collocation. Below you find 30 concordance lines for *true feelings* derived from the internet with the tool webcorp (2006). The search was restricted to the Newspaper Domains 'UK broadsheets' (accessed January 2006).

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." Benítez 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 Xiyan, 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 *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 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.

(“Bin Laden taunts the West: ‘I’m ready to die’”, Jason Burke, The Observer, Nov 11, 2001, Copyright Guardian Newspapers Limited 2001)

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*:⁵

‘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 feeling* packages a stretch of text, namely the words of Simpson that are presented in the form of a quotation. In this sense *true feel-*

ings 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 *view* 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 *view* 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 preferences. 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.

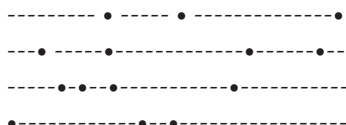
Still, the textual analysis that can be presented in the limited space of this article is only partial. The variability of lexical items is linked to a number of factors, and the lexical patterns into which *true feelings* enters will not be the same in every text. The realisation of the components of 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 seems 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 of 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 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 concordances 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”.

2. On the definition of ‘corpus-driven’ see Tognini-Bonelli (2001).

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 view*. 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

- Beaugrande, R. A. de & Dressler, W. U. (1981). *Introduction to Text Linguistics*. London: Longman.

- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Cook, G. (1989). *Discourse*. Oxford: Oxford University Press.
- Francis, G. (1994). Labelling discourse: an aspect of nominal-group lexical cohesion. In M. Coulthard (Ed.), *Advances in Written Text Analysis* (pp. 83–101). London: Routledge.
- Francis, G. (1986). *Anaphoric Nouns* (Discourse Analysis Monograph 11). English Language Research, Birmingham: University of Birmingham.
- Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hasan, R. (1984). Coherence and cohesive harmony. In J. Flood (Ed.), *Understanding Reading Comprehension* (pp. 181–219). Delaware: International Reading Association.
- Hatch, E. (1992). *Discourse and Language Education*. Cambridge: Cambridge University Press.
- Hoey, M. (2005). *Lexical Priming. A new theory of words and language*. London: Routledge.
- Hoey, M. (2001). *Textual Interaction. An Introduction to Written Discourse Analysis*. London: Routledge.
- Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Howatt, A. P. R. & Widdowson, H. G. (2004). *A History of English Language Teaching*. Oxford: Oxford University Press.
- Hunston, S. & Francis, G. (2000). *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam/Philadelphia: Benjamins.
- Johns, T. (1991). Should you be persuaded: two examples of data-driven learning materials. *English Language Research Journal*, 4, 1–16.
- Lewis, M. (1993). *The Lexical Approach. The State of ELT and a Way Forward*. Hove: Language Teaching Publications.
- Lewis, M. (Ed.) (2000). *Teaching Collocation. Further Developments in the Lexical Approach*. Hove: Language Teaching Publications.
- Mahlberg, M. (2005). *English General Nouns: A Corpus Theoretical Approach*. Amsterdam/Philadelphia: Benjamins.
- Martin, J. R. (1992). *English Text. System and Structure*. Amsterdam/Philadelphia: Benjamins.
- McCarthy, M. (2001). *Issues in Applied Linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. (1991). *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.
- Morley, J. (1999). Sticky business: a case study of cohesion in the language of politics in the *Economist*. In L. Lombardo, L. Haarman, J. Morley & C. Taylor. *Massed Medias. Linguistic Tools for Interpreting Media Discourse* (pp. 18–83). Milano: LED.
- Mukherjee, J. (2002). *Korpuslinguistik und Englischunterricht. Eine Einführung*. Frankfurt a. Main: Peter Lang.
- Müller-Hartmann, A. & Schocker-von Dittfurth, M. (2004). *Introduction to English Language Teaching*. Stuttgart: Klett.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Salkie, R. (1995). *Text and Discourse Analysis*. London: Routledge.

- Sinclair, J. McH. (2004). *Trust the Text. Language, corpus and discourse*. London: Routledge.
- Sinclair, J. McH. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive Lexical Semantics* (pp. 1–24). Amsterdam/Philadelphia: Benjamins.
- Sinclair, J. McH. (1996). The search for units of meaning. *Textus*, 9 (1), 75–106.
- Sinclair, J. McH. & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140–160). London & New York: Longman.
- Stubbs, M. (2001a). Computer-assisted text and corpus analysis: lexical cohesion and communicative competence. In D. Schiffrin, D. Tannen & H. E. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp.304–320). Malden Massachusetts, Oxford: Blackwell.
- Stubbs, M. (2001b). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Teubert, W. (1999). *Corpus linguistics — a partisan view*. Available at: http://tractor.bham.ac.uk/ijcl/teubert_cl.html, last accessed: January 04.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam/Philadelphia: Benjamins.
- webcorp (2006). Research and Development Unit for English Studies. Available at: <http://www.webcorp.org.uk/>, last accessed: February 2006.
- Werlich, E. (1988). *Student's Guide to Text Production*. Berlin: Cornelsen.
- Whitsitt, S. (2005). A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics*, 10 (3), 283–305.
- Willis, D. (2003). *Rules, Patterns and Words: Grammar and Lexis in English Language Teaching*. Cambridge: Cambridge University Press.
- Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. London: Collins COBUILD. Out of print, available at: <http://www.cels.bham.ac.uk/resources/LexSyll.htm>, last accessed: September 2005.
- Winter, E. O. (1977). A clause-relational approach to English texts: a study of some predictive lexical items in written discourse. *Instructional Science*, 6, 1–92.